

Für eine Bayern-Karte der Bebauungsplanung: Entwicklung einer Methodik zur KI-gestützten Extraktion von textlichen Informationen

Towards a Bavarian Map of Urban Development Planning: Development of a Methodology for AI-assisted Extraction of Textual Information

Felicitas Sommer | Sebastian Riznik | Rebekka Schade |
Giulia Sophie Roßberg | Flora Hirche | Hope Ewudor

Zusammenfassung

Der Artikel befasst sich mit der systematischen, Gemeinde übergreifenden Auswertung von Bebauungsplänen, die aufgrund heterogener Formate und fehlender Digitalisierung bisher erschwert ist. Ziel der im Rahmen des Projektes »GreenDIA« entwickelten Arbeit ist die KI-gestützte Extraktion textlicher Informationen aus bayerischen Bebauungsplänen zur Erzeugung strukturierter, maschinenlesbarer Daten. Die Methodik kombiniert Webscraping zur Dokumentenerfassung, optische Zeichenerkennung (OCR) zur Textextraktion und große Sprachmodelle (LLMs) zur inhaltlichen Analyse. Inhaltlich fokussiert sich die automatisierte Extraktion auf klar definierte planungsrechtliche Festsetzungen zu den Themen Klimaanpassung, Flächenverbrauch und Hochwasserschutz. Zur Qualitätssicherung und systematischen LLM-Validierung wurde ein von menschlichen Annotator:innen erstellter Goldstandard-Datensatz entwickelt. Dieser Benchmarking-Datensatz dient dazu, die Leistung verschiedener Modell- und Pipeline-Kombinationen hinsichtlich ihrer Fehlerquoten zu messen und Prompts empirisch zu optimieren. Als zentrales Ergebnis der Arbeit entstanden vier Datensätze, darunter ein Metadatensatz öffentlich zugänglicher Pläne in Bayern sowie eine stratifizierte Stichprobe nach Region, Alter und Dokumentenlänge. Die extrahierten Zielinformationen und räumlichen Daten werden anschließend auf einem interaktiven Web-Portal als durchsuchbare Bayern-Karte visualisiert. Als Limitationen der Studie werden unter anderem Herausforderungen bei der Georeferenzierung sowie die fehleranfällige Informationsextraktion bei älteren und stark verdichteten Plänen benannt. Abschließend skizziert der Beitrag konkrete Anwendungsfälle der generierten Datenbasis für das kommunale Monitoring, die Infrastrukturplanung und die empirische Raumforschung.

Schlüsselwörter: Bebauungsplanung, Digitalisierung, Sprachmodelle und KI-gestützte Extraktion, Datenqualität

Summary

This article presents a methodology for the AI-supported analysis of Bavarian development plans. Since these documents are highly heterogeneous and poorly standardized, the authors combine web scraping, optical character recognition (OCR),

and Large Language Models (LLMs) to automatically extract regulatory text on topics like climate adaptation, land take and flood protection. To ensure data quality, a human-validated gold-standard dataset was developed to benchmark the accuracy of various AI model pipelines. The result is an interactive map of Bavaria that visualizes spatial control patterns of municipalities. Despite technical limitations with older plans, this approach opens up valuable data-driven opportunities for strategic urban development, infrastructure planning, and evidence-based spatial research.

Keywords: urban planning, digitalisation, language models and AI-powered extraction, data quality

1 Einleitung

Die Digitalisierung der Planung erlaubt es, mit Ansprüchen an die Raumnutzung anders umzugehen. In der Praxis können strukturierte Daten aus Bebauungsplänen über Gemeindegrenzen hinaus analysiert werden und damit interkommunale Kooperation und Infrastrukturplanung unterstützen. Planende können auf einen Blick geeignete Standorte oder Risiken für Maßnahmen erkennen. Außerdem entstehen neue Möglichkeiten für empirische Untersuchungen in der Raumforschung: Veränderungen in der Praxis der Bebauungsplanungen verschiedener Gemeinden können über die Zeit erhoben und verglichen werden. Mit zusätzlichen Daten kann untersucht werden, wie die Planungspraxis im Wechselspiel mit sozio-demografischen, politischen oder Umweltfaktoren steht. Beispielsweise kann verglichen werden, wie sich die Bebauungsplanung von Gemeinden nach Novellierung der Wasserrahmenrichtlinie 2009 in Hochwassergefahrengebieten unterscheidet. Damit ist es möglich, aus Bebauungsplänen kommunale Steuerungsentscheidungen abzulesen und ihre sozialen, ökologischen und wirtschaftlichen Auswirkungen zu untersuchen. Kommunen können so aus einer Analyse bestehender Bebauungsplanung lernen, wie die verschiedenen Nutzungsbedarfe besser vereint werden

können, die durch Anforderungen an die Klimafolgenanpassung, die Energiewende, den Biodiversitätsschutz und veränderte Wohnungsbedarfe entstehen.

