

# LLM gesteuerte Planung zur Reduktion des CO<sub>2</sub>-Fußabdrucks im Tiefbau

André Matutat <sup>1</sup>, Lena Golin <sup>1</sup> und Hans Brandt-Pook <sup>1</sup>

**Abstract:** Der Beitrag beschreibt ein prototypisches KI-System zur Unterstützung nachhaltiger Planungsprozesse im Tiefbau. Im Zentrum steht ein Retrieval-Augmented-Generation-System (RAG), das große Sprachmodelle (engl.: Large Language Models, LLMs) mit unternehmensspezifischem Fachwissen verknüpft, um Nutzer dialogisch bei der Informationssuche und Projektplanung zu unterstützen. Am Beispiel des mittelständischen Unternehmens Nauw wird gezeigt, wie ein solcher Ansatz sowohl textbasierte Wissensquellen als auch interaktive Tools – etwa einen CO<sub>2</sub>-Kalkulator – integrieren kann. Besonderes Augenmerk gilt der Einbindung visueller Inhalte sowie der Anbindung externer Applikationen. Durch die Kombination aus Wissensmanagement, Tool-Integration und dialogischer Interaktion entsteht ein Assistenzsystem, das nicht nur die Planungsqualität verbessert, sondern auch zur Reduktion des CO<sub>2</sub>-Fußabdrucks beitragen kann.

**Keywords:** Large Language Models, Retrieval-Augmented Generation, CO<sub>2</sub>-Bilanzierung, Tiefbau, Nachhaltigkeit, Assistenzsysteme, Wissensmanagement, Tool-Integration

## 1 Einleitung

Die Baubranche steht vor der dringenden Herausforderung, ihren Beitrag zur Reduktion von Treibhausgasemissionen zu leisten. Insbesondere im Tiefbau entstehen große Mengen an CO<sub>2</sub>. Gleichzeitig wachsen die Anforderungen an Planung und Dokumentation – sowohl im Hinblick auf ökologische Standards als auch auf Effizienz und Qualitätssicherung.

Die mittelständische Nauw GmbH befasst sich mit der Herstellung von Geobaustoffen, insbesondere in den Bereichen Umweltgeotechnik, Wasserwirtschaft, Verkehrswegebau und unterirdisches Bauen. Seit vier Jahren hat man in verschiedenen Projekten den Einsatz von KI-Lösungen evaluiert. Im vorliegenden Beitrag wird eines dieser Projekte vorgestellt: das prototypische RAG-gestützte Assistenzsystem ermöglicht nicht nur Wissensabfragen, sondern integriert auch konkrete Werkzeuge zur Projektplanung im Tiefbau – wie etwa einen CO<sub>2</sub>-Kalkulator. Ziel ist es, durch die Verbindung von Wissensmanagement und interaktiver Assistenz die Nachhaltigkeit im Tiefbau messbar zu verbessern und gleichzeitig die Planungsqualität zu erhöhen.

Im Rahmen der KI-Initiative wurde das Nauw-Wiki, ein unternehmensinternes RAG, entwickelt. Die erste Implementierung des RAG-Systems erfolgte nach bekannten Ansätzen

<sup>1</sup> Hochschule Bielefeld, Interaktion 1, 33619 Bielefeld, Deutschland,  
andre.matutat@hsbi.de,  <https://orcid.org/0009-0001-9362-7704>;  
lena.golin@hsbi.de,  <https://orcid.org/0009-0002-2963-8430>;  
hans.brandt-pook@hsbi.de,  <https://orcid.org/0009-0002-6668-2684>

[Ga23]. Eine rollenbasierte Zugriffssteuerung stellt zudem sicher, dass verschiedene Nutzergruppen jeweils nur die für sie relevanten und zugänglichen Informationen erhalten. In der folgenden Iteration des Wikis rückten zwei Aspekte in den Mittelpunkt: die besondere Bedeutung von Grafiken in den Wissensdokumenten und die Integration bestehender unternehmensinterner Applikationen.

