

Wie vermittle ich meiner KI Branchenwissen aus dem Bauwesen?

KI-Sprachmodelle stoßen bei bauspezifischen Aufgaben schnell an ihre Grenzen, etwa bei der Bemessung eines Holzträgers nach Eurocode. Ursache ist meist das Fehlen von Fachwissen im Trainingsdatensatz. Durch Prompt Engineering lässt sich die Ausgabequalität verbessern, indem der KI präzise Rollen und Kontexte vorgegeben werden. Ergänzend ermöglicht RAG (Retrieval Augmented Generation) den Zugriff auf externe Datenquellen wie Normen oder Richtlinien, ohne das Modell selbst zu verändern. Fine-Tuning hingegen integriert dauerhaft spezifisches Wissen direkt in das Modell. Eine Kombination aller Ansätze könnte eine fundierte Grundlage für den erfolgreichen Einsatz von KI in der Bauwirtschaft schaffen.



Markus Nussbaum

Die Bauherrenhilfe stellt allen Mitgliedern und Partnern einen „KI-Beauftragten“ zur Seite! Markus Nussbaum hat die HTL für Hochbau sowie Bauingenieurwesen studiert und bereits eine jahrelange Berufserfahrung hinter sich gebracht. Er ist zudem zertifizierter KI-Manager.

Kontakt:
m.nussbaum@bauherrenhilfe.org

In der Interaktion mit bekannten KI-Chatbots wie ChatGPT oder DeepSeek merkt man relativ schnell, dass die Qualität des Outputs (also der Ausgabe der KI) an ihre Grenzen stößt, wenn es um bauspezifische Themen geht. In einem Leistungstest habe ich mehrere KI-Modelle (z. B. GPT-4o oder Llama 3.3) mit der Aufgabe betraut, einen Holzträger gemäß Eurocode zu bemessen. Schon beim richtigen Einsatz von Beiwerten kam es zu ersten Problemen. Zur Erinnerung: Solche KI-Modelle werden auch als LLMs (Large Language Models bzw. große Sprachmodelle) bezeichnet. Ein LLM erhält durch meine Chateingabe einen Input, der verarbeitet und ausgewertet wird, um zu „verstehen“, welcher Output am wahrscheinlichsten zu meiner Eingabe passt. Diese Eingabe wird auch als „Prompt“ bezeichnet, da ich damit einen Befehl oder eine Anfrage an das KI-Modell richte. Ein kleines Beispiel zur Veranschaulichung:

Meine Anfrage (Eingabe) an ein beliebiges KI-Modell:

„Was ist die Hauptstadt von Österreich?“

Die Ausgabe des KI-Modells:

„Die Hauptstadt von Österreich ist Wien.“

Die Antwort ist selbstverständlich richtig, das ist doch logisches Allgemeinwissen, oder? Stellen wir uns ein KI-Modell einmal wie ein Kind vor, dem die unterschiedlichsten Dinge beigebracht werden. Wenn ich diesem Kind nie erkläre, dass Wien die Hauptstadt von Österreich ist, wird es diese Frage auch niemals beantworten können. Ähnlich funktioniert es bei KI-Modellen.

Wenn wir in ChatGPT (der Chatoberfläche) mit dem KI-Modell (z. B. GPT-4o) interagieren, „sprechen“ wir mit einem trainierten KI-System. Analog zum Kind kann man sagen: Es hat die Schule abgeschlossen und geht nun hinaus in die weite Welt. Die Schule „trainiert“ das Kind. Bevor wir also mit einem KI-Chatbot sinnvoll interagieren können, muss er trainiert worden sein, genauso wie ein Kind ein gewisses „Weltverständnis“ aufbauen muss. In diesem Training werden dem KI-Modell unzählige Daten über die Welt gezeigt (Internetseiten, Bücher, Allgemeinwissen, Mathematik usw.). Unter anderem auch Informationen über die Hauptstadt Österreichs. Das Modell hat im Training erkannt, dass in den meisten Daten die Hauptstadt „Österreichs“ mit „Wien“ verknüpft wird und merkt sich das entsprechend.

