Land Use Policy 164 (2026) 107960

Contents lists available at ScienceDirect

Land Use Policy

journal homepage: www.elsevier.com/locate/landusepol

Assessing ChatGPT's legal reasoning in statutory land consolidation: The case of Cyprus

Demetris Demetriou

Department of Civil Engineering and Geomatics, Technical University of Cyprus, Limassol, Cyprus

ARTICLE INFO

Keywords:
Land consolidation
Land policy
Legal reasoning
Artificial intelligence (AI)
Large language models (LLMs)
ChatGPT
Rural development

ABSTRACT

Land consolidation remains a cornerstone of rural development, but its implementation is often hindered by complex statutory frameworks and lengthy procedures. In Cyprus, where land consolidation has historically reduced fragmentation, the process faces new challenges amid declining agricultural importance and evolving policy priorities such as climate change resilience, sustainable development, and urban land readjustment. At the same time, artificial intelligence (AI) and large language models (LLMs) like ChatGPT are increasingly being considered as tools to support legal and planning processes. This study provides the first systematic evaluation of ChatGPT's capacity to interpret and respond to legal questions derived from Cyprus's Land Consolidation Law. Using a corpus of 100 questions across four levels of legal complexity, responses were assessed with a rubric measuring correctness, completeness, clarity, and interpretive depth. The results show strong performance in basic factual and procedural questions (98% accuracy for Type 1), moderate reliability in procedural and hypothetical reasoning (84% and 82% respectively), but significant decline in complex interpretive tasks (55% for Type 4). These findings highlight both the potential and the limitations of LLMs in statutory interpretation: they can provide accessible explanations and procedural guidance but cannot yet replace expert legal reasoning in ambiguous or high-stakes cases. The study contributes to the emerging discourse on AI in land policy and rural development, offering methodological insights for evaluating LLMs in domain-specific legal contexts and outlining implications for their responsible integration into planning, cadastral, and governance workflows.

1. Introduction

Rural development programs include land consolidation (Jiang et al., 2022; Chen et al., 2024) as their fundamental component for addressing agricultural production and sustainable land management challenges caused by land fragmentation (Ntihinyurwa, de Vries, 2020). The agricultural landscape is fragmented because single holdings possess numerous scattered parcels which vary in size and shape and exist in distant locations across multiple ownership domains (Rao, 2019; Zheng et al., 2023). Current conditions create inefficient land use, raise production expenses, and make investments less attractive.

Land consolidation projects have reduced fragmentation in Cyprus since the 1970s by focusing primarily on agricultural production. It seems that land consolidation in Cyprus has reached an end, since its aim of agricultural development has weakened due to the continuous decline of the agricultural sector from 17% of Gross Domestic Product (GDP) in the 1960s to around 10% in 1980, 6% in 1992, 2% in 2010, and only 1.2% in 2025. Therefore, there is an economic question as to whether it is worth investing in this sector. Thus, there is a need to shift land consolidation in new and future directions, such as the UN Agenda 2030 Goals, climate change, resilience, leveraging innovative technologies, developing new business models, and integrating with smart agriculture (Demetiou, 2025). It is worth mentioning that in July 2025, the Parliament of the Republic of Cyprus voted for legislation to apply urban land consolidation (land readjustment) (Mahto, 2018).

Suitable legal frameworks are necessary to implement land consolidation with transparency and equity (Van, 2024). However, land consolidation projects in Cyprus and other countries (Nestorović et al. 2023) faced significant delays because of complex procedures, rigid legal frameworks, and prolonged planning and negotiation times with landowners. The average duration of projects in Cyprus extends from six to ten years or even more. The law includes multiple interconnected provisions which planners must interpret and apply precisely at each phase of land reallocation, starting from the identification of landowners for the study area, the voting of landowners, and determining an optimum land reallocation, that is, acceptable parcel shapes and rural road planning.

Investigating emerging technologies which support legal and

E-mail address: demetris.s.demetriou@cut.ac.cy.

https://doi.org/10.1016/j.landusepol.2026.107960
Received 19 August 2025; Received in revised form 16 January 2026; Accepted 27 January 2026

procedural understanding is essential because they offer the potential to enhance transparency, reduce errors in interpretation, and improve access to complex legal information for both professionals and the public. EU acted in this direction through the AI Act (Regulation [EU] 2024/1689, 2024). Modern artificial intelligence through its large language models (LLMs) including ChatGPT offers new possibilities to improve legal (Wang et al., 2023; Wrzesniowska, 2023; Kang et al., 2023), and geospatial and planning decision-making processes (Hwang et al., 2024; Mooney et al., 2023; Xu et al., 2025). These models show strong potential to handle natural language understanding duties, including summary creation, question responding, and argument development. The reliability of these models remains uncertain when used for critical legal interpretation tasks or legislative compliance applications (Fahrani and Djajaputra, 2024).

Research on Large Language Models (LLMs) has expanded, but there is a lack of comprehensive evaluations of their performance in answering complex legal questions within specific jurisdictions (Shui et al., 2023; Fei et al., 2023). The accuracy of ChatGPT in legal reasoning has been studied in previous research (Prakken, 2024; Weichbroth, 2025); however, there is a complete absence of evaluations of its ability to understand national legislative systems through land consolidation. The absence of such research is crucial because AI systems are becoming increasingly prevalent in public sector operations and workflow planning (Zhou, 2023).

This study evaluated ChatGPT's ability to answer land consolidation legislation questions through a methodical evaluation process. The evaluation system avoids using accuracy as its final goal because it focuses on demonstrating how the system performs at different levels of legal reasoning complexity. In particular, this research used 100 questions derived from Cyprus's Land Consolidation Law together with related legal documents as the basis for its analysis. The questions span from basic factual recall to abstract legal interpretation and demand synthesis between multiple legal provisions (legislation articles, administrative circulars, and legal opinions) across four difficulty levels (explained in Section 3.1). The ChatGPT answers received a five-point grading rubric (from 0 to 1.0 in increments of 0.25) which assessed the response quality by measuring correctness, clarity, completeness, and legal reasoning.

The evaluation results showed an average accuracy rate of 84%, which decreased when moving to more complex questions. The evaluation showed that ChatGPT achieved 98% in the simplest question category (Type 1), whereas it scored only 59% in the most complex question category (Type 4). The observed performance trend matches previous LLM evaluations which show diminished accuracy and confidence in legal outputs when dealing with complex ambiguous legal questions (Tan et al., 2023; Hosseini et al., 2023).

This study investigates two essential aspects of ChatGPT: its ability to deliver reliable information on land consolidation legislation and its performance differences across different legal complexity levels. We examined the performance of generative AI models when used in real-world legal systems to answer questions about their capabilities and limitations, thus supporting ongoing research on AI's responsible implementation of AI in legal-geospatial workflows.

Accordingly, this study is guided by the following research questions:

- How does ChatGPT's performance vary across increasing levels of statutory reasoning complexity in the context of land consolidation laws?
- What types of reasoning errors and limitations emerge when a general-purpose large language model is applied to jurisdiction-specific codified legal frameworks? and
- What are the implications of these findings for the safe and responsible integration of language models into legal–geospatial and land administration workflows?

These research questions are examined through a focused, domain-specific evaluation of large language model performance within statutory land consolidation. While the findings have implications for the broader evaluation of large language models in legal contexts, the primary contribution of this study lies in its assessment of LLM performance in this legally and institutionally specific land management domain.This paper is structured as follows: Section 2 examines the current research on AI applications within legal frameworks and land consolidation processes. The research design in Section 3 explains the methodological approach for developing questions, determining their difficulty levels, and assessing their answers. Section 4 presents the analysis results using comprehensive performance metrics. Section 5 discusses the implications of these findings. Section 6 examines the implications and broader consequences for the planning systems. The paper concludes with Section 7 which reflects on the contributions of this study and outlines potential directions for future investigation.