Bebauungspläne sind in die übergeordnete Flächennutzungsplanung sowie die Regionalplanung eingebettet und dienen als zentrales Instrument zur lokalen Steuerung der Nutzungen und sind rechtsverbindliche Satzungen, mit denen Städte und Gemeinden die Nutzung und Bebauung von Grundstücken regeln; Baugesetzbuch (BauGB) und Baunutzungsverordnung (BauNVO) stellen die rechtliche Grundlage für die Ausgestaltung dar. Bebauungspläne bestehen aus dem zeichnerischen Teil, den textlichen Festsetzungen (Abb. 1) und der Begründung. Die Aufstellung und der Beschluss der Bebauungspläne liegt gemäß der kommunalen Planungshoheit in der Verantwortung der Städte und Gemeinden. Bebauungspläne werden zur Entwicklung einzelner Baugebiete aufgestellt – doch gebündelt bilden sie räumliche Muster vergangener rechtsverbindlicher Regelungen eines Gebietes ab. Als Informationsgrundlage können sie über strategische Entscheidungen und Risikomanagement informieren und sind damit für Praxis und Forschung über die kommunale Ebene und eine Ausrichtung am Einzelgrundstück hinaus relevant.

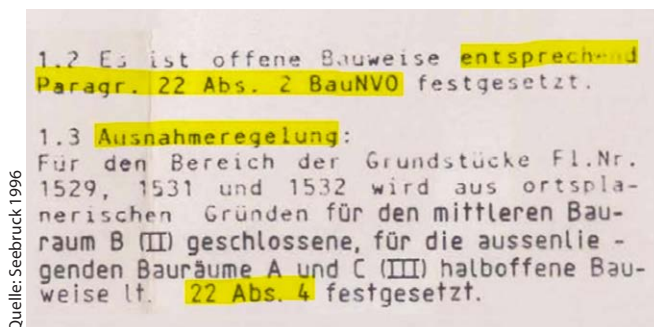


Abb. 1: Textliche Festsetzungen in einem Bebauungsplan

Es gibt jedoch einige Herausforderungen für die systematische, gemeindeübergreifende Auswertung und Verwendung dieser planungsrechtlichen Informationen. Die dezentrale Veröffentlichung und unvollständige digitale Bereitstellung von Bebauungsplänen, die variierenden Dokumentenformate, heterogene Bildqualität und variierende Symboliken und Begriffsverwendungen limitieren die maschinelle Auswertung und schaffen einen hohen Bearbeitungsaufwand. Selbst wenn die Planzeichenverordnung (PlanZV) Vorgaben für textliche oder zeichnerische Festsetzungen macht, sind Schreib- und Symbolvarianten in Bebauungsplänen weiterhin vielfältig. Durch die aufwendige Digitalisierung alter Bebauungspläne und die Standardisierung neuer Bebauungsplanung mit dem Datenstandard XPlanung entstehen in einigen Städten dennoch bereits maschinenlesbare und räumlich skalierbare Informationsangebote: Im Geoportal Hamburg sind Festsetzungen online objekt-referenziert abrufbar. Doch in ländlichen Regionen und Kleinstädten existieren die Kapazitäten für eine flächendeckende Digitalisierung häufig

nicht. Die Zentralisierung, Digitalisierung und Analyse von Bebauungsplänen muss daher teilweise mit KI-gestützten Methoden ermöglicht und beschleunigt werden (Hersperger et al. 2021). Zu KI-gestützter Extraktion aus lokalen Planungsdokumenten wurden Artikel mit Fallstudien aus Singapur (Silvennoinen et al. 2023), Polen (Kaczmarek et al. 2022) und Australien (Hosseini et al. 2025) veröffentlicht. Das Planungsbüro Syte hat ein Portal der Bebauungsplanung aufgebaut, welches nach eigenen Angaben 25 % der Fläche in Deutschland abbildet (Syte 2026). Es bindet auf seiner Plattform maschinenlesbare Bebauungspläne in Deutschland ein mit Schwerpunkt auf urbanen Räumen. In dem Projekt ARGOS des Fraunhofer Institutes für Integrierte Schaltungen wird die Potenzialflächenanalyse dank einer KI-gestützten Suche in Bebauungsplänen ermöglicht (Dürrbeck et al. 2025). Die Hochschule Luzern erarbeitet aktuell ein regulatorisches Informationssystem mit Hilfe einer KI-gestützten Datenextraktion (Kraft et al. 2023) in der Schweiz und in den USA wurde ein Tool für die Auswertung von Regionalplänen entwickelt (University of California, Davis 2024).

Für eine Nutzung der Daten in der Praxis muss jedoch abgeschätzt werden können, wie präzise Sprachmodelle die gewünschten Informationen finden und wiedergeben. Je heterogener die Dokumente sind, desto aufwendiger ist es, die »Accuracy« (auf Deutsch »Genauigkeit«, definiert als der Anteil der korrekten Extraktionen) verlässlich zu prüfen. Für jeden Aspekt der Heterogenität (Format, Bildqualität, inhaltliche Struktur und Semantik) müssen Lösungen entwickelt werden, die einer interdisziplinären Zusammenarbeit und einer engen Mensch-KI-Interaktion (Human-in-the-Loop) bedürfen. Human-in-the-Loop bezeichnet einen Ansatz, der Menschen aktiv in automatisierte KI-Prozesse einbindet, etwa zur Validierung, um Genauigkeit und Verlässlichkeit zu gewährleisten. Zuletzt muss auch die Datenverwendung darauf abgestimmt sein, welche Accuracy die KI-gestützte Extraktion für welche Dokumentengruppen erreichen kann.

