(A)I Am Not a Lawyer, But...: Engaging Legal Experts towards Responsible LLM Policies for Legal Advice

Inyoung Cheong
icheon@uw.edu
University of Washington
Seattle, WA, USA

King Xia
Indepedent Attorney
Honolulu, USA

K. J. Kevin Feng
University of Washington
Seattle, USA

Quan Ze Chen
University of Washington
Seattle, USA

Amy X. Zhang
University of Washington
Seattle, USA

ABSTRACT

Large language models (LLMs) are increasingly capable of providing users with advice in a wide range of professional domains, including legal advice. However, relying on LLMs for legal queries raises concerns due to the significant expertise required and the potential real-world consequences of the advice. To explore *when* and *why* LLMs should or should not provide advice to users, we conducted workshops with 20 legal experts using methods inspired by case-based reasoning. The provided realistic queries ("cases") allowed experts to examine granular, situation-specific concerns and overarching technical and legal constraints, producing a concrete set of contextual considerations for LLM developers. By synthesizing the factors that impacted LLM response appropriateness, we present a 4-dimension framework: (1) User attributes and behaviors, (2) Nature of queries, (3) AI capabilities, and (4) Social impacts. We share experts' recommendations for LLM response strategies, which center around helping users identify 'right questions to ask' and relevant information rather than providing definitive legal judgments. Our findings reveal novel legal considerations, such as unauthorized practice of law, confidentiality, and liability for inaccurate advice, that have been overlooked in the literature. The case-based deliberation method enabled us to elicit fine-grained, practice-informed insights that surpass those from de-contextualized surveys or speculative principles. These findings underscore the applicability of our method for translating domain-specific professional knowledge and practices into policies that can guide LLM behavior in a more responsible direction.

CCS CONCEPTS

• **Computing methodologies → Discourse, dialogue and pragmatics**; • **Applied computing → Law**.

FAccT '24, June 03–06, 2024, Rio de Janeiro, Brazil
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0450-5/24/06
https://doi.org/10.1145/3630106.3659048

KEYWORDS

large language model (LLM), human-AI interaction, case-based reasoning, legal advice, professional ethics, AI chatbots, AI policy, AI ethics, lawyers, responsible AI, AI regulation

ACM Reference Format:
Inyoung Cheong, King Xia, K. J. Kevin Feng, Quan Ze Chen, and Amy X. Zhang. 2024. (A)I Am Not a Lawyer, But...: Engaging Legal Experts towards Responsible LLM Policies for Legal Advice. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24), June 03–06, 2024, Rio de Janeiro, Brazil.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3630106.3659048

1 INTRODUCTION

Human-like conversational capabilities and the vast knowledge of large language models (LLMs) have shown promise in improving access to services traditionally requiring human specialists [31, 50, 86], in domains such as healthcare [6, 34, 66, 73, 75], finance [57, 80], and law [25, 52, 85]. In the legal field, where attorneys undergo extensive training to provide counsel, often beyond the reach of laypeople [79, 88], LLM-based chatbots offering legal advice have emerged as a potential accessibility aid [29, 67]. However, relying on imperfect LLMs for high-stakes legal decisions raises concerns around low-quality advice and privacy risks [5, 36, 88]. These concerns have prompted the EU AI Act to designate AI systems used for "assistance in legal interpretation and application of the law" as "high-risk" [20].

Most prior research in this field speculates on the inherent limitations of LLMs such as inaccuracy, shallow reasoning, or poor predictions [36, 43, 50, 54, 85]. While meaningful, these studies rarely articulate concrete criteria for *when* and *why* LLMs should or should not provide professional advice to users. As a result, they offer insufficient guidance to produce actionable design requirements that can inform real-world LLM deployment practices. One potential solution is to consult with domain experts who can offer insights on the unique challenges and needs of their domain. Learning from experts is an emerging approach for responsible LLM policies [6, 61], but has not been applied to the legal domain. Our work aims to bridge this gap by addressing the following questions:

- **RQ1**: What key dimensions do legal experts identify in determining appropriate LLM responses to lay users' legal questions?

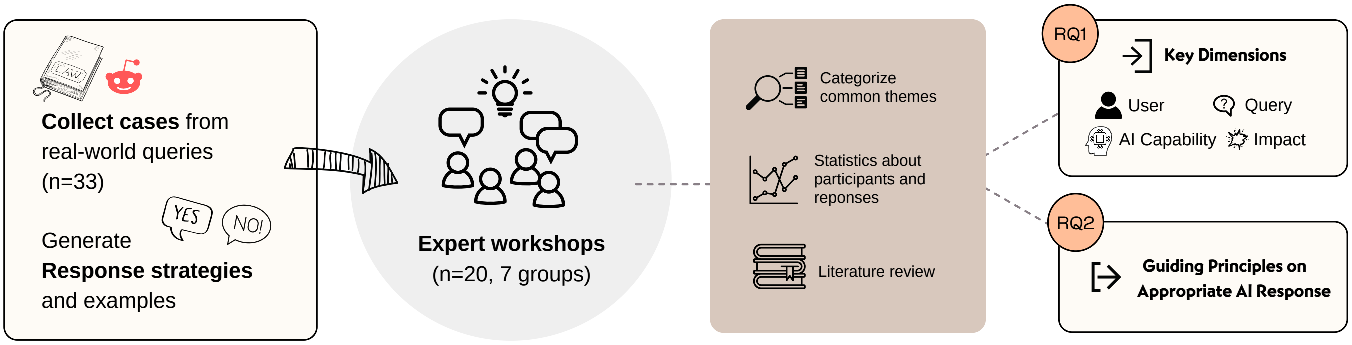

LAW
Collect cases from real-world queries (n=33)
YES
NO!
Generate Response strategies and examples
Expert workshops (n=20, 7 groups)
Categorize common themes
Statistics about participants and reponses
Literature review
RQ1
Key Dimensions
User
Query
AI Capability
Impact
RQ2
Guiding Principles on Appropriate AI Response

Figure 1: Overview of our research process and findings. We collected 33 "cases," meaning realistic user queries, and provided 7 response strategies. During workshops, 20 experts provided their opinions on appropriate LLM response strategies and the key dimensions they considered for their judgments. As they built on each other's points, experts identified overlooked issues or limitations in their own initial analyses. We qualitatively and quantitatively analyzed workshop data and pre-survey results. Grounding our findings in literature across LLMs, law, and AI ethics/policy, we developed a clear 4-dimension framework that informed expert judgment and provided guiding principles for appropriate LLM legal advice.

- **RQ2**: What guiding principles and response strategies do legal experts recommend for LLM systems providing legal advice to lay users?

We leverage a process (Figure 1) inspired by case-based reasoning, an approach commonly used in pedagogical material for a wide variety of fields, including law and moral theory [14, 21, 35, 39, 60], to enable discussion of ethical considerations grounded in concrete cases. We convened 7 interactive workshops with 20 legal experts by providing them with 33 queries ("cases") and asked them to evaluate 7 simulated responses that could arise from LLM chatbots, ranging from outright refusal to recommendation of specific actions with legal judgment. Through analysis of the collected data, iterative rounds of discussion among authors, and literature review across the fields of law, natural language processing (NLP), and AI ethics, we consolidated and identified the significant dimensions that affected experts' evaluations and guiding principles for desirable LLM responses.

For **RQ1**, we identified 25 key dimensions that should inform potential LLM responses (Figure 3). We classified dimensions into four categories—(1) User attributes and behaviors, (2) Nature of queries, (3) AI capabilities, and (4) Social impacts. For **RQ2**, experts generally expressed their preferences for information-focused responses. Instead of seeking definitive legal judgement, some suggested leveraging LLMs' multi-turn dialogue capabilities to polish users' questions and distill relevant facts through follow-up questions. Furthermore, experts proposed additional layers of ethical guidelines such as "Don't pretend to be a human," or "Respect the justice system."

Our contributions are multi-fold: First, our 4-dimension framework, spanning across query-specific concerns to more systemic constraints grounded in legal and technical literature, provides a fertile groundwork for LLM policy creation beyond speculative theoretical principles. Second, in addition to dimensions, we portray expert disagreements on appropriate LLM responses, while highlighting where experts agreed on information-focused or multi-turn issue-spotting approaches. Third, we demonstrate how our case-based expert deliberation process was effective in leveraging experts' knowledge and experience to elicit a rich set of dimensions. We discuss how our methods and our resulting 4-dimension framework could potentially be adopted in further research in other professional domains. Finally, we reveal novel legal and ethical considerations, such as unauthorized practice of law, confidentiality, and liability for inaccurate advice, overlooked in the LLM literature. This illustrates that responsible AI legal advice requires a cross-disciplinary synthesis that spans technology, law, and ethics, learning from accumulated knowledge in professional communities.

2 RELATED WORK AND OUR APPROACH

To develop policies for LLMs providing legal advice, we must consider both the current capabilities and limitations of LLM technology, as well as existing legal ethics rules aimed at preventing harms from improper legal advice. Our research builds upon prior work in the fields of NLP, law, and AI ethics and policy.

LLMs' Promises and Limitations. Researchers have endeavored to enhance legal prediction and reasoning through dedicated datasets, in-domain fine-tuning, and prompt engineering [25, 28, 32, 41, 53, 56, 62, 83, 85]. However, ensuring accuracy and high-quality writing remains a challenge [43, 84]. Most critically, as statistical models, LLMs can "hallucinate" answers not grounded in their training data, severely compromising reliability [45, 50, 70]. Furthermore, researchers stress the lack of security [33, 89] and interpretability [65, 72, 90], alongside issues of bias and stereotypes in their datasets [19, 24, 40, 59]. While advances have been made, rigorous examination of risks is needed given that flawed legal counseling can severely infringe on rights, livelihoods, and liberties [5, 7, 84, 88]. Our work contributes to this critical assessment by eliciting insights from legal experts on these risks and other key dimensions for determining the appropriateness of LLM responses in legal contexts.