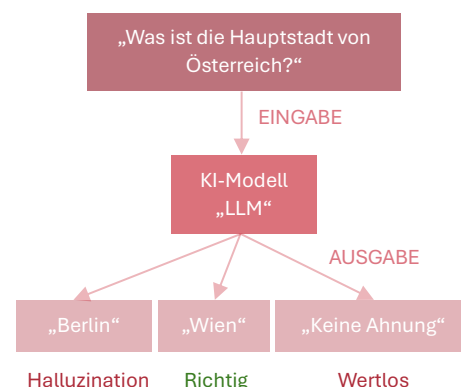


Abbildung 1: Mögliche Outputs des LLM für einen Input

Wenn der vollständige Eurocode nicht im Datensatz für das Training enthalten war, kann das KI-Modell gar nicht wissen, an welcher Stelle welcher Beiwert einzusetzen ist. Problematischer ist jedoch die Tatsache, dass die KI dennoch selbstbewusst den Holzträger mit irgendwelchen Beiwerten

bemisst und dabei Nachweise als erfüllt betrachtet, die in Wirklichkeit gar nicht erfüllt sind, das Modell „halluziniert“. Der KI ist der grobe Inhalt des Eurocodes natürlich bekannt, da dieses „Wissen“ teilweise im Internet verfügbar ist und das Modell mit solchen Internetdaten trainiert wurde. Fachwissen aus dem Bauwesen ist jedoch leider nicht so offen zugänglich wie zum Beispiel Wissen aus der Informatik. In der Informatik werden Inhalte deutlich häufiger „open source“ miteinander geteilt als im Bauwesen. Wir halten fest: Daten aus dem Bauwesen werden eher zurückhaltend (privat) behandelt. Das beginnt schon damit, dass gesetzlich verpflichtende Eurocodes (als Normen bzw. Standards) nicht frei im Internet einsehbar sind. Wie kann man einem KI-Modell also neues Wissen über das Bauwesen „vermitteln“?

Prompt Engineering

Es gibt verschiedene Möglichkeiten, einem KI-Modell spezifisches Bauwissen zu vermitteln. Die erste Variante hat nicht direkt mit „Wissensvermittlung“ zu tun, vielmehr geht es um die bereits erwähnten Prompts, also die Anweisungen, die man dem KI-Modell zu Beginn gibt. Wenn im KI-Modell irgendwo Wissen über das Bauwesen verankert ist, kann ich dieses durch einen gut formulierten Prompt abrufen. Ein Prompt liefert der jeweiligen KI den nötigen Kontext: Was will ich? In welchem Stil? Wie ausführlich soll die Antwort sein? Ein Beispiel soll zwei unterschiedliche Prompts (einen schlechten und einen guten) gegenüberstellen:

Prompt A: „Bemesse mir einen Holzträger auf Biegung als Einfeldträger mit einer Spannweite von 6 m und mit einer üblichen Belastung.“

Prompt B: „Du bist Bauingenieur und arbeitest in einem Ingenieurbüro für Tragwerkslehre. Führe eine Tragfähigkeitsbemessung nach ÖNORM EN 1995-1-1 und ÖNORM B 1995-1-1 für einen Holzträger im Grenzzustand der Tragfähigkeit durch. Gegeben sind:

- Träger Brettschichtholz GL 24h mit 120/280 mm

- Statisch bestimmter Einfeldträger, gelenkig gelagert
- Spannweite: 3,80 m
- Bemessungswert der Streckenlast: 15,8 kN/m
- Nutzungsklasse: NKL 1
- KLED: kurz

Ich benötige eine nachprüfbare Nachweisführung gemäß der ONR 24005.

Führe den Nachweis für Biegung Schritt für Schritt durch und gib an, ob der Biegespannungsnachweis erfüllt ist.“

Es lässt sich erahnen, dass Prompt B eine deutlich bessere Ausgabequalität erzielen wird, auch wenn es sehr wahrscheinlich ist, dass die KI mit der ONR 24005 nichts anfangen kann. In diesem Prompt wird die KI gezielt in die Rolle eines Bauingenieurs versetzt und es wird mehr Kontext bereitgestellt. Die gezielte Optimierung des Outputs durch die reine Veränderung des Inputs wird als „Prompt Engineering“ bezeichnet. Mit dieser Technik arbeitet man mit dem, was das fertig trainierte KI-Modell bereits „von sich aus“ zu bieten hat, also beispielsweise ohne zusätzliches Training.

RAG als externe Wissensquelle

Eine Möglichkeit, einem KI-Modell externes Wissen bereitzustellen, ist RAG (Retrieval-Augmented Generation). Dabei wird das eigentliche KI-Modell nicht verändert. Dem LLM wird ein erweiterter (augmentierter) Input bereitgestellt, also zusätzlicher Kontext, der die jeweilige Benutzeranfrage (die Eingabe) beantworten kann. Abbildung 2 veranschaulicht diese externe Informationsbereitstellung.

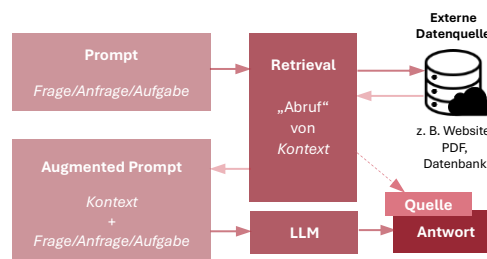


Abbildung 2: Externe Informationsbereitstellung für das LLM durch RAG¹

¹ Fraunhofer IESE (2024): Zuverlässiger Einsatz von Large Language Models (LLMs), Teil 2: Retrieval Augmented Generation (RAG), S. 9.