Dieser Beitrag stellt eine Methodik vor, welche im Rahmen des Konsortialforschungsprojektes GreenDIA (Green Data, Indicators, Algorithms) für den B-Plan Extractor entwickelt wurde. GreenDIA ist eine Zusammenarbeit von Lehrstühlen der TU München, der LMU München und der Universität der Bundeswehr München, untersucht das Zusammenwirken von Dateninfrastruktur, Indikatoren und Algorithmen im Bereich Nachhaltigkeit und wird vom Bayerischen Institut für Digitale Transformation (bidt) gefördert. Ziel des Projektes ist die Entwicklung einer Methodik zur KI-gestützten Extraktion der textlichen Informationen aus Bebauungsplänen, primär fokussiert auf Pläne aus dem Bayernatlas. Hierfür wurden Workflows aus Webscraping (dem automatisierten Auslesen von Daten auf Webseiten), optischer Zeichenerkennung (OCR) und Extraktion durch große Sprachmodelle (LLMs) zur Erzeugung strukturierter, maschinenlesbarer Daten implementiert. Zur Qualitätssicherung wurde ein »Goldstandard«, eine manuell geprüfte Referenzanno-

tation der gesuchten Informationen, geschaffen. Dafür wurde eine Stichprobe von Bebauungsplänen erstellt, die potenzielle Fehlerquellen für die KI-Extraktion enthalten, z. B. aufgrund von heterogenen Formaten, Bildqualität, Seitenzahl und Layout-Komplexität. Aus dieser Stichprobe wurden die Zielbegriffe zunächst KI-gestützt extrahiert und die Ergebnisse von Studierenden der Raumplanung validiert und kommentiert. Im Rahmen des Projektes wurden in einer Region in Oberbayern sämtliche über den Bayernatlas zugängliche Bebauungspläne abgerufen, Informationen für klar definierte Themen (Hochwasserrisiken, Klimaanpassung und Neuversiegelung) ausgewertet und die Informationen anschließend auf einem öffentlich zugänglichen Portal visualisiert.

2 Methodik und Durchführung

2.1 Dokumente erfassen und strukturieren

In Bayern wird ein Großteil der gescannten und digitalen Bebauungspläne auf dem Bayernportal des Bayerischen Ministeriums für Digitales angeboten (Freistaat Bayern 2026). Viele Städte und Gemeinden bieten darüber hinaus auf eigenen Portalen Dokumente oder Geodatenervices an. Um Bebauungspläne automatisiert und georeferenziert abzufragen, wurden Skripte geschrieben, die neben dem Download der Metadaten auch die Koordinaten der URL auf der Karte erfassen. Für einige weitere Portale, wie beispielsweise die Stadtplanungskarte in Passau (Stadt Passau 2026), wurden zusätzliche Skripte entwickelt. Anschließend wurden sämtliche Links abgerufen und die vorhandenen Daten als PDFs oder Bilddateien heruntergeladen. Unter Einhaltung der Nutzungsbeschränkungen wurden die Koordinaten mit einem Datensatz des Bayerischen Staatsministeriums für Wirtschaft, Landesentwicklung und Energie verknüpft, welcher die Geltungsbereiche der Bebauungspläne enthält. Der Datensatz mit den identifizierten PDFs wurde für die nächsten Schritte aufbereitet.

2.2 Entwicklung der Extraktionspipeline

Anschließend wurde ein Verfahren entwickelt, bei dem die PDFs entweder direkt mit einem Vision Language Model (VLM) als Bilddateien oder mit einer Kombination aus optischer Texterkennung und LLM analysiert wurden. Mit jeweils drei verschiedenen Varianten, bei denen 1) das Input-Format (separate Seiten vs. gesamtes Dokument), 2) das LLM/VLM (Gemini 3 Flash vs. GPT-4o) und 3) das OCR-Modell (EasyOCR vs. PaddleOCR) variiert wurden, wurde untersucht, welche Applikationen und Kombinationen für welche Dokumententypen die besten Ergebnisse erzielen. Die Gesamtheit dieser Verarbeitungsschritte wird als Extraktionspipeline bezeichnet.

Anschließend wurden verschiedene Prompts, also textliche Aufgabenanweisungen an das Modell, mit den Verfahrensvarianten getestet. Die Extraktionspipeline wird in Abb. 2 dargestellt.

2.3 Konzeption und Prüfung

Die automatisierte Auswertung von Daten, ihre Nutzbarkeit und Qualität für bestimmte Anwendungsfälle sind zentrale Kriterien für die Entwicklung des strukturierten B-Plan-Datensatzes. Dieser enthält die Metadaten wie URL und Koordinaten sowie die extrahierten Zielinformationen aus Bebauungsplänen, beispielsweise zu zulässigen Bebauungsmaßen, geltenden Rechtsvorschriften und Hochwasserschutzmaßnahmen. Im Folgenden wird beschrieben, wie die Nutzbarkeit und Datenqualität definiert wurden und in die Methodik der Datensatzerstellung eingeflossen sind. Das Ziel der automatisierten Bewertung der Modellleistung ist es, die falsch negativen (übersehenen) und falsch positiven (fehlerhaften) LLM-extrahierten Werte zu erkennen und damit die Leistung verschiedener Modellkombinationen, Verarbeitungsschritte und Prompts zu messen.