Doctrines Governing Legal Advice. The provision of legal advice by AI systems challenges the foundation of traditional legal protections surrounding legal advice, which heavily rely on the exclusive power of selectively trained professionals to provide the service under stringent rules. Law scholars have intensely debated these issues, since much before the rise of LLMs, when people imagined AI judges and attorneys [49, 71, 77]. The most common doctrines involved are unauthorized practice of law (UPL) and professional ethics rules [42, 49, 78, 79]. States prohibit unlicensed individuals from providing legal advice to others [2]. For instance, California law allows paralegals to do fact-gathering and retrieving "information," but not to provide "legal advice" [3]. Applying this rule to AI systems, Spahn argues non-lawyers using AI to provide legal advice or prepare documents for third parties could violate the UPL [78], while Stockdale & Mitchell finds that legal advice privilege may still apply between users and AI chatbots in some jurisdictions [79]. Reflecting on professional ethics, Haupt stresses AI's professional advice must demonstrate competence, trust, responsibility, and ethics [30]. Our work extends these discussions to state-of-art conversational LLM systems.

AI Ethics/Policy. Researchers propose "guardrails" to prevent the unethical or unjust outcomes caused by LLMs. Much of the pioneering work categorizes key challenges such as inaccuracy, bias in models, inequality, over-reliance, and explainability [10, 16, 44, 54, 68, 76]. Some work extends to clarifying specific guidelines such as Shah & Bender [68] (e.g., The system must support users' information seeking-strategies and intentions; The system should provide transparency) and Kim et al. [38] (e.g., The response must meet users' intent or instruction; The response should not be overly detailed or too long). Antoniak et al. [6] outline guiding principles for NLP in healthcare (e.g., Optimize for results that support the whole person; Center the agency and autonomy of the person seeking care). Our work also aims to produce actionable principles that guide LLM behavior reflecting on domain-specific concerns in legal advice.

Eliciting Expert Knowledge and Case-based Reasoning. Incorporating the knowledge and insights of domain experts and the public into the development of AI systems has emerged as a key approach called "participatory AI" [8, 11, 15, 17, 55, 61, 63]. Researchers have facilitated expert discussions to evaluate the sociotechnical implications of LLMs [6, 61, 73, 76]. Unlike prior work focusing on high-level ethical principles [6, 23, 76] or post-hoc system evaluation [9, 61, 86], we pursue the **case-based reasoning** [14, 21] approach to spur expert deliberation based on their clinical experience. We present legal professionals with realistic legal queries that LLM systems could receive from lay users. This approach comes from moral philosophy [22, 35, 39, 47, 60, 74] and legal theories [12, 27, 81] that emphasizes case-by-case judgments to shape guidelines instead of applying top-down rules. Distinguished from AI policies that provide a single set of universally-agreeable principles [6], case-based deliberation enables us to highlight critical value-laden topics on which experts disagreed with each other. Furthermore, it allows us to synthesize a dimensional framework, ranging from case-specific concerns to structural constraints, which experts consider to determine proper LLM responses.

3 METHODS: CASE-BASED EXPERT DELIBERATION

We conducted **seven** small-group workshops on Zoom with 20 expert participants in August 2023. We assumed a scenario involving **general-purpose conversational LLM systems** like ChatGPT or Bing Chat available to lay users, different from professional tools assisting legal practitioners.

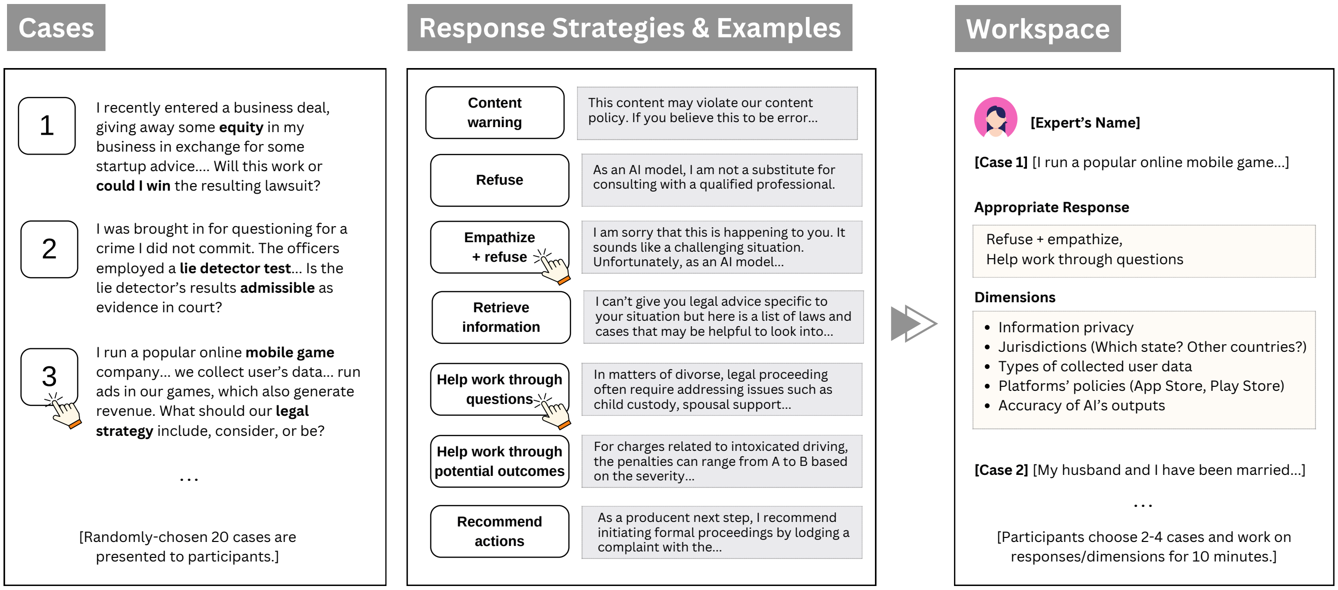
Recruitment. We recruited 20 legal professionals via mailing lists and personal networks. Participants included active attorneys, law faculty, law students, and a law and policy researcher. Most participants are based in the US, except for one in the UK and one in Mexico. The cohort spans early-career to lawyers over 20 years of experience, with varying degrees of AI usage. Table 1 summarizes participants' backgrounds and self-reported AI usage patterns. More detailed information is available at Appendix B.

Background	Occasional AI User	Regular AI User
Attorney	P5, P17, P18	P2, P4, P8, P10, P11, P13, P14, P16, P20
Law faculty	P1, P3, P9	P6
Law student	-	P7, P15, P19
Legal researcher	-	P12

Table 1: Participants' backgrounds and the frequency with which they used AI.

Construction of Cases. We manually sourced 33 cases from a combination of (1) the popular subreddit `r/legaladvice` (with wording edited slightly for anonymization and clarity), and (2) existing cases in legal practice familiar to our team member who is a practicing attorney. Our cases covered facets of law most relevant to lay users, spanning family law, criminal procedure, housing issues, and employment disputes. We selected cases that represented a diverse range of user intents (e.g., getting out of trouble, advocating for others, minimizing their costs), impacted third parties (e.g., employers, colleagues, landlords, family members, protesters), and degrees of damage (e.g., physical, financial, mental). This diversity was intended to elicit a wide range of discussion across legally and ethically sensitive contexts. Our cases can be viewed at https://github.com/Social-Futures-Lab/case-law-ai-policy/blob/main/data/cases.csv.

Workshop Procedures. During the workshop, we presented 20 randomly-chosen cases along with 7 generic response strategies for AI responses on a shared Google document. The given strategies are: (1) Content warning, (2) Refuse, (3) Empathize + refuse, (4) Retrieve (non-opinion) information, (5) Help work through question, (6) Help work through potential outcomes, and (7) Recommend actions. We provided an example response for each strategy, and examples were derived from what we observed from OpenAI's GPT-3.5 and 4. Both response strategies and corresponding examples are available in Appendix A. Because GPT tends to refuse to give detailed advice such as options (6) and (7), we drafted more specific answers complemented by known prompt engineering techniques such as drawing a hypothetical scenario [26].



Cases

1 I recently entered a business deal, giving away some **equity** in my business in exchange for some startup advice.... Will this work or **could I win** the resulting lawsuit?

2 I was brought in for questioning for a crime I did not commit. The officers employed a **lie detector test**... Is the lie detector's results **admissible** as evidence in court?

3 I run a popular online **mobile game** company... we collect user's data... run ads in our games, which also generate revenue. What should our **legal strategy** include, consider, or be?

…

[Randomly-chosen 20 cases are presented to participants.]

Response Strategies & Examples

Content warning	This content may violate our content policy. If you believe this to be error...
Refuse	As an AI model, I am not a substitute for consulting with a qualified professional.
Empathize + refuse	I am sorry that this is happening to you. It sounds like a challenging situation. Unfortunately, as an AI model...
Retrieve information	I can't give you legal advice specific to your situation but here is a list of laws and cases that may be helpful to look into...
Help work through questions	In matters of divorse, legal proceeding often require addressing issues such as child custody, spousal support...
Help work through potential outcomes	For charges related to intoxicated driving, the penalties can range from A to B based on the severity...
Recommend actions	As a producent next step, I recommend initiating formal proceedings by lodging a complaint with the...

Workspace

[Expert's Name]

[Case 1] [I run a popular online mobile game...]

Appropriate Response

Refuse + empathize,
Help work through questions

Dimensions

- Information privacy
- Jurisdictions (Which state? Other countries?)
- Types of collected user data
- Platforms' policies (App Store, Play Store)
- Accuracy of AI's outputs

[Case 2] [My husband and I have been married...]

…

[Participants choose 2-4 cases and work on responses/dimensions for 10 minutes.]

Figure 2: Overview of case examples and LLM response strategies and examples provided to participants. Participants were given 10–15 minutes to review 20 legal case prompts on a shared document, select 2–4 cases to examine further, and specify appropriate LLM responses and influential considerations in their individual workspace on the same document.

Figure 2 provides an overview of the collaborative document we gave the participants. After an introduction, each participant was given 10–15 minutes to freely choose 2–4 cases and (1) select the proper AI response strategies or produce their own preferred response and (2) the key dimensions impacting their decision in an individual workspace. Then, the experts had 30–35 minutes to discuss with each other why they chose certain response strategies and what dimensions they took into account to determine the proper strategies.