2. Literature review

AI technology, together with large language models (LLMs), including ChatGPT, has triggered extensive discussions about their capability to handle legal interpretation (Fei et al., 2023; Shui et al., 2023) and spatial planning functions (Rane et al., 2024; Li et al., 2024). The legal sector has witnessed AI transformation from rule-based expert systems to probabilistic and neural-network-based models which interpret and generate legal texts while providing summaries (Chahal, 2024; Nithya et al., 2024). The most significant transformation in generative transformer models occurred with ChatGPT which responds to elaborate natural language prompts. The system was originally created for standard tasks but has gained significant adoption within legal environments. Wrzesniowska (2023) assessed ChatGPT's performance by comparing its responses to those of law students and practising lawyers when answering Dutch legal case questions. The research revealed that ChatGPT succeeded in answering elementary legal cases but demonstrated weaknesses in grasping complex legal meanings and jurisprudential reasoning. According to Ryan and Hardie (2024), the tool serves two purposes: as an efficiency tool and a possible threat to obtaining correct legal counsel in educational and clinical settings.

The fundamental issue remains unresolved because LLMs lack reliable navigation of context-sensitive legal interpretation alongside precedent references and jurisdiction-specific details from their training data based on unverified Internet corpora. According to Spaić and Jovanović (2024), ChatGPT delivers syntactically correct responses but fails to produce the precise logical structure needed for legal clarity, especially within civil law systems which follow codified legislation without case law support found in common law traditions.

The implementation of AI systems for geospatial analysis (Rana and Bhambri, 2024; Darwish, 2024), land consolidation planning support (Demetriou, 2013), and land administration has become a growing interest alongside legal AI development. In addition to the EU's AI Act (Regulation [EU] 2024/1689), the INSPIRE Directive (Directive 2007/2/EC; European Parliament and Council of the European Union, 2007) contributes to this broader effort by promoting harmonised spatial data infrastructures across the EU. Demetriou (2013) and subsequent studies showed how spatial data merges with decision-making models in land consolidation processes through tools like LACONISS and optimisation algorithms such as genetic algorithms (Demetriou et al., 2012). The primary objective of these tools is to perform technical land parcel reallocation which promotes spatial efficiency and fairness.

The wider geospatial planning community acknowledges that AI contributes value to the extraction of data while enabling simulation activities and optimisation capabilities. Aidaoui et al. (2024) examined the integration of GeoAI (Richter and Scheider, 2023) technologies in urban land management, emphasising predictive modelling and data analysis for monitoring landscape changes. Such applications work within structured data frameworks but do not directly address legal interpretation or text analysis. The application of LLMs in this field

remains in its early stages, facing untested challenges regarding cadastral law and land tenure systems which require the joint examination of legislative texts and geospatial reasoning (Padiu et al., 2024; Zeng et al., 2024).

The demand for thorough assessments of AI-generated legal outputs has become more prominent in recently increased. The standard evaluation process uses rubrics which examine accuracy and fullness, as well as interpretive quality through numeric or ordinal rating systems. Tan et al. (2023) used a binary correctness scale to evaluate ChatGPT's bar exam-type question responses, while human experts validated the results. Luo et al. (2023) evaluated ChatGPT's performance in detecting factual inconsistency in text summarization tasks. Their study assessed ChatGPT in a zero-shot setting using tasks such as binary entailment inference, summary ranking, and consistency rating. While ChatGPT outperformed previous metrics in general, the authors noted limitations, including lexical bias, false reasoning, and challenges in instruction comprehension.

To date, no study has systematically assessed the legal provisions of land consolidation in any country or region using evaluation frameworks, despite the domain's inherent complexity, which combines statutory language, spatial planning norms, rural development strategies, and cadastral databases. The evaluation approach in this study (Section 3) expands existing rubric-based assessment methods by applying them to questions extracted directly from national land consolidation legislation to contribute both methodological and substantive value to the literature.

Historically, AI applications for land consolidation have concentrated on spatial optimisation, automated parcel design, and integrated system support. Demetriou et al. (2012), Demetriou (2013) established spatial models that integrated legislative constraints (e.g. road access, ownership, and land value) into GIS-based systems. Land reallocation research shows that it exceeds spatial analysis because it combines spatial needs with legal regulations that define proper arrangements and compensation protocols. There are also other relevant studies dealt with agricultural land planning (Peng, 2024).

The study of land tenure law and cadastre reveals that standardised legal interpretations become more challenging because of jurisdictional differences (Alterman, 2010). European land law systems, including Cyprus, use mixed legal foundations to enable semi-administrative agencies for land law implementation. Standard AI models face an increased risk of overgeneralisation when they extract rules from foreign jurisdictions or follow prevalent case reasoning in their training data.

The progress of AI research remains fast-paced; however, there is a significant lack of empirical evidence regarding how LLMs function in the land law and rural legislation domains. Most benchmark datasets for legal AI stem from commercial and criminal or constitutional law sources from English-speaking nations (López Espejel et al., 2023). The current state of knowledge regarding model capabilities for understanding legislation in less common domains and jurisdictions, such as land consolidation in Cyprus, remains unknown.

This study fills the existing knowledge gap by examining a specialised legal field that produces significant procedural effects. This research contributes to the understanding of LLM reliability in understudied legal subfields through a structured evaluation of ChatGPT responses. The difference in how well LLMs perform on easy versus hard questions helps researchers understand the limits of their reasoning, especially when moving from simple fact-finding to more complex legal interpretations.

3. Methodology

This section explains the method used to evaluate the ChatGPT responses regarding land consolidation legal issues. This section includes (i) a detailed description of the question development process and categorisation system, (ii) an explanation of the assessment criteria and scoring system, (iii) a description of the assessment procedure, and the AI model configuration and testing environment. The evaluation framework aimed to establish methodological clarity and repeatability using legal domain-specific parameters.

3.1. Development of the question corpus

3.1.1. Dataset construction protocol

The construction of the 100-question dataset followed a structured and transparent protocol designed to ensure legal coherence, procedural coverage, and alignment with practitioner-level reasoning requirements. Rather than relying on random sampling, the dataset was developed using a coverage-based approach that reflects the practical application of statutory land consolidation law. The document universe from which questions were derived consisted of the consolidated Cyprus Land Consolidation Law (1969–2018), together with its subsequent amendments, official administrative circulars, procedural guidelines, and explanatory documents issued by the former Land Consolidation Department and its successor unit within the Land and Surveys Department. These sources collectively define the legally binding framework governing land consolidation procedures, institutional responsibilities, and decision-making constraints in the study area.

Legal provisions were selected to ensure coverage across the full lifecycle of a land consolidation project, including initiation and decision-making mechanisms, institutional composition and powers, land valuation principles, design and redistribution rules, appeal procedures, and registration outcomes. In addition, provisions were selected to reflect the responsibilities of different institutional actors involved in the process, such as the Land Consolidation Committee, Valuation Committee, Appeals Committee, and competent administrative authorities. This approach ensures that the question corpus captures the functional scope of the legislation rather than concentrating on isolated provisions.

Each question was explicitly linked to at least one legal provision (article, circular, or guideline), which served as an authoritative basis for defining the expected answer. Questions requiring synthesis across multiple provisions were deliberately included to reflect real-world legal practice, where statutory interpretation frequently involves cross-referencing articles, procedural rules, and exceptions.

Question complexity was operationalised using a four-tier typology based on the cognitive and legal reasoning demands required to answer each question. Type 1 questions require the direct recall of definitions, actors, or procedural thresholds derived from single provisions. Type 2 questions involve sequential or conditional reasoning across procedural steps and may include exception clauses. Type 3 questions require the integration of multiple legal provisions or instruments to resolve hypothetical scenarios. Type 4 questions require abstract interpretation, ambiguity resolution, or the application of exceptional cases involving institutional discretion. Two domain experts independently validated that each question was unambiguous, legally coherent, and matched its intended complexity level, ensuring alignment between typology and practitioner-level difficulty.

For each question, the legal provisions forming the authoritative basis for the expected answer were documented during dataset construction, ensuring that all items were directly traceable to specific articles, circulars, or procedural guidelines. Taken together, the question construction protocol was designed to ensure that the dataset is representative of the procedural and interpretive challenges encountered in statutory land consolidation practice, rather than providing exhaustive coverage of all legal provisions. By selecting questions across core project stages and stratifying them according to increasing legal and cognitive demands, the dataset reflects the progression from routine statutory inquiries to integrative reasoning tasks involving procedural dependencies, statutory exceptions, and institutional discretion. This design enables a systematic evaluation of model performance under conditions that closely mirror real-world legal decision-making in land consolidation contexts.