## 2 Grafiken im RAG

Ein besonderer und wichtiger Aspekt des Naue-Wiki liegt in der Einbindung grafischer Inhalte, der mit spezifischen Herausforderungen einhergeht – insbesondere in der Vorverarbeitung der Wissensbasis, konkret bei der Dokumentvorverarbeitung von PDF-Dokumenten. Gängige Tools zur Segmentierung, wie die Adobe PDF Extract API (<https://developer.adobe.com/document-services/docs/overview/pdf-extract-api/>), erzeugen dabei eine strukturierte Repräsentation der Dokumente, in der Textabschnitte und Grafiken als separate Einheiten mit Metadaten wie Position und Layoutinformationen erfasst werden.

Abhängig vom Aufbau des jeweiligen Dokuments kann es dabei vorkommen, dass Abbildungen von ihrem zugehörigen Textabschnitt getrennt werden. Bei der anschließenden Umwandlung der Chunks in Vektorrepräsentationen ist es daher notwendig, eine korrekte Zuordnung zwischen Text- und Bildelementen herzustellen, um die inhaltliche Integrität und den Kontextbezug zu wahren. Die Zuordnung erfolgt dabei, indem das Bild als Metadaten im zugehörigen Text-Vektor hinterlegt wird. Wird dieser Vektor bei einer semantischen Suche vom RAG-System ausgewählt, kann das Bild direkt mit dem Text angezeigt werden.

Im Projektverlauf wurden verschiedene Szenarien identifiziert, die im Folgenden vorgestellt und mit entsprechenden Lösungskonzepten erläutert werden.

### 2.1 Eindeutig zusammengehörige Bild-Text-Paare

In vielen Dokumenten befinden sich erklärende Texte direkt über oder unter einer zugehörigen Abbildung. Diese Zusammengehörigkeit wird auch in der strukturierten Repräsentation der Dokumente abgebildet. In solchen Fällen ist die Beziehung zwischen Text und Bild durch die räumliche Nähe klar definiert. Die Zuordnung kann auf Basis der Dokumentenstruktur mit hoher Zuverlässigkeit vorgenommen werden.

### 2.2 Zusammengehörige Bild-Text-Paare ohne explizite Verknüpfung

Dieser Fall ähnelt dem zuvor beschriebenen Szenario, unterscheidet sich jedoch dadurch, dass die räumliche Nähe zwischen Grafik und Text zwar vorhanden ist, jedoch nicht explizit in der strukturierten Repräsentation des Dokuments als verknüpft angegeben wird.

Grafik und zugehöriger Textabschnitt erscheinen dort lediglich als aufeinanderfolgende Elemente. Die Verknüpfung wird hier durch die Erkennung einer typischen Bildunterschrift hergestellt. Solche Bildunterschriften folgen häufig bestimmten sprachlichen Mustern, die sich zuverlässig mithilfe regulärer Ausdrücke oder eines Sprachmodells identifizieren lassen. Wird so eine Struktur erkannt, kann das entsprechende Textelement als beschreibender Kontext zur benachbarten Grafik interpretiert und als zugehörig markiert werden.

### **2.3 Abbildungen mit explizitem Verweis im Text und Bildunterschrift**

Der Text enthält Verweise wie „siehe Abbildung 1“, während sich die zugehörige Grafik mit entsprechender Beschriftung an anderer Stelle befindet. Solche Referenzen müssen korrekt erkannt und aufgelöst werden. Für Grafiken mit Bildunterschriften erfolgt die Zuordnung, indem jedes Textelement in der Datenstruktur daraufhin untersucht wird, ob es sich um die passende Bildunterschrift zum gesuchten Bild handelt. Die Entscheidung darüber trifft ein LLM. Sobald die korrekte Bildunterschrift identifiziert wurde, kann über die oben beschriebenen Verfahren das entsprechende Bild gefunden und dem ursprünglichen Textchunk zugeordnet werden.

### **2.4 Verweise ohne Bildunterschrift, aber mit positionsbezogenen Hinweisen**

In weniger strikt formulierten Texten kommt es zu vagen Formulierungen wie „siehe Abbildung oben“ oder „siehe Abbildung unten“, ohne dass eine spezifische Bildunterschrift oder Nummerierung vorhanden ist. Da die Abbildung selbst nicht näher beschrieben wird, ist eine inhaltliche Zuordnung über Referenzen nicht möglich.