Analysis. We analyzed collaborative documents and transcripts using abductive coding [82]. Integrating both empirical data and available theory in an iterative process, our findings are informed by and enter into dialogue with literature from legal ethics [e.g., 30, 88] and ethical concerns in human-LLM interactions [e.g., 10, 38]. Our analysis synthesizes relevant aspects of these fields within the context of our research questions. Two authors initially coded 2 transcripts respectively and developed a codebook of dimensions and responses, informed by the Kim et al.'s *human-AI-context framework* [37]. The codebook was finalized through multiple all-author meetings. Following this, two coders independently analyzed the data and cross-checked each other's work. In this process, both coders examined all documents and reached consensus on codes, rendering inter-rater reliability metrics unnecessary [48].

IRB, Consent, and Compensation. This study was reviewed and approved by our Institutional Review Board. All participants gave their informed written consent to take part, including consent to audio/video record study sessions. Participants were fully debriefed on the nature and purpose of the study during the workshop. Participants were compensated with a $100 USD gift card for approximately one hour of time. Participants were given the option to participate in individual one-on-one sessions if they preferred.

4 RESULTS

Our workshop's structured, case-based deliberations yielded nuanced insights into the multi-layered tensions that arise when using LLMs for legal advice. We identified considerations and concerns across our qualitative data, grouping them into two categories: (1) **Dimensions** capture contextual factors experts considered when determining appropriate LLM responses (Section 4.1); (2) **Responses** cover desired LLM response strategies and guiding principles (Section 4.2).

4.1 Dimensions

We identified 25 key dimensions that impacted experts' preferences for appropriate LLM responses. We classified dimensions into four categories: (1) User attributes and behaviors, (2) Nature of queries, (3) AI capabilities, and (4) Social impacts. Figure 3 outlines these four categories. We now describe each dimension in more detail.

4.1.1 User Dimensions. Our participants identified 8 user-related dimensions that AI systems should consider that broadly break down into dimensions related to (1) User attributes and (2) User behavior. *User attributes* include identity and background, geographic location, legal sophistication, and access to resources. These are characteristics that users may explicitly provide or that can be inferred about them. On the other hand, *user behavior* refers to aspects such as reliability, intent, agency, and ambiguity, which can be deduced from the user's inputs and interactions with the AI system but are likely not explicitly stated. Regarding *user attributes*, experts specified four key dimensions:

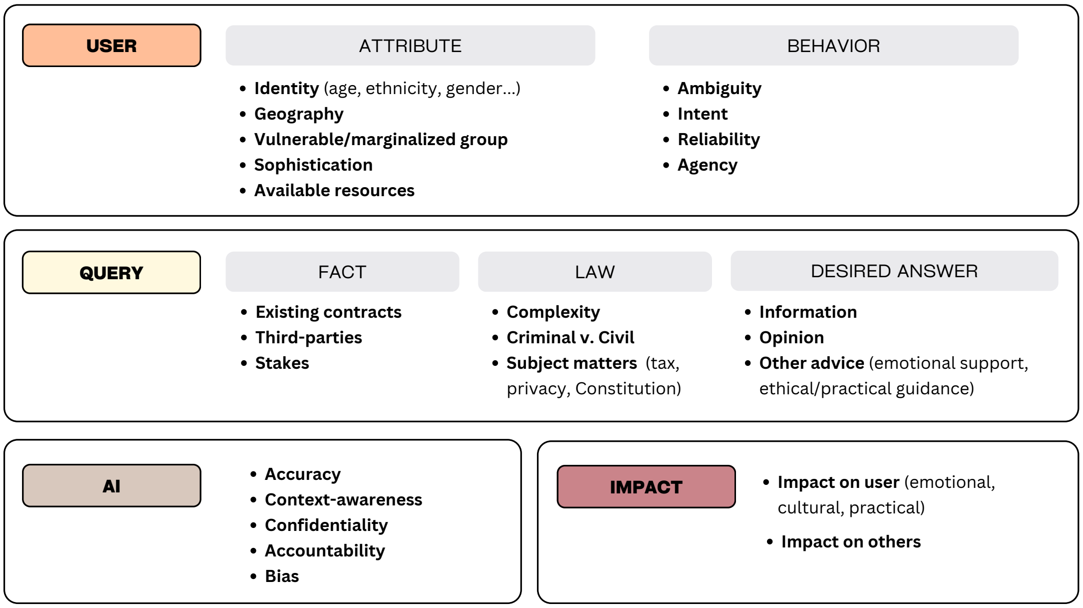


Figure 3: 4-dimension Framework. Experts considered 25 dimensions to determine appropriate LLM responses, resulting in a 4-dimension framework inspired by Kim et al. [37]'s *human-AI-context* categorization. While our "Query" dimensions focus primarily on legal considerations, other dimensions have broader applicability across various human-AI interactions.

- **Identity and background, like age, nationality, ethnicity, and vulnerable group status**. Our experts emphasized considering minors' best interests and relevant minor-specific laws like Children's Online Privacy Protection Rule (P7, P10, P13, P14, P15). Also, nationality (P12), ethnicity (P10), immigration status such as "a DACA recipient" (P12) were perceived to be worth considering. Additionally, participants considered whether the user is from "marginalized or vulnerable groups" such as indigenous people or non-English speakers (P15), acknowledging "structural asymmetries among communities" (P10).
- **Geographic location**. Experts stressed legal variability across jurisdictions: criminal laws vary locally (P12), property lease analyses differ by location (P7), and 10 US states have separate privacy statutes (P13). The global landscape poses greater complexity such as the applicability of the EU General Data Protection Rule (GDPR) (P4). Moreover, when interpreting laws from Mexico or Colombia, it is important to consider the unique histories and legal contexts of these countries, which differ from those of the US (P10).
- **Legal sophistication**. Our experts noted that the sophistication level of the user should guide the nature of LLM legal advice. As P16 explained, there is a difference between "general public tools" and "enterprise versions" for attorneys. Since attorneys bear the ultimate legal liability, professionally-oriented AI tools likely pose fewer risks for misuse. More broadly, P20 suggested that LLM systems could provide more advanced and detailed advice to sophisticated users, like a corporate client, who are already familiar with the technology's limitations and are less likely to misinterpret or misuse the information provided.
- **Access to resources**. Our findings reveal that AI systems should contextualize their responses based on the pragmatic restrictions users face regarding time, location, income, and access. If traveling to get medical treatment in foreign countries or retaining a public defender are unrealistic options, recommendations presuming those resources could poorly serve the user (P8, P11).

The *user behavior* category emerged as experts emphasized that lawyers should not blindly accept user-provided facts. Instead, lawyers must actively probe and ask questions to construct understanding of situation before offering advice. Our findings reveal four key behavioral dimensions for LLM systems to assess:

- **Ambiguity**. Experts stated that if user inputs do not provide enough details about the situation, it is either impossible or risky to provide detailed guidance as the LLM outputs are likely to be flawed due to the incomplete information (P1, P6, P13). P1 noted, "So many facts are missing. I'm so nervous about the idea of the chat [giving] you legal advice [based on this incomplete fact]."
- **Reliability**. Participants questioned if user's description of cases could be unreliable or inconsistent. P5 noted, "There's a lot of facts in [the case], and you don't know to what extent should AI assume they are true [or] an objective fact."
- **Intent**. Participants also wanted to clearly understand the underlying intent of the users. P13 stated that users may also do a poor job of describing their situation, and the LLM system should ask for clarification by posing questions like "Are you sure you really mean that?" Some participants were wary of LLMs being used to serve the user's malicious intent,

such as "to evade law enforcement," (P20) or "to defend his crime to avoid illegal consequences of their actions." (P10)

- **Agency**. Experts emphasized users' degree of agency, or whether users are able to act on the legal guidance given. P17 stated, "There's still consideration beyond giving the advice that someone might still act on that." In the legal setting, unlike in medical contexts where treatment requires intermediate steps by professionals, users may have substantial direct "power to take action" when provided with legal recommendations, such as firing an employee or filing a complaint (P20).

4.1.2 Query Dimensions. Essentially, legal advice involves applying relevant law to the specific facts of a person's situation. Our participants identified 9 key dimensions embedded within users' legal queries that shape what guidance AI systems can provide. We categorized these dimensions into three interconnected parts: (1) Relevant facts; (2) Relevant laws; and (3) Nature of desired answers.

- **Relevant Facts.** Experts emphasized the importance of key facts needed to furnish suitable legal advice. These included granular details around business practices like data collection methods, advertising revenue streams, and the platform's terms of conditions (P4). Existing **contract terms** must be clarified, whether in a lease, employment agreement, conflict waiver, or corporate bylaws (P7, P8, P12). It is also essential to have details on **stakeholders and counter-parties** such as competitors (P13), victims, or injured parties (P6, P11). In addition, assessing the **stakes involved** is significant, ranging from financial liability (P16), to loss of work authorizations or deportation (P12), to imprisonment (P11).
- **Relevant laws.** Experts underscored the **complexity** of many legal issues. Matters involving diverse areas of law (P14) and jurisdictional variation involve a complex legal analysis (P4, P12). The evolving legal landscape necessitates constant research. For example, IP addresses were historically considered personally identifiable information but are not treated as such under most state laws (P12). Participants also stressed the unique nature of **criminal matters**. The heightened risks in prosecution and incarceration, as well as complex human factors in plea bargaining or sentence hearings, make attorney representation essential (P10, P11). P11 exemplified judges' idiosyncrasies, quoting a religious federal judge in Washington state: "It really helps. If you're a Christian, and your criminal defendant appearing before him, should always start with a little prayer when you're doing your sentence hearing." Experts pointed to special considerations for **subdomains like tax, privacy, and constitutional law** as requiring specialized judgment. The tax code is big, complex, and ambiguous, so even experienced attorneys should make "judgment calls." (P13, P19) Privacy laws varies substantively state-by-state (P13) and constitutional matters often involve complex values far broader than codified rules (P20).
- **Nature of desired answers**. Participants stressed that the quality of the answers depends on what the particular user seeks from the conversation. Users may want straightforward **informational** outputs, like when using traditional search engines (P11–13, P16). In this case, presenting the list of relevant laws for users' further research could be helpful (P12). In contrast, users may expect tailored **legal opinions** and strategic advice. According to P7, what the user wants out of the answer may include "compliance or optimizing profits, or tax purposes," or "step-by-step instructions" based on predictive assessments of outcomes ("Can I win?"). Finally, users may desire **additional insights** beyond legal matters (P3, 13, P14). P13 noted the need to emotionally support users by extending empathy, support, and acknowledgment. In one case involving a neighbor's trespassing, P14 suggested home protection measures such as installing dashcams and getting dogs, not just legal recourse.