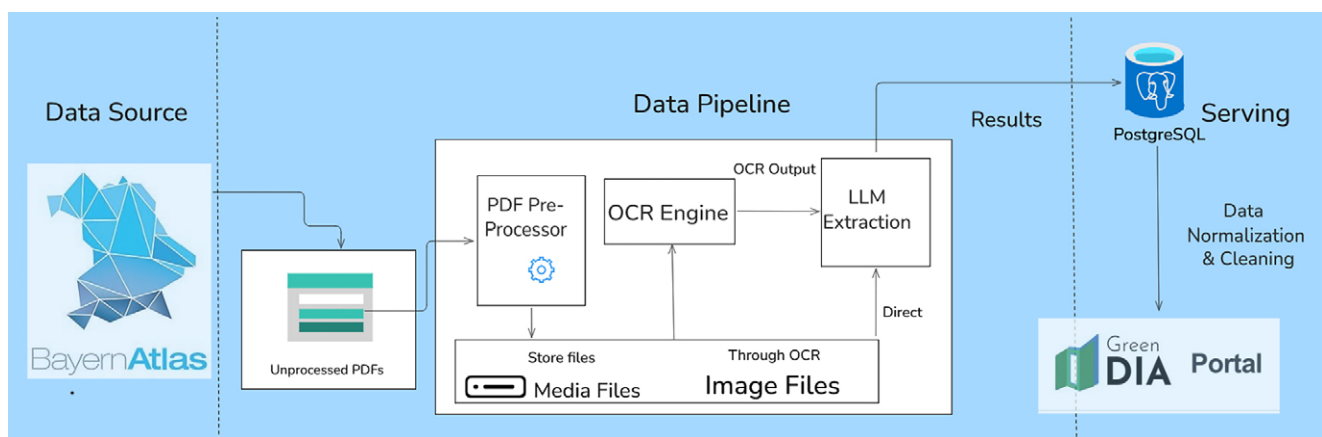


Abb. 2: Extraktionspipeline des B-Plan-Projektes

2.4 Vorgehen zur Identifikation der Schlüsselbegriffe

Eine der großen Herausforderungen bei der automatisierten Informationsgewinnung ist es, »räumliche Daten mit Menschen und Computern zu klassifizieren und zu teilen« (Barramou et al. 2020, S. 697; eigene Übersetzung). FAIR-Prinzipien (findability, accessibility, interoperability, reusability) sichern die Auffindbarkeit, Zugänglichkeit, Interoperabilität und Wiederverwendbarkeit für maschinenlesbare Daten. Der Datenstandard XPlanung, eine ausführliche Datensatzbeschreibung und die transparente Regelung von Zugangsrechten ermöglichen dies bereits ohne KI-gestützte Extraktion. Die Digitalisierung alter Bebauungspläne kann beschleunigt werden, wenn Informationen KI-gestützt extrahiert und im XPlanung-Datenstandard abgebildet werden. Dafür müssen die Objekte und rechtlichen Attribute eindeutig identifizierbar und zuzuordnen sein. Anschließend können regulatorische Informationen einheitlich extrahiert und validiert werden, in einer Datenbank gespeichert, eindeutig bestimmten Themen und räumlichen Grenzen eines Bebauungsplans in Form von Polygonen zugeordnet und in einem Geoportal visualisiert werden.

Um für aktuelle Fragestellungen relevante Planungs-begriffe in der textlichen Festsetzung (vgl. Abb. 1), teilweise in der zeichnerischen Festsetzung und der Begründung zu identifizieren, wurden Literaturrecherchen und Experteninterviews durchgeführt und Bebauungspläne gesichtet. Dabei wurden üblich verwendete Bebauungsmaße wie die Grundflächenzahl und Firsthöhe, rechtliche Grundlagen, beispielsweise Paragraphen des BauGB und der BauNVO und Schlüsselbegriffe wie Versickerung und Hochwasser identifiziert.

Die Liste der Schlüsselbegriffe für die automatisierte Extraktion wurde anschließend reduziert, da die menschliche Validierung pro Begriff aufwendig und kostspielig ist. Dafür war ausschlaggebend, wie häufig und einheitlich Begriffe vorkommen und wie viele Annotationsstunden pro Begriff für die Validierung nötig sind. Ausgewählt wurden häufige und möglichst klar definierte, identifizierbare und abgrenzbare Begriffe mit geringen sprachlichen Abweichungen. Begriffe mit variierenden Formulierungen und inhaltlicher Ambiguität wurden nicht nur ausgeschlossen, um eine robustere Extraktion durch die Sprachmodelle zu erreichen. Einheitliche Schreibweisen, strikte Ja/Nein-Klassifikationen oder numerische Werte verbessern auch die Qualität der menschlichen Validierung, erhöhen die Übereinstimmung zwischen Annotator:innen und erleichtern den Vergleich zwischen Annotations- und KI-Extraktionsergebnissen. Ausgewählte Zielbegriffe sind beispielsweise BauGB/BauNVO, GRZ, GFZ, Firsthöhe, Überschwemmungsgebiet und die jeweiligen numerischen Angaben. Aktuell nicht vollständig validiert sind Hochwasseranpassungsmaßnahmen.