Stattdessen erfolgt die Verknüpfung auf Basis der Layoutinformationen: In der zugrundeliegenden Datenstruktur wird ermittelt, welches Bild dem Verweis im Text am nächsten steht – entweder unmittelbar davor („oben“) oder danach („unten“). Auf diese Weise kann eine plausible Zuordnung auch ohne explizite Beschreibung oder Bildunterschrift erfolgen.

### **2.5 Herausforderungen**

Viele Probleme ergeben sich aus der unsystematischen Erstellung der Dokumente. Funktionen wie Bildunterschriften- oder Tabellenformate werden häufig nicht genutzt; stattdessen erscheinen Beschriftungen als Fließtext oder sind in Screenshots eingebettet, was ihre automatisierte Verarbeitung erschwert oder verhindert.

Besonders herausfordernd ist der Umgang mit Bildserien oder thematisch zusammenhängenden Abbildungen. Für eine korrekte semantische Einordnung müssen diese gemeinsam im Kontext interpretiert werden. Ein verlässlicher Ansatz zur automatisierten Erkennung solcher Bildgruppen existiert bislang nicht.

### **3 Anbindung von Applikationen: Dialog und Tooling mit RAG**

Ergänzend zu der oben beschriebenen Wissensbereitstellung durch ein RAG-System wurde im Projekt evaluiert, wie sich interaktive Applikationen an ein solches System anbinden lassen, um Aufgaben jenseits der reinen Wissensabfrage zu realisieren. In der Fachliteratur werden derartige Architekturen, die generative Sprachmodelle mit externen Funktionalitäten wie Datenbanken oder Tools kombinieren, unter dem Begriff Augmented Language Models (ALMs) zusammengefasst [Mi23]. Als exemplarische Anwendung wurde ein CO<sub>2</sub>-Kalkulator entwickelt, der Anwender bei der Bilanzierung des CO<sub>2</sub>-Fußabdrucks konkreter Bauvorhaben unterstützt.

Funktional lässt sich das Gesamtsystem als Agentensystem beschreiben. Basierend auf Nutzereingaben wählt dieses passende Aktionen (Wissensabfrage vs. Tool-Nutzung) und orchestriert damit das Zusammenspiel von RAG und angeschlossenen Applikationen. Im Unterschied zu Tool-Augmented Language Models (TALMs), bei denen das Sprachmodell eigenständig lernt, wann und wie externe Tools einzusetzen sind, basiert die hier entwickelte Lösung auf einer fest implementierten, regelbasierten Entscheidungslogik.

Während neuere Entwicklungen im Bereich der Tool-Augmented Language Models (z.B. [Sc23]) darauf abzielen, die Nutzung externer Tools vollständig durch das Sprachmodell selbst steuern zu lassen, zeigt die hier vorgestellte Lösung die Vorteile einer domänen spezifischen, regelbasierten Entscheidungslogik: Sie erlaubt eine hohe Kontrolle, Robustheit und Nachvollziehbarkeit – zentrale Anforderungen im industriellen Anwendungskontext. Eine perspektivische Erweiterung des Systems um selbstlernende Komponenten wäre möglich, müsste jedoch mit Blick auf Verlässlichkeit und Sicherheit sorgfältig evaluiert werden.

Der Chatbot sammelt relevante Parameter aus der Konversation, um eine Massenkalkulation sowie die CO<sub>2</sub>-Bilanzierung einzuleiten. Das System nutzt die API-Definitionen der unternehmensinternen Kalkulationswerkzeuge, um die je nach Typ des Bauvorhabens benötigten Parameter zu erkennen.

Durch die Verzahnung von dialogischer Interaktion, Wissensbereitstellung und angebundenen Applikationen wie dem CO<sub>2</sub>-Rechner entsteht ein System, das sowohl beratende als auch planende Funktionen erfüllt. Nutzer können ein Bauvorhaben schrittweise im Dialog mit dem Chatbot konkretisieren, der dabei sowohl auf das unternehmensinterne Wissen zur Beantwortung fachlicher Rückfragen als auch auf spezifische Kalkulationswerkzeuge für Mengen, Materialeinsatz und Transportwege zugreift.