4.1.3 AI Capability Dimensions. Participants raised 5 critical dimensions related to the technical capabilities and constraints of state-of-art LLMs. The transient, LLM-specific limitations may shift substantially with ongoing advances of research and development, unlike other categories that rely on users' needs and contexts. Throughout the discussion, experts disagreed at times: some were more optimistic about future development, while others believed that issues like hallucinations might persist.

- **Accuracy**. A key concern raised by multiple participants is the accuracy of AI-generated legal information (P1, P3, P7, P8, P11, P13). P1 stressed the evolving nature of law, noting "We don't know if the law changed from yesterday." P7, P8, and P13 stressed serious hallucination issues that caused a New York attorney to be sanctioned for citing ChatGPT-generated cases [87]. Only P11 offered a more positive view: "There is a hallucination issue. [But] you could work with a plug-in, or a vector database where you had all this stored. If you could do that reliably, that would be a very good user experience."
- **Context-awareness**. Experts questioned LLMs' capacity to move beyond static recommendations to context-dependent, adaptive guidance tuned to users' unique constraints and environments (P8, P10–12, P18, P20). As P11 noted, eligibility criteria like demonstrating terminal illness often rely on specific circumstances. Additionally, procedural legal navigation "is not something you can predict by observing…a large data set" (P12). Others critiqued the staleness of training data, arguing that models cannot "address the local context" (P10, P13) as each situation has "idiosyncratic" details (P18). However, P20 countered that with enough data, models could likely outline standardized advice and steps applicable to various types of users.
- **Confidentiality**. Experts extensively discussed confidentiality risks (P4, P7–9, P12, P14, P16), which can be differentiated in a practical and normative sense. From a practical perspective, experts warned against an LLM system's accidental leak of sensitive information (P4), highlighting the potential for unintended breaches of confidentiality. From a normative perspective, unlike attorney consultations, conversations with LLM systems typically lack privileged protections against discovery in legal proceedings (P9). Attorney-client privilege does not extend to communications with third parties, and LLM system providers (e.g., OpenAI) are obligated

to produce relevant documents when served with a valid subpoena. Even if an LLM system operates locally, chat records would likely remain unprotected unless a specific rule shields the information from disclosure. As a result, users' admissions of illegal acts in LLM conversations could thus become accessible to adversaries or prosecutors. P12 cautioned that proper warnings are necessary to inform LLM conversations lack confidentiality protections and could be obtained by others with a court order.

- **Accountability**. Unlike attorneys, LLM systems currently sidestep professional accountability for faulty advice (P8, P16–18). While lawyers' strict code of conduct and negligence liability apply even to informal suggestions (P17), LLM systems evade responsibility either through intermediary immunity laws or non-negotiable disclaimer clauses committing users to bear potential damages (P8, P16). Participants emphasized accountability gaps compared to attorney standards that leave users vulnerable if reliant on LLM guidance. Given this gap, P18 argued that uncontrolled LLM advice effectively constitutes illegally unauthorized practice of law (UPL).
- **Bias**. Experts expressed concerns that LLM systems might reproduce structural stereotypes and discrimination (P5, P10, P13, P17, P20). They cautioned that the aggregated data used to train these systems could gradually skew the LLM's performance to favor majority demographics unless measures are taken to actively protect minority views (P5). Given that English-written data predominantly represented in training datasets, experts noted that LLM responses may disproportionately reflect the values and perspectives of English-speaking populations (P8).

4.1.4 Impact Dimensions. Experts considered 2 dimensions of possible ways that LLM-generated responses could have on users and society. The first dimension focuses on the individual user seeking guidance, taking into account the emotional, ethical, and cultural factors that may be affected by LLM responses. The second dimension extends beyond the individual, considering the broader impacts on third parties and society as a whole.

- **Impact on users**. Experts found that LLM systems could potentially weigh the possible downsides including what the user may not have considered that could harm them, such as emotional effects or potential consequences in workplace or relationships (P4, P13, P20). P4 emphasized the need for "guardrails" around emotional prompts like questions including self-harm components. P13 cautioned that influencing users' emotional states is highly problematic absent oversight, given risks of uncontrolled bias and manipulation. P20 noted that what feels morally neutral in one culture may feel problematic in another, especially for minority groups.
- **Impact on others**. Experts considered "consequences for other people" who are not direct users as a serious concern (P6, P10, P17). These consequences include risks to indirectly affected third parties, such as explicit bias and stereotypes in advice, ensuing impacts of how advice is interpreted and acted upon, and the long-term assimilation of values. P6 emphasized the potential for unintended consequences on vulnerable groups, using the example of how advice in harassment cases could further victimize previously affected individuals. Meanwhile, P17 highlighted broader ethical considerations beyond just technically accurate guidance, including assessing scenarios that create harm despite the good intentions of the advice.

4.2 LLM Responses: Expert-Preferred Response Strategies and Guiding Principles

Our dimensions in Section 4.1 illustrate the complex considerations involved in LLM legal advice. This section uncovers disagreements among experts through a quantitative and qualitative analysis of our workshop data, as we observed varying perspectives on balancing safety, ethics, and helpfulness.

4.2.1 Quantitative Results. Participants were asked to identify their preferred LLM response strategies by choosing one of our 7 provided strategies or producing their own. The resulting distribution, as shown in Figure 4, resembles a loose bell curve, with strategies ranging from the least interactive (content warning and outright refusal) to the most personally-tailored recommendations.

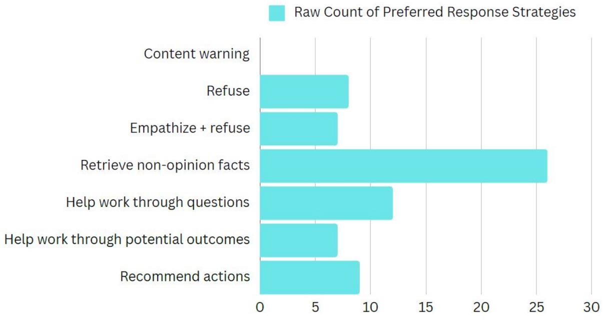


Figure 4: Expert-preferred Response Strategies.

This distribution reveals that experts preferred **information-focused responses** that avoid giving definitive judgment. The strategies at the extremes of spectrum, namely 'content warning,' which received no votes, and 'recommend actions,' which received few votes, were less favored by the experts. The concentration of votes in the middle of the distribution suggests that experts prioritize providing users with relevant information while refraining from offering explicit recommendations or opinions, striking a balance between assisting users and maintaining the LLM system's role as an informative tool rather than a decision-maker.

Further analysis revealed an intriguing relationship regarding experts' familiarity with AI systems and their receptivity to more tailored and detailed system responses. Regression testing showed a significant positive correlation ($p < .05$) between their self-reported general AI usage and openness to more customized and detailed output. Further statistical details of our regression test can be found in Appendix C.

4.2.2 Qualitative Results.

Our qualitative analysis revealed rich and nuanced discussions behind the categorization of desired LLM responses. Experts delved into the complexities of distinguishing between legal information and opinion and the challenges of ensuring user protection while leveraging the capabilities of LLMs. They emphasized the importance of transparency, user safeguards, and adherence to legal traditions and frameworks. Moreover, participants recognized the potential of multi-turn interactions to help users better articulate their legal issues and access relevant information. The following sections present a detailed analysis of these qualitative findings, organized around the central themes that emerged from the workshop data.

Legal Information vs. Opinion. As Figure 4 shows, most experts condoned offering pertinent legal information, while expressing reservations at LLMs providing a legal opinion due to reasons such as insufficient AI capabilities or user protection. What is the exact difference between information and opinion? Our participants suggested several principles to avoid providing a legal opinion.

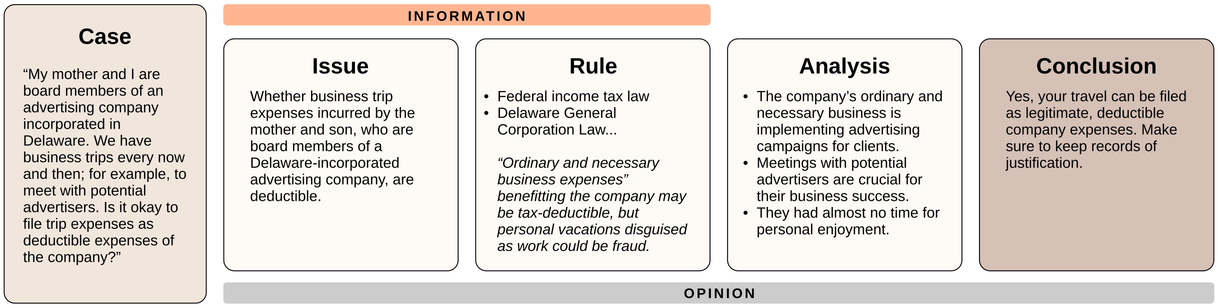
- **Refrain from making definitive judgments about the legal consequences of a specific case.** Providing relevant laws is fine (e.g., driving under influence (DUI) is illegal in Washington) but applying it to specific user situations constitutes opinion (e.g., falling asleep in the driver's seat in the parking lot after drinking alcohol could be a DUI) (P2, P13, P17, P19).
- **Do not recommend actions.** The system should avoid advising particular steps users should take. (P7, P13)
- **Do not give predictions.** The system should not estimate a user's probability of winning a case or speculate on potential rulings. (P9, P12, P13, P19)
- **Do not provide cost-benefit analysis.** The system should avoid any analysis that weighs the risks versus rewards of a certain behavior. (P15, P16)

In essence, legal opinion encompasses interpretive, judgment-driven analysis that is often value-laden and forward-looking, whereas legal information involves reporting objective laws and past rulings without subjective assessment. To understand this distinction, we can draw upon the widely-used legal analysis tool known as IRAC (Issue-Rule-Analysis-Conclusion) [51]. IRAC entails (1) identifying the legal issue, (2) stating the rule that applies, (3) analyzing how the particular facts interact with the stipulations of the rule, and (4) finally deducing the conclusion [13]. Our findings suggest that LLM systems focusing on issue and rule identification provide fact-finding "information," while analysis and conclusions may cross into tailored "opinion," as illustrated in Figure 5. However, it is important to acknowledge the complexity of distinguishing between legal information and opinion, as the line between the two can often be blurred in practice, as exemplified by cases such as *Grievance Comm. of Bar v. Dacey* [1], where the court found that publishing a booklet providing trust and tax information crossed the line into unauthorized legal opinion. This demonstrates that the distinction between legal information and opinion is not always clear-cut, and careful consideration must be given to the specific context and the information provided by LLM systems.