2.5 Entwicklung der Umfrage zur Validierung der KI-gestützten Extraktion

Menschliche Annotation bezeichnet das Kodieren von Texten oder Bildern, um mit den klassifizierten Daten Algorithmen oder KI-Befehle zu verbessern. Entweder werden Trainingsdaten direkt von Menschen annotiert oder zunächst KI-gestützt extrahiert und dann menschlich validiert. Der Ansatz der LLM-Validierung eignet sich, da Annotator:innen die KI-Fehler kommentieren und damit zur Promptoptimierung beitragen können. In jedem Fall muss eine Annotationsrichtlinie entwickelt werden. Im zweiten Fall muss auch eine Methode für die LLM-Validierung entwickelt werden. Im Folgenden wird die gewählte Annotationsmethode begründet und erläutert, wie das Design und die Richtlinien des Annotationsprozesses für die LLM-Validierung entwickelt und überarbeitet wurden, um einen »Goldstandard-Datensatz« für die LLM-Validierung zu generieren.

Die Annotationsumfrage besteht aus mehreren Teilen: (1) der Einführung in die Struktur des Bebauungsplans und den von den Annotator:innen zu verwendenden Definitionen und Labels; (2) dem zu annotierenden Bebauungsplan; (3) einer Abfrage zur Vollständigkeit der LLM-extrahierten Begriffe mit Möglichkeit zur Ergänzung nicht-gefundener Begriffe; (4) einer Prüfung der Genauigkeit und der jeweiligen Fundstelle sowie (5) einem Kommentar zum Fehler.

Mit dem Online-Umfrageprogramm Limesurvey wird das Umfragemodul für die Annotation automatisiert für jeden Bebauungsplan repliziert. Dabei werden die Originalpläne mit den entsprechenden LLM-Extraktionen präsentiert. Zur Erhöhung der Übereinstimmung zwischen den Annotator:innen wird nach einer Testphase eine Annotationsrichtlinie erstellt, nach welcher die LLM-Extraktionen einheitlich zu bewerten und zu korrigieren sind, wodurch der kontinuierliche Austausch zwischen den Annotator:innen sichergestellt wird. Der resultierende Metadatensatz enthält die extrahierten Werte, ob diese korrekt sind (Ja/Nein) und gegebenenfalls den korrigierten Wert. Zusätzlich enthält der Datensatz die Dokumentation von Fehlerquellen, den »Fundort«, um Schwachstellen im Prompt oder im Modell zu erkennen, und einen Kommentar zur Einordnung des Fehlertyps. Zu den Fehlertypen gehören Text-Interpretationen, Text-Zusammenfassungen, Layout-Interpretationen, Verwechslungen von Zielinformationen, beispielsweise zwischen GFZ- und GRZ-Werten oder Symbolen, und Unvollständigkeit. Die Annotator:innen beschreiben zusätzlich, was den LLM-Fehler hervorgerufen haben könnte, wie beispielsweise komplizierte Grafiken oder geringe Bildqualität. Diese Informationen ermöglichen es, Pläne zu klassifizieren, die potenziell fehleranfällig sind, und die Prompts für diese Gruppen gezielt zu optimieren. Im Folgenden werden die Erstellung des Benchmarkingdatensatzes und die Bewertung der Modulleistung vorgestellt.

2.6 Stichprobenziehung

Zur automatisierten Bewertung der Modellleistung muss berücksichtigt werden, dass die Dokumentenformate, Textinhalte und die Bildqualität in Bebauungsplänen sehr heterogen sind. Um einen Benchmarking-Datensatz zu gewinnen, muss daher aus der Gesamtheit der Pläne eine ausgeglichene und ausreichend große Stichprobe gezogen werden, es müssen ausgewählte Zielinformationen extrahiert und durch fachkundige Annotator:innen bewertet und korrigiert werden. Anschließend wird durch den Vergleich der Datensätze berechnet, wie häufig Begriffe nicht gefunden wurden und wie häufig Begriffe erfunden wurden. Die übersehenen Instanzen bezeichnen den Recall, die erfundenen oder falschen Begriffe die Precision. Auf Basis der Fehlerhäufigkeit bestimmter Dokumente im menschlich annotierten Evaluierungsdatensatz werden Dokumentenklassen für die finale Stichprobe definiert. Die Stichprobe ist in der Tabelle weiter spezifiziert (Tab. 1). Die Merkmale Alter und Dokumentenlänge bilden nun die Klassen in dem finalen Bebauungsplan-Datensatz. Die Bildqualität der Pläne nimmt mit höherem Alter ab und die korrekte Informationsextraktion wird erschwert. Darüber hinaus ändern sich Begrifflichkeiten über die Zeit durch Gesetzesreformen, beispielsweise wurde BBauG zu BauGB. Die Seitenzahl der Dokumente ist relevant, weil bei langen Dokumenten die Gefahr höher ist, dass Begriffe nicht gefunden werden, während einseitige Pläne aufgrund ihrer Layout-Komplexität und Unübersichtlichkeit zu fabrizierten Informationen, sogenannten Halluzinationen, führen können.