3.1.2. The types of questions

The assessment started with 100 questions extracted from the Cyprus land consolidation legal framework, focusing on the Land Consolidation Law (1969–2018) and its modifications. The question corpus was established through a thorough examination and professional understanding of legal texts, departmental circulars, and procedural guidelines defined by the ex-Land Consolidation Department, which dissolved in 2018 and joined the Land and Surveys Department as a Land Consolidation Branch.

The questions covered all possible levels of legal complexity which appear in actual legal situations, from basic definition understanding to procedural analysis and comprehensive legal understanding. They developed four types of questions to demonstrate various difficulty levels of interpretation, which are explained below:

Type 1: Basic Legal Facts and Definitions (27 questions)

These tests recall legal terms, the identification of actors, and procedural timelines. Examples: (i) What are the three possible methods for implementing land consolidation in Cyprus? (ii) Summarise the main steps leading to the decision to implement land consolidation using the "majority decision by landowners" method. (iii) What are the two main conditions for making a majority decision? (iv) Summarise the powers and duties of the Land Consolidation Committee.

Type 2: Procedural Order and Sequential Reasoning (31 questions)

These assess procedural order understanding and reasoning across the steps. Examples: (i) Can any landowner be elected as a member of the Land Consolidation Committee? (ii) What principles regarding land valuation must the Valuation Committee follow, and which law governs them? (iii) Within the consolidation plan, is it possible to alter river boundaries relative to private land under the principle of exchange? What is the prescribed procedure for this? (iv) When is compulsory expropriation carried out, and what are the minimum thresholds?

Type 3: Cross-Article or Clause Integration (32 questions).

This requires combining data from different articles or clauses. Examples: (i) What is the allocation of expenses, and on which principle is the calculation based? (ii) A landowner received a plot of 6800 m^2 in an irrigated consolidated area. Can it be subdivided? If yes, what is the minimum area and number of plots required? Is approval required, and by whom is it required? (iii) A landowner owns 8000 m^2 in an irrigated area—can it be transferred to two children after consolidation? If not, what is the solution? (iv) After a landowner dies post-consolidation, can three heirs receive the property in undivided shares, and under what conditions?

Type 4 – Abstract and Interpretive Questions (10 questions).

These demands abstract legal understanding, ambiguous interpretations, and exceptional cases. Examples: (i) If a landowner's properties are excluded but they still benefit from the road network, can they be included among the contributors and share costs if they consent? (ii) Can new evidence or expert testimony be admitted after the appeal deadline? (iii) After redistribution, may a plot be granted as co-ownership to landowners who request it, even if they were not co-owners before the redistribution? (iv) During appeals, does the Appeals Committee have the authority to request clarifications or consultations with landowners and assign tasks to the Department or the Land Consolidation Committee?

For all questions, internal validation by two experts in land consolidation confirmed that each question was free of ambiguity and matched the intended difficulty level. The question corpus contained jurisdiction-specific content while ensuring legal coherence and representing the types of enquiries which occur in legal consultancy and land consolidation planning. Table 1 shown below, summarises the coverage of the land consolidation legal framework within the question corpus, showing the distribution of questions across procedural phases, institutional actors, and primary legal sources

Table 1
Coverage of the land consolidation legal framework within the question corpus.

Procedural phase/legal topic	Institutional actor (s)	Number of questions (n)	Primary legal sources
Initiation & decision mechanisms	Landowners, Land Consolidation Committee, District Office	**24**	Law Articles 4, 7–9; related Circulars
Institutional composition & powers	Committees, Authorities (Department, Minister)	**18**	Law Articles 10–13, 46; related Circulars
Valuation principles & compensation	Valuation Committee, Department	**17**	Law Articles 14–15, 26; Cap.224; related Circulars
Design, road planning & redistribution	Committee, Department	**21**	Law Articles 20–21, 25–30; First & Second Schedules; Guidelines
Appeals & dispute resolution	Appeals Committee, Minister, Courts	**12**	Law Articles 13, 20, 26, 34; related Circulars
Registration & post-consolidation effects	Department of Lands & Surveys	**8**	Law Articles 28–29; Cap.224
Total		**100**	

3.2. Scoring rubric and evaluation criteria

The scoring system consisted of five levels, from 0.0 (incorrect/poor) to 1.0 (excellent), with 0.25 increments. Other studies used specific scoring systems fitting their scope (Zhou, 2023; Kaplan et al., 2024). More precisely, the scoring rubric assessed four core response quality factors: (i) Correctness: The extent to which the answer provides factually and legally accurate information; (ii) Completeness: All the key elements of a legal query are addressed; (iii) Clarity: Linguistic coherence and logical structuring of answers; and (iv) Interpretive Depth: The ability to recognise nuances, cross-reference legislation, or interpret procedural intent.

To cover the quality factors, they were defined as the following five score definitions: Excellent (1.0); Good (0.75); Satisfactory (0.50); Needs improvement (0.25) and Poor/Incorrect (0.0). "Excellent" means that the answer was completely correct, clear, and well explained. The answer shows comprehensive legal knowledge while maintaining a logical organisation with both direct references and implied reasoning. The answer contained no incorrect information or missing details. "Good" indicates that the answer is mostly correct but may omit minor details. The answer shows robust comprehension; however, it does not deliver complete interpretive understanding. The response may contain limited incorrect information which does not impact the legal meaning. "Satisfactory" suggests that the answer is partially correct but incomplete or lacking depth. This may include significant omissions, unclear language, and moderate misunderstandings. "Needs Improvement" indicates that the answer was largely incorrect or based on major misinterpretations. The answer fails to show a logical structure and contains fundamental mistakes. "Poor/Incorrect" indicates that the answer is completely wrong, irrelevant, or nonsensical. This reflects a lack of understanding of the legal question or an entirely misapplication of legal principles. This rubric allowed for a nuanced evaluation that reflected both the precision and the interpretive nature of legal analysis.

3.3. Evaluation procedure and model settings and operational environment

The author, with 25 years of experience implementing land

consolidation in Cyprus, first evaluated each ChatGPT-generated answer. The second review of the evaluation process involved a current expert working on land consolidation in Cyprus for approximately 40 years. The evaluation consistency of the 25 answers was measured through inter-rater agreement (the degree to which two or more independent evaluators gave consistent ratings, classifications, or judgments when assessing the same set of items). This resulted in an 88% observed agreement and a Cohen's kappa (a statistical measure of inter-rater reliability (or agreement) for categorical data) score of 0.82 which indicates a strong agreement; hence, their agreement is very strong and not due to chance. In cases of disagreement during the evaluation process, the researchers referred to the relevant legal framework to ensure a consistent and accurate interpretation. To prevent prompt engineering bias or answer conditioning, the ChatGPT interface received each question as a separate input. The evaluation process avoided any system instructions and contextual primers because it aimed to reproduce the actual usage behaviour of nonexpert users.

In this context, the GPT-4 model generated all responses through the ChatGPT interface provided by OpenAI (as of April 2024). The author accessed the model through a professional subscription account using the default system parameters (Grigorov, 2025; Peng et al., 2023). Specifically, the "temperature" was set to 0.7 to allow moderate variability in the output while preserving determinism. "Temperature" is a critical setting that controls the randomness or creativity of the model response. Typically, it ranges from 0.0 to 1.0 (occasionally up to 2.0 in some models). A low "temperature" (a value of 0.1 or 0) makes the model more deterministic, providing the most likely and safest answer. A higher "temperature" (e.g., a value of 0.7 or 1.0) increases variability and creativity; the model takes more "risks" and produces more diverse or imaginative answers. Thus, they are used for low "temperature" values when we wish for reliable, consistent, factual output, and they are used high temperature when we wish brainstorming, creative writing, or varied suggestions.