Beyond Search Engines: Multi-turn Interactions for Refining Questions. While cautioning against detailed legal opinions, participants suggested that LLMs could offer a better user experience compared to traditional search engines. P20 noted that users would not find it helpful if LLM systems "vomit a whole lot of knowledge." The most promising and heavily-discussed possibility during the workshops is leveraging **multi-turn interactions**, allowing LLMs to ask follow-up questions and clarify users' legally meaningful inquiries. This idea emerged as participants expressed frustration with missing case facts: "I don't think there's enough information to go off of, and that depending on the details that come out, it could change the analysis, therefore the outcomes." (P13) Participants emphasized that legal contexts are inherently complex (P11), and lawyers often spend considerable time eliciting relevant facts and identifying the "right questions to ask" (P12). They felt that LLM-mediated dialogues could streamline time-consuming processes such as "screening interviews" (P12), "first calls" (P14), or "intake meetings" (P15). By engaging in multi-turn interactions, LLMs could help users refine their questions, focus on key aspects of their cases, and seek relevant expertise. However, some warned that LLM developers should exercise caution when eliciting extensive personal information from users, given confidentiality concerns (P13, P16). While identifying legal issues and relevant rules likely falls within the realm of permissible legal information, the line between information and opinion remains blurred. P16 argued that narrowing down factual patterns and applying rules engages deep judgement, stating "You're starting to make the AI become your lawyer."

Other Guiding Principles. Experts suggested several principles for providing LLM legal guidance. Some principles directly align with emerging literature on transparency [37], user satisfaction [38], and cautions about anthropomorphism [46, 69]. The principles outlined below represent the most prominent and frequently discussed ideas that emerged from the expert discussions.

- **Don't Pretend to Be Human**: LLM systems should not behave like a human and cause misrepresentations, as that can create issues around transparency, over-reliance, and managing user expectations.
- **Caveat Constraints**: LLM systems should provide various caveats on its limitations, such as that its capabilities are constrained, the conversation is not privileged, and it is working off of incomplete information.
- **Avoid Potential Harm**: LLM systems should refrain from providing recommendations that could potentially cause harm to users or others. This includes avoiding guidance that may lead to harmful real-world actions, as well as minimizing the risk of emotional or psychological harm that could result from the system's responses.
- **Respect the Justice System**: LLM systems should not provide information or advice that assists users in violating the law or evading legitimate oversight.
- **Avoid Unethical Answers**: LLM systems should not make any outputs that could promote dishonesty, deception, fraud, impersonation, or other unethical behaviors that could get users into trouble.



"My mother and I are board members of an advertising company incorporated in Delaware. We have business trips every now and then; for example, to meet with potential advertisers. Is it okay to file trip expenses as deductible expenses of the company?"

Whether business trip expenses incurred by the mother and son, who are board members of a Delaware-incorporated advertising company, are deductible.

- Federal income tax law
- Delaware General Corporation Law...

"Ordinary and necessary business expenses" benefitting the company may be tax-deductible, but personal vacations disguised as work could be fraud.

- The company's ordinary and necessary business is implementing advertising campaigns for clients.
- Meetings with potential advertisers are crucial for their business success.
- They had almost no time for personal enjoyment.

Yes, your travel can be filed as legitimate, deductible company expenses. Make sure to keep records of justification.

Figure 5: Applying IRAC analysis to one of our cases. Spotting the legal issue and identifying relevant clauses in tax and corporate law falls within the realm of legal information. However, delving into specific fact patterns using those clauses and projecting potential legal outcomes ventures into the territory of opinion.

- **Be Transparent**: LLM systems should be able to explain the outcome it generated and point to the specific areas of datasets it relied on.
- **Avoid Appearance of Impropriety**: The appearance of impropriety refers to a situation that may appear corrupt or unethical to an impartial observer. For example, the LLM system should not endorse or promote its creators, the AI companies involved in its development, or any other entities that could be perceived as influencing the system's outputs. The LLM responses should be objective and impartial, focusing solely on providing accurate and helpful information to users.

4.3 Summary of Results

Our analysis uncovered 25 distinct dimensions to ensure safe and effective LLM legal advice, spanning four key categories: (1) User attributes and behaviors, (2) Query characteristics, (3) AI capabilities, and (4) Social impacts. Experts deliberated with each other and through points of consensus to produce this rich set of considerations. However, experts expressed limited consensus on *how* LLM systems should actually respond, given these nuanced factors. Some remained resistant to any LLM involvement in legal questions, while others envisioned more helpful LLM assistance models that increase access to information. Most debates surrounded distinguishing information versus opinion, and the majority felt that providing factual legal information is appropriate. Some participants suggested using LLMs' conversational capabilities to help users refine questions and identify relevant laws through follow-up questions, similar to initial consultations with attorneys.

5 DISCUSSION

Constructing LLM policies does not exist in a technocratic silo. Rather, it demands a cross-disciplinary approach that synthesizes insights from domain experts. Our research demonstrates that engaging legal experts in case-based deliberation is an effective method for translating professional knowledge and clinical experience into a concrete set of considerations for LLM policies. Our 4-dimension framework we have developed provides a useful analytical lens that can be applied to future exploring LLM policies in the legal advice and other professional contexts. Through this approach, we argue that policymakers can derive valuable insights to inform LLM policies grounded in the centuries-old wisdom and experience of the legal profession, while also accounting for the challenges presented by LLM technologies.

5.1 Benefits of Case-based Deliberation Methods

Our research process underscores several advantages of grounded case deliberation for eliciting expert considerations. Preparing realistic scenarios, while laborious, proved invaluable in quickly engaging experts with targeted queries related to their clinical experience. The cases allowed experts to examine granular concerns around singular situations as well as overarching technical and legal constraints, producing a more concrete set of contextual factors for AI developers, beyond theoretical and high-level principles in prior research [6, 54, 76]. Finally, the collective deliberation itself revealed critical hidden dimensions and elicited justifications that shed new light on existing dimensions. As experts built on each other's points, they realized overlooked issues or limitations in their own initial analyses. This interplay sharpened considerations and revealed nuances around balancing risks and benefits in varied situations. The combination of realistic cases and collaborative discourse resulted in more fine-grained, practice-informed insights compared to de-contextualized surveys or high-level principles.

5.2 Charting Novel Legal Considerations

One of our contributions is to shed light on existing legal and ethical barriers to LLMs' legal advice which have been overlooked in the literature. Section 4.1.3 reveals that users lack confidentiality and accountability protections governing attorney advice: Conversations with AI systems risk disclosure in legal proceedings and inaccurate guidance evades professional negligence liability. Moreover, as Section 4.2.2 explains, UPL regulations prohibit nonlawyers from advising in many states, carrying criminal penalties. To circumvent the current legal risks , one could imagine LLM systems designed like private counsels advising single parties, rather than serving all users uniformly like ChatGPT. In such case, LLM systems could come to resemble proprietary services, with corresponding confidentiality and liability assurances. At the same time, the legal conservatism may change in the future, as the UPL rules have already faced criticism for limiting affordable access to

Type	Permissible questions	Impermissible questions
Procedure	Can you tell me how to file a small claims action?	Can you tell me whether it would be better to file a small claims action or a civil action?
Definition	What does "certificate of service" mean?	My neighbors leave their kids at home all day without supervision. Isn't that child neglect?
Forms	I need to file for divorce and I have no idea where to begin. Is there some place I can go to find out how to get started?	The self-help divorce petition says I should list any gifts as my separate property. Should I list the money that my parents gave me last month as my separate property?
Options	What can I do if I cannot afford to pay the filing fee?	My ex-husband hasn't paid the debts that he agreed to pay in our divorce settlement. Can I be made responsible for this debt?

Table 2: Examples of impermissible questions that requires legal opinion [58]. Remarkably similar to how red-teaming in LLM development identifies harmful user inputs [4, 23], this edited list (compiled from Texas law clerk resources) distinguishes between permissible and forbidden questions Texas court personnel can answer.

legal help [7, 18, 64]. The EU AI Act's categorization of AI legal assistance tools as "high-risk," which subjects them to heightened responsibilities instead of banning them outright, may speak to this potential shift [20].

5.3 Learning from Time-tested Wisdom

Leveraging accumulated expertise in professional communities can help sidestep painful mistakes [6]. In our research, UPL does not only constrain the possibility of LLM legal advice but also provides long-standing distinction criteria between information versus opinion, as merely providing legal information has not been historically punished as a UPL violation [2, 3, 18]. For example, the Texas Court provides guidelines for court staff and illustrative examples like in Table 2. These examples show subtle differences between information and opinion, which resembles the red-teaming approach to distinguish harmful user prompts [4, 23]. Furthermore, legal scholars have explored legally justifiable AI advice under UPL, attorney-client privilege, and other doctrines [30, 78, 88]. Wendel stated that the "core lawyering functions" such as recommending the course of action or drafting contracts cannot be delegated to AI agents due to technical limitations and accountability deficits [88]. This demonstrates how principles accumulated over centuries of legal scholarship now inform responsible LLM systems and the call for cross-disciplinary collaborations.