2.7 Pipeline und LLM-Validierung

Für die Bewertung der Modellleistung sind Genauigkeit und Auffindungsrate sinnvolle Evaluationsmetriken bei binärer Klassifikation (Ja/Nein) (Naidu et al. 2023). Die gängige Methode ist die Berechnung eines sogenannten F1-Scores als harmonisches Mittel aus Genauigkeit (Precision) und Auffindungsrate (Recall). Mit dem menschlich annotierten Benchmarking-Datensatz der ursprünglichen Stichprobe wurden acht Pipeline-Modell-Kombinationen aus den Modellen (Gemini und OpenAI) und zwei Applikationen der Optischen Zeichenerkennung (PaddleOCR und DeepSeek-OCR) für jeweils ganze Dokumente oder Einzelseiten getestet.

Mit String Matching wurde die Übereinstimmung zwischen menschlichen Annotationen und KI-Extraktion geprüft, um die Qualität der einzelnen Extraktionen zu bewerten und die Pipeline-Variationen darin zu vergleichen. Abschließend wurde untersucht, wie die Accuracy der Pipelines abhängig von der Seitenzahl und dem Erstellungsjahr der Pläne variiert. Die Kombinationen mit den geringsten Fehlern wurden ausgewählt und Dokumente identifiziert, die in allen Kombinationen schlecht abschneiden. Leicht unterscheidbare Merkmale dieser Pläne wurden neben den Kommentaren aus der Annotation für die Promptoptimierung herangezogen. Das Alter der Pläne ist ein Indiz für die Bildqualität, die Seitenzahl für die Komplexität des Plans, da einseitige Pläne wesentlich größer und komplizierter strukturiert sind als mehrseitige A4-Dokumente. Eine Fehleranalyse über alle Pipeline-Variationen hinweg bestätigte die beobachteten Schwierigkeiten bei der Extraktion von einseitigen Plänen. Der niedrigste Recall zeigte sich bei kurzen Plänen von ein bis zwei Seiten, da deren Layout-Komplexität aufgrund der Informationsdichte

Tab.1: Anzahl Bebauungspläne und Aufschlüsselung nach Alter in der Stichprobe

Phase	Periode	Anzahl	Anteil	Bedingungen
BBauG Ära (≤1986)	≤1986	8.376	22,8 %	BBauG Vokabular; geringe OCR Qualität · PDF verfügbar: 5.247 (14,3% of 36.811) · kein PDF: 3.129 (8,5%)
BauGB (1987–2009)	1987–2009	14.909	40,5 %	Gemischt BBauG/BauGB; frühe digitale Scans; Layoutvariationen · PDF verfügbar: 9.558 (26,0% of 36.811) · kein PDF: 5.351 (14,5%)
BauGB (2010+)	2010+	8.049	21,9 %	Standardisiertes Vokabular; vorwiegend maschinenlesbare Pläne · PDF verfügbar: 5.258 (14,3% of 36.811) · kein PDF: 2.791 (7,6%)
Pool ohne Datum	NULL	5.477	14,9 %	Kein Datum im Metadatenatz; daher vom Evaluationskorpus ausgeschlossen · PDF verfügbar: 3.696 (10,0% of 36.811) · kein PDF: 1.781 (4,8%)