Another parameter is the "maximum token limit". That is, the maximum number of tokens (word pieces) that the model can use in a single response (or a request + response combined, depending on the setup). Tokens are word components. For example, "ChatGPT" = 1 token; "unbelievable" = 2 tokens ("un", "believable") Set at 1024 tokens per response to ensure the completeness of answers. The Model version used was GPT-4-0314 (the latest stable release at the time of testing). The author initially uploaded the relevant legislation and documents for the administrative circulars and then submitted each query to the new browser session without storing data to prevent both caching and learning effects before saving the answers directly to plain text documents. The model performed without any training on Cypriot legal documents, and it received no legal corpus information before the evaluation. The results reflect the base model performance because the testing did not involve any customisations which is crucial for generalising the results.

The results and analysis were based on a structured spreadsheet using standard statistical methods. The evaluation system used the following key metrics to assess performance: Average score per question type; Distribution of score frequencies across difficulty levels; Performance deviation from means per level; Standard deviation and variance in scoring.

Qualitative assessments were conducted to identify the patterns of AI response errors and omissions. The system produced Type 4 answers containing relevant legal principles from other jurisdictions instead of using local legislative requirements because it preferred basic legal terminology. The Results (Section 4) combines quantitative data with output examples to show performance patterns. In addition, it should be noted that the study contained no personal data, human subjects, or confidential legal cases.

4. Results

4.1. Overall performance and breakdown by question type

The overall performance of the model achieved a score of 84% on 100 legal questions about land consolidation, indicating strong performance in Cypriot legal question answering. The model achieved remarkable performance despite the lack of training data on Cypriot legal documents and land law administrative procedures in this jurisdiction. The score distribution revealed that the model achieved a perfect score of 1.0 on 54 questions, while it failed on eight questions which had a score of 0.0. The model demonstrated strong performance in answering questions effectively and providing clear and complete responses. The model provided consistent legal definitions, procedural stages, and logical relations in legislative texts despite Type 4 high-complexity challenges.

Further, the overall performance was evaluated by question type. In other words, the extent of ChatGPT's performance differences based on difficulty levels and cognitive complexity, the 100 questions were classified into four distinct categories: The performance data from these categories appears in Table 2.

Type 1 – Basic Legal Definitions: ChatGPT obtained an average score of 98% which was the highest among all question categories. The model produced 1.0 scores for 24 questions, but it delivered a 0.75 (out of 1.0) score for the remaining three questions. The model excels in answering questions that demand terminology knowledge and definitions of legal terms, as well as understanding institutional roles and procedural requirements, such as quorum and area thresholds. The model displays consistent behaviour because it receives pre-training data with explicit encoding of such concepts.

Type 2 – Procedural Logic: The performance score reached 84% among 32 questions, even though the model performed slightly worse than the previous results. Although the model provided mostly accurate answers, it made mistakes when analysing sequential procedural steps and exception clauses. The higher standard deviation (0.32) indicates that the response quality depends on specific procedural details.

The model achieved a performance score of 81% for Type 3 Hypothetical Reasoning questions. The model was used to implement legal concepts in theoretical situations that included various land conditions and partial authorisation elements. Type 4 – Complex Interpretation: Among the 10 questions in this category, the lowest performance occurred at 55%, with a standard deviation of 0.50. The model faced major difficulties when handling questions that required multiple layers of interpretation and context-specific reasoning, as it failed on four out of ten questions. The model either applies to generalisations that exceed legal boundaries or merges distinct procedural elements that should remain separate from each other.

4.2. Statistical analysis and qualitative assessment

Based on the above statistical analysis and observations, the results showed that more complex questions received lower-quality answers in

Table 2
ChatGPT performance by question type.

Question Type	Number of questions	Average Score	Standard Deviation	Min	Max
Type 1 – Basic legal definitions	27	98.1%	0.067	0.75	1.0
Type 2 – Procedural logic	32	83.6%	0.321	0.00	1.0
Type 3 – Hypothetical reasoning	31	81.5%	0.288	0.00	1.0
Type 4 – Complex interpretation	10	55.0%	0.497	0.00	1.0

direct proportion. The model achieved near-perfect results when performing basic recall duties; however, it demonstrated rising mistakes when asked to combine legal information or make complex interpretive assessments. The standard deviation levels increased with difficulty because the model maintained 0.07 in Type 1 but reached 0.50 in Type 4. The model experiences declining response stability when performing tasks that require complex legal and cognitive processing skills.

Furthermore, the results showed that 1.0 scores were achieved by 56%, all questions and 66% of Type 1, Type 2 and 48% of Type 3 questions, respectively. The research lacked any instances of 0.0 scores in Type 1, but Type 2 recorded four instances, and Types 3 and 4 had three and four instances, respectively. The highest score obtained by Type 4 questions was 1.0 which indicates that the best possible answers in this category did not follow a standard pattern. The quantitative evaluation was supplemented by qualitative assessment and error typologies, which showed how the model applied its reasoning mechanics in making mistakes based on five factors explained immediately below.

(i) Overgeneralisation: The model applied generic principles in Types 3 and 4 responses to support its answers but did not reference the legal provisions specified in the questions. The results showed that the Cypriot Legal Corpus was not well represented. (ii) Omission of Exceptions: When answering procedural questions (Type 2), the model failed to consider exceptions and fallback procedures which resulted in partially correct responses. (iii) Ambiguity in Language: Language output was smooth, legal precision was occasionally impaired. The model utilised terms such as “sometimes” and “generally” in its responses, which created problems for legal interpretation. (iv) Misclassification of Authority: The model made small errors in authority classification by attributing decisions to “Minister” when the decisions belonged to the Land Consolidation Board or General Attorney; (vi) Inferred but Unverifiable Statements: The model produced plausible yet non-existent legal procedures in its responses. These hallucinations occur when the model encounters interpretive ambiguities or rare legal provisions.

Based on the above results, the average accuracy score is impressive, but the model performs poorly in challenging scenarios, raising concerns about its ability to function independently for legal interpretation. The results show that ChatGPT demonstrates two functions in legal applications: it performs well in memory-based and simple reasoning tasks but struggles with complex legal combinations and subtle judgments. The practical implementation of this tool in land consolidation planning must be restricted to supporting functions that require expert supervision only. In addition, the high standard deviation in the hardest category shows unstable results rather than predictable logic. Users who depend on such models for decision-making need to evaluate the outputs carefully, especially in situations where interpretive consequences are high, such as landowner rights and dispute resolution.

Section 5 interprets the results by identifying performance strengths, limitations, and performance drivers. Section 6 (Implications for policy and practice) converts these interpretations into actionable advice about safe and unsafe practices and governance and policy development.

5. Discussion

The empirical results from the previous section are analysed, that is, scientifically interpreted, in this section to answer the research question about how well ChatGPT serves as a general-purpose large language model (LLM) for providing reliable legal answers in land consolidation legislation. The results are analysed, and the reasons for the results are contextualised by examining the evidence in prior work, identifying strengths and limitations across question types, and explaining performance drivers. This section establishes the warrants for the practical recommendations that follow the study. The discussion follows an analytical framework based on five elements: (i) reasoning capabilities, (ii) strengths, (iii) weaknesses, (iv) causes of performance variation, and (v) comparisons with human or traditional legal research.

5.1. Limitations and comparison with human performance

The research design of this study consists of a descriptive benchmark which evaluates a single LLM through a 100-item instrument based on statutory content. The percentages in this study function as relative indicators of item difficulty between question types, but they do not represent absolute legal competence measures.

Internal control (applied in this study):. The scoring rubric implemented a rule that prevented the author's key from being considered perfect by requiring answers to demonstrate correct legal authority identification for high scores. The scoring system limited all responses without proper authority identification to lower marks, regardless of their overall plausibility. The source-anchored rule serves as an internal control mechanism to differentiate between answers that reference legal sources and those that do not.

Comparative stance:. Research studies involving human-LLM interactions in legal settings show that human professionals excel at complex interpretation tasks, but LLMs match human performance on straightforward definitional and procedural questions [add 1–2 citations from Section 2]. The observed results match those of previous research because the model performs better on Types 1 and 2 but shows a significant decline when handling Type 4 questions. The results of this study present relative performance data between question types instead of providing human reference points.

The following step involves establishing human and multiple model performance baselines through identical testing conditions with the same instrument and evaluation criteria, although this study did not require baseline assessment. This research provides a standardised evaluation method which allows future studies to perform comparative assessments.