5.4 Applicability to Other Professional Domains

While each possessing unique dimensions, domains like medicine, mental health, law, and finance share common threads around high-stakes real-world impact and historical reliance on licensed specialists for advice. We believe that our research methods and 4-dimension framework give illustrative guidance to further research in other professional domains. As this research demonstrates how case-based deliberation methods can unravel complex professional ethics, researchers could adopt similar processes engaging mental health counselors, financial advisors, or medical professionals. Tapping into the clinical experience and integrity of practitioners through structured deliberation based on realistic cases can help produce tailored dimensions and guidelines for responsible LLM advice respective to each profession. Building upon this foundation, our 4-dimension framework—(1) user, (2) query, (3) AI capabilities, and (4) impact—could be adapted and applied across various professional domains. The (1) user, (2) AI, and (4) impact dimensions can be applied in other domains with minimal modifications. However, the query dimension requires more customization to address the typical requests of clients, terminology, and satisfactory responses in each field.

5.5 Limitations and Future Research

Our study has several limitations. First, our expert sample predominantly focused on practitioners familiar with the US legal system. Ethical considerations around appropriate AI assistance may differ across different legal systems and cultures. Second, our participants' responses are conditioned by their prior experience with state-of-the-art LLM technology, such as ChatGPT empowered by GPT-4. Experts' evaluations of the appropriateness of LLM legal advice may evolve in the future, based on technological innovations, which could be an avenue for future research. Third, we did not engage end-users like clients of legal services. Future work can specifically investigate end-user perceptions to compare and contrast with our expert-informed results. Finally, while our taxonomy conceptualizes a concrete set of dimensions, how these dimensions could change the appropriateness of LLM responses remains unexplained. This may require larger-scale empirical analysis on public assessments across diverse pairings of cases and responses.

6 CONCLUSION

Today, LLM chatbots are increasingly capable of providing users with advice in a wide range of professional domains, including legal advice. However, what constitutes an appropriate LLM-generated response to legal queries, where both required expertise and resulting consequences are high? To explore this, we conducted workshops with 20 legal experts using methods inspired by case-based reasoning to encourage deliberations around appropriate LLM responses to legal queries in practice. Our contributions are threefold. First, we presented a set of 25 key dimensions, synthesized from expert deliberations, that impacted LLM response appropriateness in the legal domain. Second, we shared experts' recommendations for LLM response strategies and guiding principles for generating appropriate responses—these centered around helping users identify and prepare salient information for legal proceedings rather than recommending specific legal actions. Finally, we posit that our case-based method has utility in engaging expert perspectives on LLM response appropriateness in professional domains beyond the legal sphere. Taken together, our work sets an empirical foundation for translating domain-specific professional knowledge and practices into policies to steer real-world LLM behavior in a more responsible direction.

ACKNOWLEDGMENTS

This project was supported by OpenAI's "Democratic Inputs to AI" grant program. We thank Teddy Lee, Tyna Eloundou, Alex Beutel, Aviv Ovadya, members of the grant recipient cohort, and others at OpenAI for insightful discussions and support. We are deeply appreciative of all the legal experts who participated in our workshops, generously contributing their time, expertise, and perspectives. Additionally, we thank Attorney Daniel Keum for providing valuable feedback on confidentiality matters.

REFERENCES

[1] 1966. Grievance Comm. of Bar v. Dacey, 222 A.2d 339 (Conn.), appeal dismissed, 386 U.S. 683.

[2] 1970. Baron v. City of Los Angeles, 2 Cal. 3d 535.

[3] 2007. Cal. Bus. & Prof. Code § 6450.

[4] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. (2023). arXiv:2303.08774

[5] Benjamin Alarie and Rory McCreight. 2023. The Ethics of Generative AI in Tax Practice. *Tax Notes Federal* (2023), 785–793.

[6] Maria Antoniak, Aakanksha Naik, Carla S. Alvarado, Lucy Lu Wang, and Irene Y. Chen. 2023. Designing Guiding Principles for NLP for Healthcare: A Case Study of Maternal Health. (2023). arXiv:2312.11803

[7] American Bar Association. 1995. *Nonlawyer activity in law-related situations: A report with recommendations.* ABA, Chicago, IL.

[8] Brhmie Balaram, Tony Greenham, and Jasmine Leonard. 2018. Artificial Intelligence: real public engagement. *RSA, London* 5 (2018).

[9] Michael Balas, Jordan Joseph Wadden, Philip C. Hébert, Eric Mathison, Marika D. Warren, Victoria Seavilleklein, Daniel Wyzynski, Alison Callahan, Sean A. Crawford, Parnian Arjmand, and Edsel B. Ing. 2023. Exploring the potential utility of AI large language models for medical ethics: an expert panel evaluation of GPT-4. *Journal of Medical Ethics* (2023). https://doi.org/10.1136/jme-2023-109549

[10] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 610–623. https://doi.org/10.1145/3442188.3445922

[11] Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. Power to the people? opportunities and challenges for participatory AI. *Equity and Access in Algorithms, Mechanisms, and Optimization* (2022), 1–8.

[12] Benjamin N. Cardozo and Andrew L. Kaufman. 2010. *The Nature of the Judicial Process.* Quid Pro, LLC.

[13] Columbia Law School Writing Center. 2001. *Organizing a Legal Discussion (IRAC, CRAC, etc.).* Retrieved May 1, 2024 from https://www.law.columbia.edu/sites/default/files/2021-07/organizing_a_legal_discussion.pdf

[14] Quan Ze Chen and Amy X Zhang. 2023. Case Law Grounding: Aligning Judgments of Humans and AI on Socially-Constructed Concepts. (2023). arXiv:2310.07019

[15] Sasha Costanza-Chock. 2020. *Design Justice: Community-Led Practices to Build the Worlds We Need.* The MIT Press, Cambridge, MA London.

[16] Alexandra D'Arcy and Emily M. Bender. 2023. Ethics in Linguistics. *Annual Review of Linguistics* 9 (2023), 49–69.

[17] Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. 2023. The participatory turn in ai design: Theoretical foundations and the current state of practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization.* 1–23.

[18] Derek A Denckla. 1998. Nonlawyers and the unauthorized practice of law: an overview of the legal and ethical parameters. *Fordham Law Review* 67 (1998), 2581.

[19] Harnoor Dhingra, Preetiha Jayashanker, Sayali Moghe, and Emma Strubell. 2023. Queer people are people first: Deconstructing sexual identity stereotypes in large language models. (2023). arXiv:2307.00101

[20] European Commission. 2021. *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts.* Retrieved May 1, 2024 from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021PC0206 COM(2021) 206 final 2021/0106(COD).

[21] K. J. Kevin Feng, Quan Ze Chen, Inyoung Cheong, King Xia, and Amy X. Zhang. 2023. Case Repositories: Towards Case-Based Reasoning for AI Alignment. (2023). arXiv:2311.10934

[22] Robert K Fullinwider. 2010. Philosophy, casuistry, and moral development. *Theory and Research in Education* 8, 2 (2010), 173–185.

[23] Deep Ganguli, Liane Lovitt, Jackson Kernion, Amanda Askell, Yuntao Bai, Saurav Kadavath, Ben Mann, Ethan Perez, Nicholas Schiefer, Kamal Ndousse, et al. 2022. Red teaming language models to reduce harms: Methods, scaling behaviors, and lessons learned. (2022). arXiv:2209.07858

[24] Sourojit Ghosh and Aylin Caliskan. 2023. ChatGPT Perpetuates Gender Bias in Machine Translation and Ignores Non-Gendered Pronouns: Findings across Bengali and Five other Low-Resource Languages. (2023). arXiv:2305.10510

[25] Candida M. Greco and Andrea Tagarelli. 2023. Bringing order into the realm of Transformer-based language models for artificial intelligence and law. (2023). arXiv:2308.05502

[26] Kai Greshake, Sahar Abdelnabi, Shailesh Mishra, Christoph Endres, Thorsten Holz, and Mario Fritz. 2023. Not What You've Signed Up For: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection. In *Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security* (Copenhagen, Denmark) *(AISec '23)*. Association for Computing Machinery, New York, NY, USA, 79–90. https://doi.org/10.1145/3605764.3623985

[27] Thomas C Grey. 1983. Langdell's orthodoxy. *University of Pittsburgh Law Review* 45 (1983), 1–54.

[28] Neel Guha, Julian Nyarko, Daniel E. Ho, Christopher Ré, Adam Chilton, Aditya Narayana, Alex Chohlas-Wood, Austin Peters, Brandon Waldon, Daniel N. Rockmore, Diego Zambrano, Dmitry Talisman, Enam Hoque, Faiz Surani, Frank Fagan, Galit Sarfaty, Gregory M. Dickinson, Haggai Porat, Jason Hegland, Jessica Wu, Joe Nudell, Joel Niklaus, John Nay, Jonathan H. Choi, Kevin Tobia, Margaret Hagan, Megan Ma, Michael Livermore, Nikon Rasumov-Rahe, Nils Holzenberger, Noam Kolt, Peter Henderson, Sean Rehaag, Sharad Goel, Shang Gao, Spencer Williams, Sunny Gandhi, Tom Zur, Varun Iyer, and Zehua Li. 2023. LegalBench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models. (2023). arXiv:2308.11462

[29] Biyang Guo, Xin Zhang, Ziyuan Wang, Minqi Jiang, Jinran Nie, Yuxuan Ding, Jianwei Yue, and Yupeng Wu. 2023. How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection. (2023). arXiv:2301.07597

[30] Claudia E. Haupt. 2019. Artificial professional advice. *Yale Journal of Law and Technology* 21 (2019), 55–77.