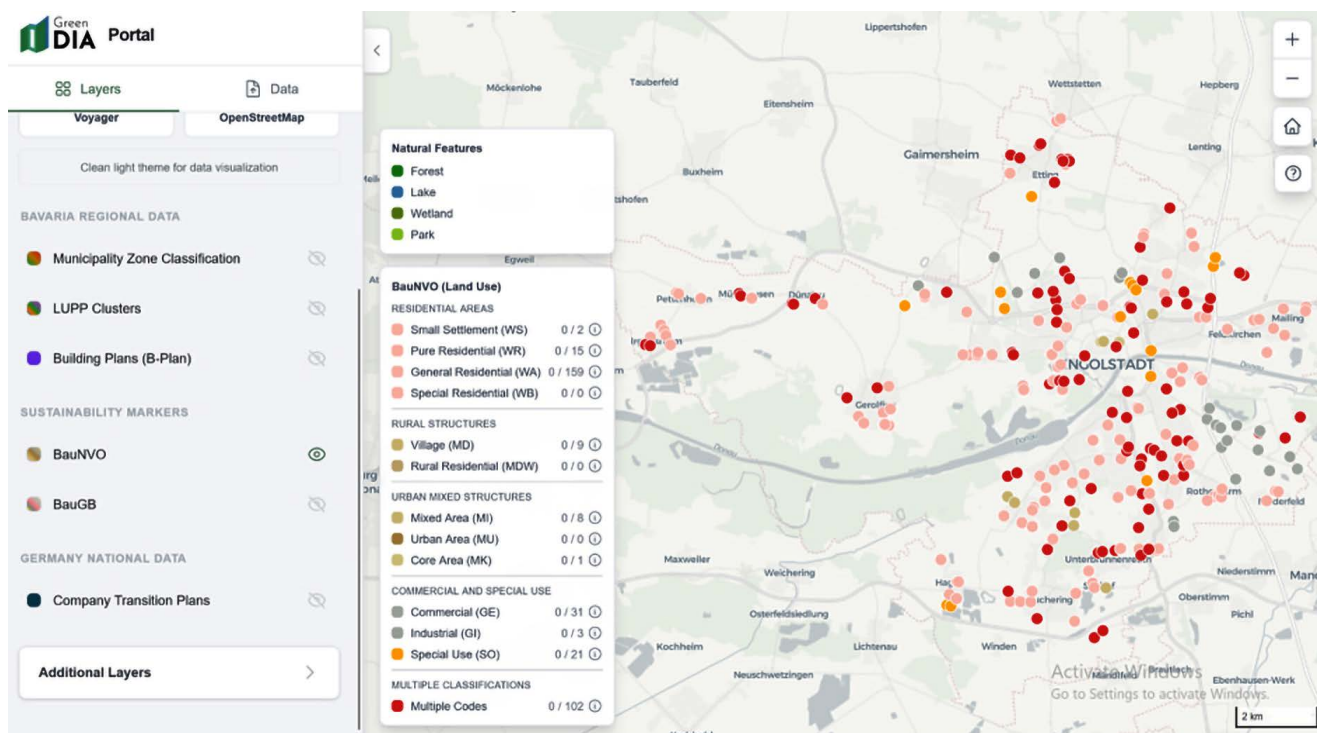


Abb. 3: Screenshot des B-Plan-Portals von GreenDIA

und durch Kartendarstellungen hoch ist, während wir die niedrigste Precision bei langen Plänen mit mehr als neun Seiten fanden. Für den ursprünglichen Prompt zur Extraktion von BauNVO-Paragrafen erreichte eine LLM-Validierung auf Basis von einem Benchmarkingdatensatz mit 80 Bebauungsplänen für Gemini-noOCR einen F1-Score von 0,5774 (Precision 0,4929, Recall: 0,6970). Der aktuelle Prompt erreicht einen Gesamtscore (F1) von 0,8895 (Precision: 0,861, Recall: 0,920). Somit wurden falsch negative und falsch positive Fehler sehr stark reduziert.

3 Datensätze

Es wurden vier zentrale Datensätze im Projekt erstellt: ein Metadatenatz der öffentlich zugänglichen Bebauungspläne in Bayern, eine Stichprobe der Bebauungspläne und ein Benchmarkingdatensatz auf Grundlage von LLM-Extraktion, menschlicher Validierung, erneuter LLM-Extraktion und einem finalen Vergleich zwischen LLM-Extraktion und menschlicher Annotation. Im Folgenden werden die vier Datensätze näher spezifiziert.

1. Der gesamte Metadatenatz enthält Titel, Planart (Qualifizierter oder Einfacher Bebauungsplan), Rechtsstand, Gemeinde, Georeferenz (Punktdaten), Download-URL, Datum des Inkrafttretens. Hinzugefügt wurden für jeden Bebauungsplan die Gemeindekennziffer und ob eine Gefahrenlage Hochwasser/Starkregen existiert.
2. Die Stichprobe für das Prompt-Experiment-Design (PAT) nach Region, Alter und Seitenlänge. Dieses enthält zusätzlich den Zeitpunkt des Inkrafttretens und die Seitenzahl (siehe Tab. 1).

3. Der dritte Datensatz ergänzt die menschlich validierten Zielinformationen: Die Art der baulichen Nutzung gemäß §§ 1–11 BauNVO, die Festsetzungen zur Geschossflächenzahl und zur Grundflächenzahl gemäß § 9 BauGB und die Firsthöhe, HQhäufig, HQ100 und HQextrem, Aussagen mit den Themen »Versickerung« und »Bauweise«.
4. Ein weiterer Datensatz, der ausschließlich Bebauungspläne in Hochwassergefahrenlagen enthält, wurde mit Polygondaten der Bebauungspläne verknüpft und ist nur für Forschungszwecke nutzbar.

Die Ergebnisse der Extraktion werden auf einem Portal veröffentlicht, wie unter Abb. 3 sichtbar. Hier werden die extrahierten Informationen visualisiert. Die Quellen der Zielinformationen im Text sowie die Original-URL können dort für die Bebauungspläne eingesehen werden.

4 Limitationen und Erkenntnisse

In diesem Artikel wurde eine Methodik besprochen, heterogene und lückenhafte Bestände von Bebauungsplänen zu sammeln und rechtliche Informationen daraus zu kartieren. Die Methodik ist jedoch nicht geeignet, rechtssichere Aussagen über einzelne Festsetzungen eines großen Gebietes zu machen, da die Genauigkeit der KI-Modelle nicht 100 % erreicht. Die technischen Herausforderungen für das Projekt ergeben sich besonders aus der Heterogenität der Bebauungspläne, ihrer Bildqualität, Struktur und Begrifflichkeiten. Limitationen bestehen besonders aufgrund folgender Aspekte: (1) Vollständigkeit: Die online verfügbaren

baren Bebauungspläne pro Gemeinde im Bayernatlas variieren stark. Nicht für alle Gemeinden waren Pläne auf dem Bayernportal zu finden. Einige Kommunen haben eigenständige Bebauungsplanportale, die für eine Anreicherung des Datensatzes genutzt wurden, wie beispielsweise Passau oder Donauwörth. (2) Georeferenzierbarkeit: Die Punktdaten sind nicht immer akkurat und müssen über einen Abgleich der Plannamen mit Polygondaten gematcht werden. Informationen können nicht direkt den verschiedenen Objekten, sondern lediglich dem gesamten Bebauungsplan zugeordnet werden. (3) Genauigkeit: insbesondere ältere und einseitige Pläne müssen aufwendig gesplittet und mit hoher Auflösung verarbeitet werden. Die dafür nötigen Rechenkapazitäten sind im Projekt nicht vorhanden.