Man kann sich das etwa wie eine Person mit einem Lexikon in der Tasche vorstellen. Wird diese Person etwas Spezifisches gefragt, holt sie sich die passenden Informationen zur Beantwortung der Frage aus dem Lexikon. Im Fall des KI-Chatbots lässt sich dieses Prinzip durch ein „manuelles RAG“ umsetzen, zum Beispiel durch das Hochladen des Eurocodes als PDF-Datei im Chat. Bevor man also der KI die Aufgabe aus Prompt B stellt, lädt man einfach den entsprechenden Eurocode in den Chat hoch. Auf diese Weise kann die KI im weiteren Verlauf der Interaktion auf die PDF zugreifen. Tatsächlich werden so, mithilfe des „manuellen RAG“, die relevanten Beiwerte bei der Bemessung im Holzbaubeispiel korrekt verwendet. Im Normalfall soll RAG jedoch automatisch im Hintergrund mit angebundene Datenbanken ablaufen und dort das benötigte Wissen abrufen. Zusammenfassend lässt sich sagen: Mit RAG nutzt man externes Wissen, das nicht explizit im KI-Modell eingebettet ist (also nicht Teil des Trainings war). Gleichzeitig kann der Output mit einer Quellenangabe versehen werden, da die Herkunft der Information in der angebundene Datenbank transparent nachvollziehbar ist. Die Struktur des KI-Modells wird dabei nicht verändert, sie wird lediglich mit eigenen Daten (z. B. Firmendaten, Baudaten) verknüpft. Das kann allerdings zu einer leicht verzögerten Antwortzeit führen, da zunächst die relevanten Informationen in der Datenbank gesucht werden müssen.

Fine-Tuning für die KI

Beim sogenannten Fine-Tuning läuft das anders ab: Hier wird das bestehende KI-Modell weitertrainiert, das „Kind“ wird sozusagen in eine Weiterbildung geschickt. Durch diesen Prozess wird die Struktur des LLM verändert und das Bauwissen wird fest im Modell „verdrahtet“. In unserem einfachen Beispiel mit der Person und dem Lexikon würde das Lexikon nun entfallen, weil die Person die Frage aus dem eigenen Wissen heraus beantworten kann, sie benötigt keine externe Wissensquelle mehr. Das Nachtrainieren erfordert jedoch deutlich mehr Rechenleistung sowie passende, bauspezifische Trainingsdaten. Das KI-Modell kann damit schnell und konsistent Fragen beantworten, ohne Informationen aus externen

Dokumenten abrufen zu müssen. Dabei ist zu beachten, dass Wissen, etwa aus einer ÖNORM, durch das Fine-Tuning dauerhaft in das Modell „eingraviert“ wird. Eine spätere Änderung dieser Norm kann nur durch erneutes (und rechenintensives) Nachtrainieren berücksichtigt werden. Beim RAG-Ansatz hingegen genügt es, die angebundene Datenbank mit der neuen ÖNORM zu aktualisieren. In einer Branche, in der sich Informationen häufig ändern, ist die einfache Integration aktueller Daten von großer Bedeutung.

Fazit

Eine KI (Künstliche Intelligenz) kann also durchaus entsprechende Fragen aus dem Bauwesen beantworten. Wichtig ist dabei, dass das KI-System (wobei das KI-Modell ein Teil davon ist) dieses Wissen auch irgendwo verfügbar hat. Ein interessanter Ansatz wäre daher, dem KI-Modell durch Fine-Tuning die „Fachsprache“ und die grundlegenden Prinzipien des Bauwesens zu vermitteln (da sich diese nur selten ändern) und zusätzlich über RAG (Retrieval-Augmented Generation) eine externe Wissensquelle für Dokumente wie Normen, Richtlinien, Tabellenwerke usw. bereitzustellen. Nicht zu vergessen ist allerdings, dass wir bisher nur die technische Machbarkeit betrachten. Inwiefern die rechtlichen Aspekte, insbesondere im Hinblick auf Urheberrecht und Datenschutz, zu berücksichtigen sind, wird aktuell noch diskutiert. Das Hochladen einer Datei in einen KI-Chat (z. B. den Eurocode als PDF) kann rechtlich durchaus heikel sein. Besonders ÖNORMEN müssen käuflich erworben werden und sind daher nicht direkt öffentlich zugänglich. Es stellt sich jedoch grundsätzlich die Frage, ob gesetzlich vorgeschriebene ÖNORMEN (wie etwa der Eurocode) überhaupt hinter einer „Paywall“ stehen sollten.