[31] Peter Henderson, Jieru Hu, Mona Diab, and Joelle Pineau. 2024. Rethinking Machine Learning Benchmarks in the Context of Professional Codes of Conduct. In *Proceedings of the Symposium on Computer Science and Law* (, Boston, MA, USA,) *(CSLAW '24)*. Association for Computing Machinery, New York, NY, USA, 109–120. https://doi.org/10.1145/3614407.3643708

[32] Quzhe Huang, Mingxu Tao, Zhenwei An, Chen Zhang, Cong Jiang, Zhibin Chen, Zirui Wu, and Yansong Feng. 2023. Lawyer LLaMA Technical Report. (2023). arXiv:2305.15062

[33] Umar Iqbal, Tadayoshi Kohno, and Franziska Roesner. 2023. LLM Platform Security: Applying a Systematic Evaluation Framework to OpenAI's ChatGPT Plugins. (2023). arXiv:2309.10254

[34] Mohd Javaid, Abid Haleem, and Ravi Pratap Singh. 2023. ChatGPT for healthcare services: An emerging stage for an innovative perspective. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations* 3, 1 (2023), 100105. https://doi.org/10.1016/j.tbench.2023.100105

[35] Albert R Jonsen. 1986. Casuistry and clinical ethics. *Theoretical Medicine* 7 (1986), 65–74.

[36] Sayash Kapoor, Peter Henderson, and Arvind Narayanan. 2024. Promises and pitfalls of artificial intelligence for legal applications. *Journal of Cross-disciplinary Research in Computational Law* 2, 22 (May 2024). https://journalcrcl.org/crcl/article/view/62

[37] Sunnie S. Y. Kim, Elizabeth Anne Watkins, Olga Russakovsky, Ruth Fong, and Andrés Monroy-Hernández. 2023. Humans, AI, and Context: Understanding End-Users' Trust in a Real-World Computer Vision Application. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 77–88. https://doi.org/10.1145/3593013.3593978

[38] Yoonsu Kim, Jueon Lee, Seoyoung Kim, Jaehyuk Park, and Juho Kim. 2023. Understanding Users' Dissatisfaction with ChatGPT Responses: Types, Resolving Tactics, and the Effect of Knowledge Level. (2023). arXiv:2311.07434

[39] Janet L. Kolodner. 1992. An introduction to case-based reasoning. *Artificial Intelligence Review* 6, 1 (March 1992), 3–34. https://doi.org/10.1007/BF00155578

[40] Hadas Kotek, Rikker Dockum, and David Sun. 2023. Gender bias and stereotypes in Large Language Models. In *Proceedings of The ACM Collective Intelligence*

Conference. 12–24.

[41] Davide Liga and Livio Robaldo. 2023. Fine-tuning GPT-3 for legal rule classification. *Computer Law and Security Review* 51 (Nov. 2023), 105864. https://doi.org/10.1016/j.clsr.2023.105864

[42] John Lightbourne. 2017. Algorithms & fiduciaries: existing and proposed regulatory approaches to artificially intelligent financial planners. *Duke Law Journal* 67 (2017), 651–680.

[43] Stephanie Lin, Jacob Hilton, and Owain Evans. 2021. Truthfulqa: Measuring how models mimic human falsehoods. (2021). arXiv:2109.07958

[44] Chen Ling, Xujiang Zhao, Jiaying Lu, Chengyuan Deng, Can Zheng, Junxiang Wang, Tanmoy Chowdhury, et al. 2023. Beyond One-Model-Fits-All: A Survey of Domain Specialization for Large Language Models. (2023). arXiv:2305.18703

[45] Nelson F. Liu, Tianyi Zhang, and Percy Liang. 2023. Evaluating Verifiability in Generative Search Engines. (2023). arXiv:2304.09848

[46] Zilin Ma, Yiyang Mei, and Zhaoyuan Su. 2024. Understanding the Benefits and Challenges of Using Large Language Model-based Conversational Agents for Mental Well-being Support. *AMIA Annual Symposium Proceedings* 2023, 1105–1114.

[47] John Leslie Mackie. 2003. *Hume's moral theory*. Routledge.

[48] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-Rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 72 (nov 2019), 23 pages. https://doi.org/10.1145/3359174

[49] Katherine Medianik. 2017. Artificially intelligent lawyers: updating the model rules of professional conduct in accordance with the new technological era. *Cardozo Law Review* 39 (2017), 1497–1532.

[50] Donald Metzler, Yi Tay, Dara Bahri, and Marc Najork. 2021. Rethinking Search: Making Domain Experts out of Dilettantes. In *ACM SIGIR Forum*, Vol. 55. ACM, New York, NY, USA, 1–27.

[51] Jeffrey Metzler. 2002. The importance of IRAC and legal writing. *University of Detroit Mercy Law Review* 80 (2002), 501–504.

[52] John J. Nay. 2023. Large language models as corporate lobbyists. (2023). arXiv:2301.01181

[53] John J. Nay, David Karamardian, Sarah B. Lawsky, Wenting Tao, Meghana Bhat, Raghav Jain, Aaron Travis Lee, Jonathan H. Choi, and Jungo Kasai. 2023. Large Language Models as Tax Attorneys: A Case Study in Legal Capabilities Emergence. (2023). arXiv:2306.07075

[54] Anam Nazir and Ze Wang. 2023. A Comprehensive Survey of ChatGPT: Advancements, Applications, Prospects, and Challenges. *Meta-radiology* 1, 2 (Sept. 2023), 100022. https://doi.org/10.1016/j.metrad.2023.100022

[55] Wilhelmina Nekoto, Vukosi Marivate, Tshinondiwa Matsila, Timi Fasubaa, Tajudeen Kolawole, Taiwo Fagbohungbe, Solomon Oluwole Akinola, Shamsuddeen Hassan Muhammad, Salomon Kabongo, Salomey Osei, et al. 2020. Participatory research for low-resourced machine translation: A case study in african languages. (2020). arXiv:2010.02353

[56] Ha-Thanh Nguyen, Wachara Fungwacharakorn, and Ken Satoh. 2023. Enhancing Logical Reasoning in Large Language Models to Facilitate Legal Applications. (2023). arXiv:2311.13095

[57] Gavin Northey, Vanessa Hunter, Rory Mulcahy, Kelly Choong, and Michael Mehmet. 2022. Man vs machine: how artificial intelligence in banking influences consumer belief in financial advice. *International Journal of Bank Marketing* 40, 6 (2022), 1182–1199.

[58] Texas Office of Court. 2015. *Legal Information vs. Legal Advice.* Retrieved May 1, 2024 from https://www.txcourts.gov/media/1220087/legalinformationvslegaladviceguidelines.pdf

[59] Shiva Omrani Sabbaghi, Robert Wolfe, and Aylin Caliskan. 2023. Evaluating biased attitude associations of language models in an intersectional context. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. 542–553.

[60] Norbert Paulo. 2015. Casuistry as common law morality. *Theoretical Medicine and Bioethics* 36, 6 (2015), 373–389.

[61] Denis Peskoff and Brandon Stewart. 2023. Credible without Credit: Domain Experts Assess Generative Language Models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 427–438. https://doi.org/10.18653/v1/2023.acl-short.37

[62] Nishchal Prasad, Mohand Boughanem, and Taoufiq Dkaki. 2022. Effect of hierarchical domain-specific language models and attention in the classification of decisions for legal cases. In *Proceedings of the CIRCLE (Joint Conference of the Information Retrieval Communities in Europe), Samatan, Gers, France*. 4–7.

[63] Organizers Of Queerinai, Anaelia Ovalle, Arjun Subramonian, Ashwin Singh, Claas Voelcker, Danica J. Sutherland, Davide Locatelli, Eva Breznik, Filip Klubicka, Hang Yuan, Hetvi J, Huan Zhang, Jaidev Shriram, Kruno Lehman, Luca Soldaini, Maarten Sap, Marc Peter Deisenroth, Maria Leonor Pacheco, Maria Ryskina, Martin Mundt, Milind Agarwal, Nyx Mclean, Pan Xu, A Pranav, Raj Korpan, Ruchira Ray, Sarah Mathew, Sarthak Arora, St John, Tanvi Anand, Vishakha Agrawal, William Agnew, Yanan Long, Zijie J. Wang, Zeerak Talat, Avijit Ghosh, Nathaniel Dennler, Michael Noseworthy, Sharvani Jha, Emi Baylor, Aditya Joshi, Natalia Y. Bilenko, Andrew Mcnamara, Raphael Gontijo-Lopes, Alex Markham, Evyn Dong, Jackie Kay, Manu Saraswat, Nikhil Vytla, and Luke Stark. 2023. Queer In AI: A Case Study in Community-Led Participatory AI. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 1882–1895. https://doi.org/10.1145/3593013.3594134

[64] Mathew Rotenberg. 2012. Stifled Justice: The Unauthorized Practice of Law and Internet Legal Resources. *Minnesota Law Review* 97 (2012), 709–742.

[65] Tulika Saha, Debasis Ganguly, Sriparna Saha, and Prasenjit Mitra. 2023. Workshop On Large Language Models' Interpretability and Trustworthiness (LLMIT). In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 5290–5293.

[66] Malik Sallam. 2023. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare (Basel, Switzerland)* 11, 6 (March 2023), 887. https://doi.org/10.3390/healthcare11060887

[67] Amy J Schmitz and John Zeleznikow. 2022. Intelligent legal tech to empower self-represented litigants. *Ohio State Legal Studies Research Paper* 23 (2022), 142–191.

[68] Chirag Shah and Emily M. Bender. 2022. Situating Search. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval* (Regensburg, Germany) *(CHIIR '22)*. Association for Computing Machinery, New York, NY, USA, 221–232. https://doi.org/10.1145/3498366.3505816

[69] Murray Shanahan, Kyle McDonell, and Laria Reynolds. 2023. Role play with large language models. *Nature* (2023), 1–6.

[70] Xinyue Shen, Zeyuan Chen, Michael Backes, and Yang Zhang. 2023. In ChatGPT We Trust? Measuring and Characterizing the Reliability of ChatGPT. (2023). arXiv:2304.08979

[71] Drew Simshaw. 2018. Ethical issues in robo-lawyering: The need for guidance on developing and using artificial intelligence in the practice of law. *Hastings Law Journal* 70 (2018), 173–214.

[72] Chandan Singh, Armin Askari, Rich Caruana, and Jianfeng Gao. 2023. Augmenting interpretable models with large language models during training. *Nature Communications* 14:7913 (2023), 1–11.