5.2. ChatGPT demonstrates legal context reasoning abilities

The model demonstrated solid basic legal reasoning abilities through its ability to perform syntactic parsing, semantic coherence, and procedural-logic recognition in Type 1–3 (low-to-mid complexity) questions. Katz et al. (2024) showed that their model achieved high fluency while effectively summarising content and replicating both statutory and judicial discourse patterns. The model fails to produce legal judgments because it depends on statistical pattern recognition and lacks both a normative framework and an institutional context which prevents it from resolving conflicting legal values or interpreting principles (Nelson, 2023).

5.3. Strengths and weaknesses

The model achieves its highest performance when processing straightforward tasks that involve recalling statutory definitions, procedural requirements, and institutional roles. The model performs well in tasks that require rule-based processing of standardised information (Hamdani et al., 2024; Surden, 2021), who state that AI tools excel in rule-based environments. The model maintains linguistic coherence when it makes small factual errors.

According to Tan et al. (2023), the model performs poorly when handling complex problems that require interpretation across multiple clauses and ambiguous situations which require multiple layers of judgment. The combination of spatial elements with environmental factors and cadastral requirements in Cypriot land consolidation requires extensive knowledge of local jurisdictional regulations. The model produces general concepts such as "fairness" and "equality of land use" instead of local legal principles specific to the local area which results in unreliable results. The system fails to properly address statutory exceptions and special procedures because it produces basic yet insufficient answers and generates plausible results that do not include the necessary legal conditions. The operational results of these findings are presented in Section 6.

5.4. Causes of performance gaps

The system encounters three main limitations which stem from its data processing capabilities and its ability to reason and establish authoritative sources. The model uses general legal traditions from common and civil law because Cypriot land consolidation law rarely appears in public web data, resulting in decreased accuracy when processing Cyprus-specific provisions that include geographic rules, administrative procedures, and cadastral definitions. The probabilistic nature of GPT-4 prevents it from performing causal and normative legal reasoning and understanding legal concepts which hinders its ability to apply proportionality principles and handle exceptions and conditional statutory rules (Jung et al., 2024). The model produces answers that seem plausible but lack strong references because it depends on limited training data which affects its ability to provide authoritative citations. According to established research, legal credibility depends heavily on traceability and source transparency (Aletras et al., 2016). Section 6 addresses operational remedies and governance aspects.

5.5. Legal applications: human comparison and domain-specific adaptation

The initial legal analysis provided by LLMs lacks human professional standards for deep interpretation, contextual understanding, and source verification. The legal decision-making process includes multiple accountability measures through peer review, appeal systems, and consensus building which generative systems cannot replicate, thus increasing the risk in critical domains such as property rights and land allocation and disputes because errors create enduring consequences (Biswas, 2023).

The technical requirements for addressing the identified gaps include two specific solutions: Fine-tuning the model with Cypriot land consolidation statutes, administrative regulations, and case law will enhance its ability to handle jurisdiction-specific legal doctrines and exceptions. The combination of retrieval-augmented generation (RAG) with authoritative sources enables better output traceability and minimises the occurrence of unsupported claims (OpenAI's "ChatGPT with Retrieval" and DeepMind's RETRO). The combined systems enhance the ability to perform tasks that require verifiable references and multiple clause analyses.

The integration of spatial-legal systems through SDSS/LIMS represents a promising technical path because of the domain's spatial requirements; however, operational deployment and oversight mechanisms require analysis (Section 6) (Dwivedi et al., 2024).

By examining how statutory land consolidation law operates in Cyprus under the current legal framework, this study moves beyond a purely descriptive evaluation of model outputs and provides analytical insight into how large language models perform when confronted with legally rigid, procedurally layered systems. The findings demonstrate that, while the model can handle basic statutory inquiries, its performance deteriorates as legal complexity increases, particularly where specific code requirements, strict procedural sequences, and interdependent legal provisions must be interpreted coherently. Situations in which different legal sections interact, conflict, or generate conditional outcomes produce responses that deviate from typical linguistic patterns, revealing limitations in probabilistic reasoning when applied to codified legal systems.

The progressive structuring of questions from simple to complex enables a systematic examination of how such models attempt to navigate legal environments characterised by procedural accumulation, structural rigidity, and ongoing legal evolution. In particular, the evaluation highlights difficulties in accounting for statutory exceptions, recent judicial interpretations, and internally inconsistent legal provisions—features that are central to real-world land administration practice. These findings provide empirical evidence of model performance under conditions of institutional and legal complexity. While the analytical insights are derived from the specific context of statutory land consolidation in Cyprus, their relevance to other jurisdictions is contingent upon the presence of comparable legal rigidity, procedural density, and governance structures, rather than implying broad generalisability across legal systems.

The findings from this section lead to practical recommendations for deployment scope, assurance, and policy development in Section 6.

6. Implications

The research findings established major changes which affect the discussion about artificial intelligence implementation in the legal and geospatial fields when these systems operate between statutory interpretation, spatial evidence, and administrative practice. The land consolidation process requires advanced technical expertise and complex legal systems which link spatial planning to cadastral administration and rural development, thus generating both significant possibilities and major threats for using the ChatGPT AI tool in operational systems. The authors use data from the previous section to create particular implications through a methodical system which assesses four essential elements: (i) stakeholder-related advantages and constraints, (ii) protected applications versus dangerous operations, (iii) monitoring systems and verification protocols, and (iv) detailed guidelines for public policy and procurement procedures. The authors define particular operational limits for this technology in this section which enables users to understand its functional range and requires human involvement for operation.

6.1. Relevance to planners, legal practitioners, and public authorities

The system design provides essential advantages to spatial planners and land surveyors who perform land consolidation work with limited resources and short deadlines within their small teams. The conversational interface enables users to directly access procedural descriptions, legal terminology, and institutional roles, making their search for information between multiple regulatory texts and technical guidelines more efficient. The system functions as a tool to assist planners during their preparation work, initial evaluations, and public outreach activities when they lack direct access to expert legal or cadastral knowledge.

The system provides legal practitioners, including lawyers, notaries, and legal scholars who work on land policy and property law, with research tools for exploratory and comparative studies instead of providing official legal interpretations. The system enables legal reasoning and academic research through its support of relevant legislative provisions, procedural steps, and conceptual distinctions while maintaining its position as a non-binding legal advice system. This distinction is essential because qualified human experts must perform statutory interpretation, dispute resolution, and rights adjudication duties when land consolidation projects establish permanent spatial and socioeconomic transformations (Jiménez, 2024).

The system delivers limited yet important value to public authorities and administrative bodies which make up its third stakeholder group. The system operates as an informational interface which enables transparency and accessibility through its guidance of citizens, landowners, and local stakeholders during consolidation phases, institutional responsibility periods, and participation processes. The deployment of these tools with proper security measures helps decrease information gaps, leading to better citizen comprehension of complicated administrative processes. However, organisations must maintain open communication about system boundaries to prevent users from making incorrect assumptions (Androniceanu, 2024).

6.2. Use cases: ChatGPT in land consolidation workflows

The system functions as a supplementary system which works together with current administrative and professional operations in land

consolidation workflows instead of making independent decisions. The system provides its most direct value through its ability to merge automated procedural templates with human monitoring which enables users to receive organised help while they remain in charge of all interpretation and validation tasks. The system enables users to find their way through legal terminology, court procedures, and official rules which exist across various legal documents and administrative rules.

The system provides functionality which enables users to perform their daily administrative work during land consolidation projects. The system produces standardised drafts for notices, invitations, and explanatory documents which create uniform output and reduce official work because duplicate documents do not need to be checked. The system provides interactive functionality which helps users understand consolidation phases, their related terms, and common situations that become essential for successful landowners and community participation in planning activities. The system operates as a facilitation tool which enables users to understand information, but users need to use it only as a supporting resource for official consultation and legal professional advice (Tugendhaft et al., 2020).

The system provides access to WebGIS, SDSS, and LIMS through its natural-language interface functionality. Users can ask parcel-related or procedural questions using normal language which produces structured summaries that lead users to official spatial and legal documents. The system is best used in agricultural and land reform portals which provide citizen-facing or advisory services because they must deliver accessible and easy-to-understand information. Regulated domains require these uses to remain non-binding because organisations must follow their current governance and risk-mitigation practices (Kalaiselvi et al., 2024; OECD, 2023).