Zur Emissionsbewertung werden standardisierte Daten aus Environmental Product Declarations (EPDs) sowie *mobitool Emissionsfaktoren* ([https://www.mobitool.ch/admin/data/files/tool/tool\\_file\\_de/5/mobitool-faktoren-v3.0-20240425.xlsx?lm=1714079448](https://www.mobitool.ch/admin/data/files/tool/tool_file_de/5/mobitool-faktoren-v3.0-20240425.xlsx?lm=1714079448)) herangezogen [Ma06]. In der aktuellen Umsetzung berücksichtigt der CO<sub>2</sub>-Rechner ausschließlich die Emissionen der Module A1–A4 gemäß der europäischen Norm EN 15804 – also jene, die bei der Herstellung und dem Transport der Produkte anfallen. Emissionen, die im weiteren

Lebenszyklus entstehen (Einbau, Nutzung, Rückbau oder Entsorgung; Module A5 bis D), werden bislang nicht einbezogen, sollen aber perspektivisch integriert werden [Um21].

## 4 Zwischenfazit und Ausblick

Mit den entwickelten Tools lassen sich schnelle, nachvollziehbare und aussagekräftige Vorhersagen über den CO<sub>2</sub>-Fußabdruck eines Bauprojekts treffen. Gleichzeitig unterstützen die Tools dabei, fundierte Entscheidungen im Hinblick auf den CO<sub>2</sub>-Fußabdruck zu treffen. Solche Vorhersagen können genutzt werden, um umweltfreundlichere Bauvorhaben gezielt zu identifizieren, zu bewerten und zu vermarkten.

RAG-Systeme bieten das Potenzial, die Informationsverfügbarkeit und Produktivität deutlich zu steigern – insbesondere, wenn sie domänen spezifisch angepasst werden [LG21]. In der praktischen Umsetzung zeigt sich jedoch, dass gerade der Umgang mit visuellen Inhalten wie technische Zeichnungen oder erklärenden Bildern komplex bleibt. Ihre sinnvolle Einbindung ist technisch anspruchsvoll und stark abhängig von der Dokumentenstruktur, der Konsistenz der Formatierung sowie dem jeweiligen Anwendungskontext.

Ein nächster Entwicklungsschritt zielt darauf ab, Bilder nicht nur als ergänzendes Medium zum Text, sondern als eigenständige Informationsquelle zu betrachten in den RAG-Prozess zu integrieren. Dazu sollen die Grafiken selbst vektorisiert werden, um sie semantisch durchsuchbar zu machen. Auf diese Weise könnten Nutzer künftig gezielt nach relevanten Abbildungen suchen – etwa durch Fragen wie „Zeige mir ein Beispiel für den Aufbau eines Regenrückhaltebeckens“. Dies eröffnet neue Möglichkeiten für die Nutzung visueller Inhalte in wissensintensiven Anwendungsfeldern. Dieses Projekt wird gefördert durch Deutsche Bundesstiftung Umwelt (<https://www.dbu.de/>) im Projekt QuKiBau (Förderkennzeichen: 39570/01-25).

## Literaturverzeichnis

- [Ga23] Gao, Y. et al.: Retrieval-Augmented Generation for Large Language Models: A Survey. arXiv preprint arXiv:2312.10997, 2023.
- [LG21] Lundborg, M.; Gull, I.: Künstliche Intelligenz im Mittelstand–So wird KI für kleine und mittlere Unternehmen zum Game Changer. Begleitforschung Mittelstand-Digital WIK-Consult, Bad Honnef, 2021.
- [Ma06] Manzini, R. et al.: Assessing environmental product declaration opportunities: a reference framework. Business strategy and the environment 15 (2), S. 118–134, 2006.
- [Mi23] Mialon, G. et al.: Augmented Language Models: a Survey. arXiv preprint arXiv:2302.07842, 2023.
- [Sc23] Schick, T. et al.: Toolformer: Language Models Can Teach Themselves to Use Tools. Advances in Neural Information Processing Systems 36, S. 68539–68551, 2023.

- [Um21] Umweltbundesamt: Environmental product declarations for construction products, 2021, <https://www.umweltbundesamt.de/en/topics/economics-consumption/products/building-products/environmental-product-declarations-for-construction#elements-of-an-epd>, Stand: 03.04.2025.