[73] Karan Singhal, Tao Tu, Juraj Gottweis, Rory Sayres, Ellery Wulczyn, Le Hou, Kevin Clark, Stephen Pfohl, Heather Cole-Lewis, Darlene Neal, et al. 2023. Towards expert-level medical question answering with large language models. (2023). arXiv:2305.09617

[74] Adam Smith. 1982. *The Theory of Moral Sentiments* (reissue ed.). Liberty Classics, Indianapolis, Ind.

[75] Centaine L Snoswell, Aaron J Snoswell, Jaimon T Kelly, Liam J Caffery, and Anthony C Smith. 2023. Artificial intelligence: Augmenting telehealth with large language models. *Journal of telemedicine and telecare* (2023), 1357633X231169055.

[76] Irene Solaiman, Zeerak Talat, William Agnew, Lama Ahmad, Dylan Baker, Su Lin Blodgett, Hal Daumé III au2, Jesse Dodge, Ellie Evans, Sara Hooker, Yacine Jernite, Alexandra Sasha Luccioni, Alberto Lusoli, Margaret Mitchell, Jessica Newman, Marie-Therese Png, Andrew Strait, and Apostol Vassilev. 2023. Evaluating the Social Impact of Generative AI Systems in Systems and Society. (2023). arXiv:2306.05949

[77] Tania Sourdin. 2018. Judge v Robot?: Artificial intelligence and judicial decision-making. *University of New South Wales Law Journal* 41, 4 (2018), 1114–1133.

[78] Thomas E. Spahn. 2017. Is Your Artificial Intelligence Guilty of the Unauthorized Practice of Law. *Richmond Journal of Law and Technology* 24 (2017), 1–47.

[79] Michael Stockdale and Rebecca Mitchell. 2019. Legal advice privilege and artificial legal intelligence: Can robots give privileged legal advice? *The International Journal of Evidence & Proof* 23, 4 (2019), 422–439. https://doi.org/10.1177/1365712719862296

[80] Sasha Fathima Suhel, Vinod Kumar Shukla, Sonali Vyas, and Ved Prakash Mishra. 2020. Conversation to automation in banking through chatbot using artificial machine intelligence language. In *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*. IEEE, 611–618.

[81] Cass R Sunstein. 2018. *Legal reasoning and political conflict*. Oxford University Press.

[82] Iddo Tavory and Stefan Timmermans. 2014. *Abductive analysis: Theorizing qualitative research*. University of Chicago Press.

[83] Dietrich Trautmann, Alina Petrova, and Frank Schilder. 2022. Legal prompt engineering for multilingual legal judgement prediction. (2022). arXiv:2212.02199

[84] Arianna Trozze, Toby Davies, and Bennett Kleinberg. 2024. Large Language Models in Cryptocurrency Securities Cases: Can a GPT Model Meaningfully Assist Lawyers? (2024). arXiv:2308.06032

[85] Josef Valvoda, Ryan Cotterell, and Simone Teufel. 2023. On the role of negative precedent in legal outcome prediction. *Transactions of the Association for Computational Linguistics* 11 (2023), 34–48.

[86] Dave Van Veen, Cara Van Uden, Louis Blankemeier, Jean-Benoit Delbrouck, Asad Aali, Christian Bluethgen, Anuj Pareek, Malgorzata Polacin, Eduardo Pontes Reis, Anna Seehofnerova, Nidhi Rohatgi, Poonam Hosamani, William Collins, Neera Ahuja, Curtis P. Langlotz, Jason Hom, Sergios Gatidis, John Pauly, and Akshay S.

Chaudhari. 2023. Clinical Text Summarization: Adapting Large Language Models Can Outperform Human Experts. (2023). arXiv:2309.07430

[87] Benjamin Weiser. 2023. ChatGPT Lawyers Are Ordered to Consider Seeking Forgiveness. *The New York Times* (June 2023). Retrieved May 1, 2024 from https://www.nytimes.com/2023/06/22/nyregion/lawyers-chatgpt-schwartz-loduca.html

[88] W. Bradley Wendel. 2019. The promise and limitations of artificial intelligence in the practice of law. *Oklahoma Law Review* 72 (2019), 21–50.

[89] Yifan Yao, Jinhao Duan, Kaidi Xu, Yuanfang Cai, Eric Sun, and Yue Zhang. 2023. A Survey on Large Language Model (LLM) Security and Privacy: The Good, the Bad, and the Ugly. (2023). arXiv:2312.02003

[90] Haiyan Zhao, Hanjie Chen, Fan Yang, Ninghao Liu, Huiqi Deng, Hengyi Cai, Shuaiqiang Wang, Dawei Yin, and Mengnan Du. 2024. Explainability for Large Language Models: A Survey. *ACM Trans. Intell. Syst. Technol.* 15, 2, Article 20 (feb 2024), 38 pages. https://doi.org/10.1145/3639372

A PROVIDED AI RESPONSE STRATEGIES AND EXAMPLES

Table 3: AI response strategies and corresponding example responses.

Response Strategies	Description	Example Response (Legal Advice)
Content violation	Displays a warning message that the user's input violates the AI's content policy.	This content may violate our content policy. If you believe this to be in error, please submit your feedback—your input will aid our research in this area.
Refuse to answer	Avoids providing a response to the input and advises the user to seek professional advice instead.	As an AI model, I am not a substitute for consulting with a trained professional. This seems like a specific concern or issue that requires expert advice. It's always advisable to speak with a professional in the relevant field.
Empathize + Refuse	Emphasizes with a user but avoids providing a response to the input.	I am sorry that this is happening to you. It sounds like a challenging situation. Unfortunately, as an AI model, I cannot provide specific legal advice…
Retrieve (non-opinion) facts	Provides a detailed response that cites specific facts relevant to the input but avoids providing a further legal analysis.	As an AI model I cannot give legal advice specific to your situation but here is a list of laws and cases that may be helpful to look into: (1) Compliance with Applicable Laws and Regulations: Laws like the Children's Online Privacy Protection Act (COPPA) and the California Consumer Privacy Act (CCPA) may be relevant…
Help work through questions	Helps users to identify relevant matters in their situations and, if necessary, responds with a question requesting more information.	In matters of divorce, legal procedures often require addresing issues such as child custody and spousal support …. To help you with this, I will need some additional information. How many children do you have?
Recommend Actions	Analyzes a user's situation under the relevant laws and provides specific further steps that a user may take.	I regret the distressing experience of verbal abuse that you endured at school. In California, potential legal recourses may encompass claims of negligence, a violation of federal civil rights laws (Title XI) …As a prudent next step, I recommend initiating formal proceedings by filing a complaint with the appropriate administrative body responsible for overseeing Title IX compliance within the school or educational institution.

B WORKSHOP PARTICIPANT INFORMATION

Table 4: Workshop Participant Information

Number	Legal Experience (yrs)	Category	AI Use (General)	AI Use (Work)
P1	> 20	Law faculty	Occasional	Occasional
P2	< 5	Attorney	Occasional	Occasional
P3	> 20	Law faculty	Regular	Occasional
P4	6-10	Attorney	Regular	Regular
P5	11-15	Attorney	Occasional	Never
P6	> 20	Law faculty	Regular	Regular
P7	< 5	Law student	Regular	Never
P8	11-15	Attorney	Regular	Regular
P9	6-10	Law faculty	Occasional	Occasional
P10	< 5	Attorney	Regular	Regular
P11	< 5	Attorney	Regular	Regular
P12	< 5	Researcher	Regular	Regular
P13	6-10	Attorney	Regular	Regular
P14	< 5	Attorney	Regular	Regular
P15	< 5	Law student	Regular	Occasional
P16	6-10	Attorney	Regular	Regular
P17	16-20	Attorney	Occasional	Occasional
P18	< 5	Attorney	Occasional	Never
P19	< 5	Law student	Regular	Occasional
P20	< 5	Attorney	Regular	Regular

Note: Years of legal experience is self-reported with years of legal education removed for consistency.

C LINEAR REGRESSION OF PARTICIPANTS' AI USAGE AND DESIRED RESPONSES

Presented in Table 5, participants' receptivity to a tailored AI response is estimated by the average of the most generous answer types per each prompt. The "content warning" is marked as 0 points, the lowest comfort level, and the "recommend action" template is marked as 6 points. For example, if a participant chose both "empathize + refusal" (2 points) and "Help work through questions" (4 points) for the first case (the higher point is 4) and chose "Recommend actions" (6 points) for the second case, we marked their receptivity level as 5 points. While P13 worked on four cases, all other participants chose two cases each. The regression results (Table 6) indicate that general AI fluency significantly predicts higher comfort levels with proactive AI responses ($p < 0.05$), whereas work AI fluency is marginally associated with lower comfort levels ($p = 0.054$). The predictors explain 25.6% of variation. Further investigation is required to substantiate these preliminary relationships with a larger sample.

Table 5: AI Use and Receptivity

Number	AI Use (General)	AI Use (Work)	Receptivity
P1	1	1	1
P2	1	1	4
P3	2	1	5
P4	2	2	4
P5	1	0	4
P6	2	2	3.5
P7	2	0	5.5
P8	2	2	2.5
P9	1	1	4
P10	2	2	4.5
P11	2	2	4
P12	2	2	1
P13	2	2	3.75
P14	2	2	4.5
P15	2	1	4.5
P16	2	2	6
P17	1	1	4
P18	1	0	4
P19	2	1	6
P20	2	2	4.5

Note: A pre-survey asked participants to describe their AI usage in both professional ("Work") and non-professional ("General") settings, using a scale where 0 represented "Never," 1 "Occasional use," and 2 "Regular use." We then estimated receptivity to more tailored responses such as opinion by averaging the most generous answer types for each case.

Table 6: Regression Results.

Predictor	Estimate	p-value
Intercept	2.4682	0.0297
AI usage in work	-0.9682	0.0543
AI usage daily	1.6773	0.0373

- Residual Std. Error: 1.199 on 17 degrees of freedom
- Multiple R-squared: 0.2557
- Adjusted R-squared: 0.1681
- F-statistic: 2.92 on 2 and 17 DF
- p-value: 0.08127