6.3. Ethical and reliability considerations

The deployment of ChatGPT in legal and land policy environments demands specific protective measures because the operational risks are well documented. In particular, the combination of hallucinations and jurisdictional misfits can produce legally incorrect yet believable responses, which can damage user trust and mislead decision-making in rights-sensitive contexts (Livermore et al., 2024). Therefore, the credibility of high-stakes use cases depends on verifiable sources and transparent reasoning; legal AI systems must provide claim provenance and clearly traceable inferential pathways rather than relying solely on fluent natural language output (Aletras et al., 2016).

These concerns extend beyond answer quality to public sector governance obligations. Public sector data management requires strict privacy regulations, access controls, and retention policies, particularly because citizen information and cadastral records may pass through cloud-based APIs and related service layers. In addition, the system must be maintained through periodic updates or, preferably, dynamic connections to official databases, so that users do not receive outdated guidance as statutes, circulars, and procedures change.

From an operational standpoint, the findings indicate that safe deployment depends on translating ethical and legal principles into concrete, institutional safeguards. First, all outputs with potential legal or property implications must remain subject to systematic human review and approval, ensuring that the responsibility for rights-affecting decisions is never delegated to the system. This requirement must be complemented by clear user-facing warnings and readily available access to expert assistance so that users are continuously reminded of the system's advisory role. In parallel, comprehensive logging of all prompts and responses is necessary to support auditability and post-hoc accountability. Reliability is further strengthened when system outputs are explicitly linked to authoritative sources, including the capability to withhold responses in cases in which such sources are unavailable or ambiguous. Finally, all system operations must comply with established institutional data management rules, ensuring alignment with public sector governance standards and the analytical performance framework discussed in Section 5.

6.4. Future AI integration in legal-geospatial systems

The deployment of LLM technology into legal–geospatial systems requires a systematic approach, including controlled system deployment and technical solutions that improve both geographic boundary accuracy and data monitoring and organizational supervision functions. The development of the Cypriot land-law corpus requires domain adaptation because it needs a systematic collection of statutes, ministerial circulars, cadastral guidelines, and court decisions which pertain to the subject matter. The LegalBERT model learns from a curated corpus which contains data that helps the model understand both jurisdictional matters and statutory exceptions and special procedures (Chalkidis et al., 2020; Yang et al., 2024).

The operational integration process needs to establish connections between LLM output and official legal and spatial data systems which enable users to monitor the source of generated statements throughout the response procedure. The system provides enhanced value when users access SDSS/LIMS/WebGIS platforms through natural-language query interfaces which enable them to obtain parcel-related legal-spatial data that matches GeoAI objectives while maintaining a connection to official documents (Grădinaru et al., 2025).

The safety of these integrations depends on starting with limited pilot tests which produce quantifiable results before expanding their application scope. Pilots must measure their performance using specific performance indicators, including document type accuracy rates, citation verification, human referral rates, and time reduction metrics, to determine the actual benefits and risks. The system requires periodic updates which must be supported by detailed technical records, sufficient maintenance resources, and audit systems to operate sustainably. The system allows users to choose from open modules which provides them with the ability to perform independent assessments and minimises their need for untraceable system components.

6.5. External validity and transferability

The research uses Cyprus land consolidation as its empirical base, yet its results apply to similar institutional frameworks which exist in other countries. The research findings demonstrate transferability because the methodological framework, safeguard logic, and deployment principles focus on structural elements which exist across different jurisdictions that implement land consolidation and land policy programs. The system needs to show its procedures openly, while different groups must work together, decisions must follow legal rules, and the system must use both physical and organizational data for its governance structure. The evaluation method, together with deployment specifications, establishes a common framework which researchers can use to study how language models affect land consolidation operations and other land management activities despite varying national laws, administrative systems, and available data. The findings show that organisations need to transform their current structures for knowledge transfer success instead of attempting to replicate the original system. The research shows that designers must develop solutions which fit the particular conditions of their target organisations to achieve successful outcomes.

6.6. Policy recommendations

The research findings show that organisations need to follow a regulated approach when deploying large language models for land consolidation and land administration operations. The current state of technology requires organisations to use AI systems for non-critical tasks which help users find information, understand procedures, and organise documents. All legal interpretations which bind and all judicial decisions must remain under human control. This restriction is significant during land consolidation operations because it determines how

property rights, spatial arrangements, and future socioeconomic results will develop.

The deployment of these systems must occur through responsible methods which require operational usage to take place within a governance framework that defines system boundaries and establishes responsibility and verification procedures. The system enables retrieval-augmented generation pipelines to access official legislative and cadastral databases through direct database access which produces outputs that remain true to their authoritative sources and enable independent verification. The system needs technical solutions which need to work together with established institutional rules that define human participation levels and establish who bears responsibility for system failures and ambiguous situations.

Organisations must develop their operational capabilities and create open systems for resource procurement to succeed in system implementation. The supervision of AI-assisted outputs requires planners, administrators, and technical staff to develop specific competencies, including bias detection, hallucination identification, and system constraint understanding. Public-sector procurement needs to focus on three essential elements: transparent operations and audit capabilities and sustainable long-term solutions through model behaviour disclosure, performance metrics, update schedules, and data management practice requirements for pilot testing before implementing them on a larger scale.

7. Conclusions

7.1. Key takeaways

The primary result showed that ChatGPT delivered outstanding results for legal tasks at both low and moderate complexity levels, but its performance weakened substantially when tackling complex interpretive legal questions. The model obtained an average score of 84% while reaching its peak accuracy of 98% in Type 1 basic legal definition questions, but its lowest performance of 55% occurred in the abstract legal scenarios of Type 4. The evaluation indicates that ChatGPT has strong capabilities to retrieve legal documents and describe procedures while creating coherent legal responses in typical situations, yet struggles with tasks that require synthesis, legal judgment, and contextual understanding.

The results align with the research in the developing field of legal AI evaluation. The performance of large language models improves in general language understanding as their parameter size and training complexity increase. Advancements in these areas do not necessarily lead to uniform success in law because this field requires textual nuances, normative interpretations, and jurisdiction-specific logic.

In addition, this study introduces an original assessment method based on empirical evidence to examine how ChatGPT handles land consolidation legal issues which are crucial for both law and geospatial domains but have received little attention in AI research. The current research evaluates LLMs using general law exams and judicial decision prediction and contract analysis, but this study evaluates the performance in actual statutory frameworks while using detailed scoring rubrics based on practical expectations.

The analysis evaluated ChatGPT using 100 realistic questions encompassing multiple aspects of land consolidation, including institutional elements and interpretive judgment. The detailed breakdown of the assessment results, together with the question-type distribution, creates a valuable methodological approach for researchers to evaluate AI models in specialised fields.

This research adds to the ongoing dialogue on how to establish reliable AI systems while ensuring trustworthiness and effective implementation in legal operational workflows. This study establishes clear distinctions between AI capabilities (language fluency, surface reasoning, and pattern replication) and its ongoing limitations (normative understanding, source grounding, and factual verifiability).

In addition, the results of this study strongly suggest the need for domain-specific fine-tuning of general-purpose language models. The model performed well in legal recall but failed to deliver when it encountered tasks that required domain-specific interpretive reasoning. The primary reason for this performance decline stems from the fact that the documents uploaded to ChatGPT were in Greek, while OpenAI was trained in English.

The model's performance would improve substantially through targeted fine-tuning which involves using carefully curated legal corpora, including statutes and case law, as well as official guidelines and administrative circulars. This methodology has proven successful in the medical field, as well as in financial operations and international legal domains. LegalBERT and CaseLawBERT achieved superior results compared to generic models when processing statute classification and case outcome prediction through training on court opinions and legal statutes (Chalkidis et al., 2020).