5 Ausblick

Für die Daten gibt es viele potenzielle Anwendungsfälle, für die jedoch berücksichtigt werden muss, dass eine KI-Extraktion nicht vollständig fehlerfrei durchgeführt werden kann. Es ist lediglich machbar, die Unsicherheiten möglichst zu reduzieren und große Schwankungen in der Precision zu verringern, sodass Gemeinden miteinander verglichen und Muster erkannt werden können. Für die Wirkungsmessung der Bebauungsplanung und zur Analyse ihrer Übereinstimmung mit Zielen und Prinzipien der Regionalplanung und Raumordnung sowie von Stellungnahmen von Trägern öffentlicher Belange über die Zeit ist der Datensatz gut geeignet. Auch für die Erstellung von Strategien und evidenzbasierten Politikentscheidungen, die interkommunale Kooperation und die Infrastrukturplanung besteht eine geeignete Grundlage, um Risiken und Potenziale im Frühstadium zu erkennen. Eine manuelle Prüfung kann mit weiteren Schritten erleichtert, jedoch nicht gänzlich ersetzt werden. Auch für die Bestandsaufnahme und Beratung von Kommunen hinsichtlich ihrer Siedlungsentwicklung können die Datensätze genutzt werden. Für die Forschung ergeben sich zusätzlich neue Möglichkeiten, die Bedeutung sowohl der Bebauungsplanung als auch der Informationstechnologien für die Entwicklung von Siedlungsstrukturen erstmals in Deutschland empirisch zu erforschen und daraus Empfehlungen für evidenzbasierte politische Entscheidungen zu geben.

Unterstützung und Finanzierungsquellen

Dieser Artikel wurde im Rahmen des Projektes GreenDIA erstellt, welches vom Bayerischen Institut für Digitale Transformation (bidt) gefördert wurde. Es wurde außerdem unterstützt durch das Bayerische Staatsministerium für Digitales und das BERD@NFDI-Konsortium im Rahmen der Arbeit der Nationalen Forschungsdateninfrastruktur (NFDI) Association. Die NFDI wird von der Bundesrepublik Deutschland und den 16 Bundesländern finanziert.

Das BERD@NFDI-Konsortium wird innerhalb der NFDI von der Deutschen Forschungsgemeinschaft (DFG) unterstützt – NFDI 27/1-2026, Projektnummer 460037581.

Literatur

- Dürrbeck, K., Fuchs, S., Distel, L. (2025): Deutschland und seine Bebauungspläne Monitoring, Analyse und generative Künstliche Intelligenz. In: Meinel, G., G. Behnisch, G., Leibniz-Institut für Ökologische Raumentwicklung (Hrsg.): Flächennutzungsmonitoring XVII: Flächenpolitik – Bodenversiegelung – Prognosen. Berlin: Rhombos-Verlag. IÖR Schriften, Band 83, 387–402.
- Hersperger, A., et al. (2021): Digitalization in land-use planning: effects of digital plan data on efficiency, transparency and innovation. *European Planning Studies*, 30, 1–17. DOI: 10.1080/09654313.2021.2016640.
- Hosseini, H., Atazadeh, B., Rajabifard, A. (2025): Artificial Intelligence for Querying Land and Property Data from Cadastral Plans.
- Kaczmarek, I., et al. (2022): A machine learning approach for integration of spatial development plans based on natural language processing. *Sustainable Cities and Society*, 76, 103479. DOI: 10.1016/j.scs.2021.103479.
- Silvennoinen, H., et al. (2023): A semantic web approach to land use regulations in urban planning: The OntoZoning ontology of zones, land uses and programmes for Singapore. *Journal of Urban Management*, 12(2), 151–167. DOI: 10.1016/j.jum.2023.02.002.

Onlinequellen und Geodaten:

- Freistaat Bayern (2026): BayernPortal – Für Bürger. Staatsregierung Bayern. <https://www.bayernportal.de>, letzter Zugriff 12.6.2026.
- Kraft, C., Steffen, D., Staub, P., Cremosnik, M. (2023, 31. August): Regulatorisches Informationssystem: Neues Forschungsprojekt zur Navigation im Dschungel der Baugesetze. Hochschule Luzern – Immobilienblog. <https://hub.hslu.ch/immobilienblog/2023/08/31/verdictungspotenzial>, letzter Zugriff 12.6.2026.
- Stadt Passau (2026): Interaktive Bauleitplan-Karte. Stadtplanung Passau. <https://www.o-sp.de/passau/karte>, letzter Zugriff 12.6.2026.
- University of California, Davis (2024): California General Plan Database Mapping Tool. Center for Regional Change. <https://plansearch.caes.ucdavis.edu>.
- Syte (2026): <https://www.syte.ms/news/neu-bei-syte-informationen-zu-bebauungsplanen-in-der-plattform>, letzter Zugriff 12.6.2026.

Kontakt

Dr. Felicitas Sommer
Lehrstuhl für Landentwicklung und Bodenordnung, Technische Universität München
Arcisstraße 21, 80339 München
felicitas.sommer@tum.de

Dieser Beitrag ist auch digital verfügbar unter <https://geodaesie.info>.