The training corpus for land consolidation should contain National and EU-level legislation on land readjustment and agricultural restructuring, planning and zoning regulations, legal commentaries and explanatory documents from the Land Consolidation Department and Case decisions involving land disputes, valuation, or project approvals. Moreover, the performance can be improved using retrieval-augmented generation (RAG) techniques which enable the model to dynamically retrieve legal databases and spatial records during the generation process. The hybrid approach that combines generative fluency with knowledge retrieval presents a promising direction by ensuring that AI responses are fluent, context-sensitive, and verifiable.

7.2. Future work

This study examines the land consolidation law. The legal system contains complex regulations relevant to zoning regulations, urban redevelopment, cadastral management, eminent domain, and land taxation. Domains that combine legal codes with geospatial data should serve as testbeds for evaluating the performance of legal AI. Building on this foundation, the assessment was conducted in Greek and focused on the legal environment of Cyprus. Future research should evaluate multilingual models in other languages to analyse how different legal systems and linguistic variations affect AI reasoning across various jurisdictions. The evaluation would prove most relevant in the European Union context because member states implement directives differently.

To implement AI in land governance operations, it is necessary to validate real-world workflows among human planners, legal officers, and AI tools. Experimental tests involved using ChatGPT for land consolidation tasks by participants to measure their efficiency, accuracy, and decision-making confidence. The “human-in-the-loop” design helps researchers determine usability barriers and integration possibilities for the workflow. Equally important is the need to examine the ethical and societal implications of this integration. Future research must focus on understanding citizen trust, alongside the legal accountability and ethical acceptance of AI systems in property and planning decisions. The process of making land-related decisions involves crucial rights and past injustices to the Indigenous peoples. AI implementation in critical domains requires socio-legal evaluation to determine legitimacy, as well as transparency standards and mechanisms for redress. Public interfaces that use AI technology require clear statements, traceable logic systems, and features that enable human intervention.

Research teams and software developers must develop methods to incorporate LLMs into spatial decision support systems (SDSS) used for land consolidation planning. Users can utilise ChatGPT to access geospatial and legal information and bespoke GIS planning tools when they query natural language. The system enables planners and land surveyors to embed legal constraints into land-use modelling while performing scenario evaluations and stakeholder communication. To support and evaluate such integrations effectively, researchers should prioritise developing benchmark datasets that assess AI performance within

specific legal domains. This will enhance reproducibility and enable comparative research through the creation of Massive Multitask Language Understanding (MMLU) datasets. A benchmark for land consolidation law would consist of a set of canonical questions at varying difficulty levels, accompanied by annotated legal references and validated scoring rubrics developed by experts in the field.

As a final reflection, it can be said that the combination of artificial intelligence systems with legal, land-based spatial planning functions has evolved from theory into practical reality. The research demonstrates that ChatGPT and similar tools can deliver practical value by providing legal explanations, procedural directions, and generating relevant text. While the development of AI from general-purpose models to domain-specific and legally responsible systems remains an active process, this study establishes that large language models (LLMs) function as predictive systems, offering strong yet imperfect capabilities to users. Their predictive nature gives them significant power and appeal but also makes them susceptible to errors and instability. In this context, their role in legal development is to support human legal reasoning by improving access to legal information, accelerating routine tasks, and introducing new methods of interacting with legal knowledge. With proper training, ethical safeguards, and intelligent deployment, the integration of such AI systems has the potential to establish AI as a core partner in the complex work of developing sustainable rural areas through more effective land management, legal frameworks, and the development of livelihoods.

This study proves that ChatGPT serves as an effective tool for supporting legal-geospatial activities, such as land consolidation. The system requires strict boundaries and strong legal monitoring, and developers must continually enhance it with domain-specific adjustments. The future of land management and spatial planning will become possible when AI achieves a trustworthy partnership status through these measures.

CRediT authorship contribution statement

Demetris Demetriou: Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of Competing Interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Aidaoui, A., Dechaicha, A., Alkama, D., Menai, I., Salah, H., 2024. Mapping tomorrow's cities: GeoAI strategies for sustainable urban planning and land use optimisation. J. Contemp. Urban Aff. 8 (1), 158–176.

Aletras, N., Tsarapatsanis, D., Preoţiuc-Pietro, D., Lampos, V., 2016. Predicting judicial decisions of the European Court of Human Rights: a natural language processing perspective. PeerJ Comput. Sci. 2, e93.

Alterman, R., 2010. Takings international: a comparative perspective on land use regulations and compensation rights. Soc. Sci. Res. Netw.

Androniceanu, A., 2024. Generative artificial intelligence: present and perspectives in public administration. Admin. Şi Manag. Public 43 (43), 105–119.

Biswas, S., 2023. Role of ChatGPT in law: according to ChatGPT. Deleted J. 1 (1).

Chahal, N.K., 2024. Transformative trends: exploring the evolving role of artificial intelligence in the legal landscape. Int. J. Sci. Res.

Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., Androutsopoulos, I., 2020. LEGAL-BERT: the Muppets straight out of Law School. In: Findings of the Association for Computational Linguistics: EMNLP 2020. Association for Computational Linguistics, pp. 2898–2904.

Chen, B., Huang, Z., He, W., Wang, M., 2024. Comprehensive land consolidation as a development strategy for rural revitalisation: political ecology mechanisms and benefits of the pastoral complex. Land 13 (6), 897.

Darwish, D., 2024. Geospatial AI future perspectives. Adv. Geospat. Technol. Book Ser. 297–324.

Demetiou, D., 2026. Fifty-five years of experience with land consolidation in Cyprus: lessons learned and future directions. Land Tenure J. in preparation soon.

Demetriou, D., 2013. LACONISS: a land consolidation integrated support system for planning and decision making. zfv – J. Geod. Geoinf. Land Manag. 138 (2), 119–131.

Demetriou, D., Stillwell, J., See, L.M., 2012. LandParcelS: A Module for Automated Land Partitioning (Working Paper No. 12/02). University of Leeds, School of Geography, UK.

Dwivedi, D.N., Mahanty, G., Dwivedi, V. nath, 2024. ChatGPT and AI in the government. Adv. Comput. Intell. Robot. Book Ser. 47–68.

European Parliament and Council of the European Union, 2007. Directive 2007/2/EC establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Off. J. Eur. Union, vol. L 108, pp. 1–14.

Fahrani, A., Djajaputra, G., 2024. Legal validity of artificial intelligence technology on GPT chat as legal aid. J. Law Poliit. Humanit. 5 (1), 54–61.

Fei, Z., Shen, X., Zhu, D., Zhou, F., Han, Z., Zhang, S., Chen, K., Shen, Z., Ge, J., 2023. LawBench: benchmarking legal knowledge of large language models. arXiv.Org abs/2309.16289.

Grădinaru, A.P., Badea, A.-C., Ene, A., Badea, A.-C., 2025. The integration of geospatial analyses based on AI in GIS in the context of the "15-min city" concept. E3S Web Conf. 608, 05008.

Grigorov, D., 2025. Building and fine-tuning LLMs. In: Intermediate Python and Large Language Models. Apress, Berkeley, CA. https://doi.org/10.1007/979-8-8688-1475-4_5.

Hamdani, R., Bonald, T., Malliaros, F.D., Holzenberger, N., Suchanek, F.M., 2024. The Factuality of Large Language Models in the Legal Domain.

Hosseini, M., Gao, C.A., Liebovitz, D.M., Carvalho, A.M., Ahmad, F.S., Luo, Y., MacDonald, N., Holmes, K.L., 2023. An exploratory survey about using ChatGPT in education, healthcare, and research. PLoS One 18 (10), e0292216.

Hwang, Y.-S., Um, J.-S., Pradhan, B., Choudhury, T., Schlüeter, S., 2024. How does ChatGPT evaluate the value of spatial information in the 4th industrial revolution? Spat. Inf. Res. 32 (2), 187–194.

Jiang, Y., Tang, Y.-T., Long, H., Deng, W., 2022. Land consolidation: a comparative research between Europe and China. Land Use Policy 112, 105790.

Jiménez, F.M., 2024. On legal expertise. Am. J. Jurisprud.

Jung, H., Oh, C., Kang, J., Sohn, J., Song, K., Kim, J., Mortensen, D.R., 2024. Mitigating the linguistic gap with phonemic representations for robust multilingual language understanding arXiv.Org. https://doi.org/10.48550/arxiv.2402.14279 (abs/2402.14279.

Kalaiselvi, G., Monisha, N., Kasthuri, T., Devadharshini, P., 2024. CitizenConnectBot (CCB): the interactive guide to government schemes. Int. J. Sci. Res. Arch.

Kang, X., Qu, L., Soon, L.-K., Trakic, A., Zhuo, T.Y., Emerton, P.C., Grant, G., 2023. Can ChatGPT perform reasoning using the IRAC method in analyzing legal scenarios like a lawyer? In: Tomeh, N., Fujita, A., Sun, A., Wang, B., Tong, R., Cotterell, R. (Eds.), Findings of the Association for Computational Linguistics: EMNLP 2023. Association for Computational Linguistics, pp. 13900–13923.

Kaplan, A., Sayan, İ.U., Şahan, H., Begen, E., Bayrak, A.T., 2024. Response performance evaluations of ChatGPT models on large language model frameworks. In: 2024 32nd Signal Processing and Communications Applications Conference (SIU). IEEE, pp. 1–4.

Katz, D.M., Bommarito, M.J., Gao, S., Arredondo, P., 2024. GPT-4 passes the bar exam. Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. 382 (2274), 20230254. https://doi.org/10.1098/rsta.2023.0254.

Li, F., Hogg, D.C., Cohn, A.G., 2024. Advancing spatial reasoning in large language models: an in-depth evaluation and enhancement using the stepgame benchmark. arXiv.Org, abs/2401.03991.

Livermore, M.A., Chau, B.K., Rockmore, D.N., 2024. Computational legal studies come of age: Surveying law-as-code and law-as-data approaches and the role of LLMs. Annu. Rev. Law Soc. Sci. 20, 171–194. https://doi.org/10.1146/annurev-lawsocsci-121723-014645.

López Espejel, J., Ettifouri, E.H., Alassan, M.S.Y., Chouham, E.M., Dahhane, W., 2023. GPT-3.5, GPT-4, or BARD? Evaluating LLMs reasoning ability in zero-shot setting and performance boosting through prompts. Nat. Lang. Process. J. 5, 100032.

Luo, Z., Xie, Q., Ananiadou, S., 2023. ChatGPT as a factual inconsistency evaluator for text summarization. arXiv preprint arXiv:2303.15621.

Mahto, P., 2018. A critical review on land readjustment technique. J. Emerg. Technol. Innov. Res. 5 (12), 732–739.

Mooney, P., Cui, W., Guan, B., Juhász, L., 2023. Towards understanding the geospatial skills of ChatGPT: taking a geographic information systems (GIS) exam. In: 6th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery (GeoAI '23). ACM, Hamburg, Germany, pp. 1–10. November 13.

Nelson, J.W., 2023. Large Language Models and the Treaty Interpretation Game. In: Springer International Publishing, Switzerland.

Nestorović, Ž., Trifković, M., Kuburić, M., 2023. On the front management of land consolidation projects. In: LIMEN 2023/9 – Leadership, Innovation, Management and Economics: Integrated Politics of Research, pp. 81–84.

Nithya, M., Harini, S., Kavyadharshini, S., Srinidhi, K., 2024. AI-Driven Legal Automation to Enhance Legal Processes with Natural Language Processing. pp. 1246–53.

Ntihinyurwa, P.D., de Vries, W.T., 2020. Farmland fragmentation and defragmentation nexus: scoping the causes, impacts, and the conditions determining its management decisions. Ecol. Indic. 119.

Padiu, B., Iacob, R., Rebedea, T., Dascălu, M., 2024. To what extent have LLMs reshaped the legal domain so far? A scoping literature review. Information 15 (11), 662.

Peng, B., Li, C., He, P., Galley, M., Gao, J., 2023. Instruction Tuning with GPT-4. arXiv. Org, abs/2304.03277. https://doi.org/10.48550/arXiv.2304.03277.
Peng, L.-P., 2024. Innovative applications of artificial intelligence in agricultural land planning. Front. Sci. Eng. 4 (6), 75–81.
Prakken, H., 2024. On evaluating legal-reasoning capabilities of generative AI. In: Proceedings of the 24th Workshop on Computational Models of Natural Argument. Vol. 3769. CEUR Workshop Proceedings, pp. 100–12.
Rana, R., Bhambri, P., 2024. Harnessing AI and machine learning for enhanced geospatial analysis. Adv. Geospat. Technol. Book Ser. 27–72.
Rane, J., Kaya, Ö., Mallick, S.K., Rane, N., 2024. Artificial intelligence-powered spatial analysis and ChatGPT-driven interpretation of remote sensing and GIS data.
Rao, X., 2019. Land fragmentation with double bonuses – the case of Tanzanian agriculture. Eur. Rev. Agric. Econ. 46 (4), 609–635.
Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), OJ L 1689, 2024.
Richter, K.-F., Scheider, S., 2023. Current topics and challenges in geoAI. K. Intell. 37 (1), 1–6.
Ryan, F., Hardie, L., 2024. ChatGPT, i have a legal question? The impact of generative AI tools on law clinics and access to justice. Int. J. Clin. Leg. Educ. 31 (1), 166–205.
Shui, R., Cao, Y., Wang, X., Chua, T.-S., 2023. A comprehensive evaluation of large language models on legal judgment prediction. arXiv.Org, abs/2310.11761.
Spaić, B., Jovanović, M., 2024. Artificial reason and artificial intelligence: the legal reasoning capabilities of GPT-4. Ann. Fac. Law Belgrade Belgrade Law Rev. 72 (3), 383–422.
Surden, H., 2021. Machine learning and law: an overview. In: Woodrow Barfield (Ed.), Research Handbook on Big Data Law. Edward Elgar Publishing, pp. 171–184.
Tan, J., Westermann, H., Benyekhlef, K., 2023. ChatGPT as an artificial lawyer? In: L.K. Branting (Ed.), Proceedings of the ICAIL 2023 Workshop on Artificial Intelligence for Access to Justice (AI4AJ) (CEUR Workshop Proceedings). Vol. 3435.
Tugendhaft, A., Danis, M., Christofides, N., Kahn, K., Erzse, A., Gold, M.R., Twine, R., Khosa, A., Hofman, K., 2020. CHAT SA: modification of a public engagement tool for priority setting for a South African rural context. Int. J. Health Policy Manag.
Van, H.D., 2024. Law on ensuring the legitimate interests of the people in the land sector. Glob. Acad. J. Humanit. Soc. Sci. 6 (05), 217–223.
Wang, Y., Qian, W., Zhou, H., Chen, J., Tan, K., 2023. Exploring new frontiers of deep learning in legal practice: a case study of large language models. Int. J. Comput. Sci. Inf. Technol.
Weichbroth, P., 2025. AI and the Law: Evaluating ChatGPT's Performance in Legal Classification. arXiv preprint arXiv:2502.12193.
Wrzesniowska, L., 2023. Can AI make a case? AI vs. Lawyer in the Dutch legal context (SSRN Legal Studies Research Paper No. 23-31). Int. J. Law Ethics Technol.
Xu, L., Zhao, S., Lin, Q., Chen, L., Luo, Q., Wu, S., Du, Z., 2025. Evaluating large language models on geospatial tasks: a multiple geospatial task benchmarking study. Int. J. Digit. Earth 18 (1). https://doi.org/10.1080/17538947.2025.2480268.
Yang, X., Wang, Z., Wang, Q., Wei, K., Zhang, K., Shi, J., 2024. Large language models for automated Q&A involving legal documents: a survey on algorithms, frameworks, and applications. Int. J. Web Inf. Syst.
Zeng, Y., Brown, C., Byari, M., Raymond, J., Hotz, R., Rounsevell, M., 2024. Exploring the Opportunities and Challenges of Using Large Language Models to Represent Institutional Agency in Land Use Modelling.
Zheng, L., Su, L., Jin, S., 2023. Reducing land fragmentation to curb cropland abandonment: evidence from rural China. Can. J. Agric. Econ. 71 (3–4), 355–373.
Zhou, W., 2023. ChatGPT legal risk and regulation study. J. Educ. Humanit. Soc. Sci. 19, 104–106.