# Can an LLM tell me if I can legally get an abortion?

Ro Encarnación
University of Pennsylvania
Philadelphia, PA, USA
rone@seas.upenn.edu

Danaé Metaxa
University of Pennsylvania
Philadelphia, PA, USA
metaxa@seas.upenn.edu

## Abstract

Following the landmark ruling in *Dobbs v. Jackson Women's Health Organization (2022)*, which effectively overturned *Roe v. Wade*, navigating the legality of abortion across the United States has become increasingly challenging. This evolving landscape is compounded by the growing use and implementation of large language models (LLMs) to navigate search results, which risks disseminating inaccurate or outdated information about state abortion laws. To assess how large language models (LLMs) interpret regionally complex and nuanced social policy issues like abortion, we developed a pilot study that examines how variations in prompt phrasing, specificity, and complexity influence the correctness, completeness, and consistency of responses generated by OpenAI ChatGPT and Google Gemini regarding abortion laws in Alabama, Pennsylvania, and New Jersey. By analyzing LLM responses to six structured binary and long-form prompts, we explore how the dissemination of pertinent legal information varies by location. Our preliminary findings indicate that responses for New Jersey are the most incorrect, followed by Pennsylvania and Alabama. Even though responses for Alabama exhibited the highest percentage of correct responses, we found they also had the worst overall consistency. Additionally, while the LLMs provided consistent responses for Pennsylvania, they were incorrect 50% of the time. In addition to these early results, we discuss future considerations for the development of context-specific evaluations of LLMs.

## CCS Concepts

• **Information systems → Evaluation of retrieval results**; • **Applied computing → Law**; • **Human-centered computing → HCI design and evaluation methods**; **Empirical studies in HCI**.

## Keywords

Large Language Models, OpenAI GPT-4, Google Gemini, LLM evaluation, abortion policy, reproductive health

## 1 Introduction

As LLMs like OpenAI ChatGPT and Google Gemini become more common sources of information, concerns about their tendency to generate false or misleading statements are becoming increasingly pressing [5, 21, 25, 28]. These concerns are particularly urgent in complex, high-stakes domains such as healthcare and legal policy, where inaccurate information can lead to serious consequences, drawing attention to the need for thorough and context-specific LLM evaluations. Despite the risks, a 2024 survey indicates that 1 in 6 adults use LLMs for critical searches such as medical information [27].

This reliance on AI-driven search is especially relevant in the post-*Dobbs* era, where the patchwork of state-specific abortion laws creates a complex and often unclear legal landscape for people seeking abortion care. Due to growing legal hostility and risks associated with accessing abortion care [6, 30], people may turn to the internet before consulting professionals. In fact, following the *Dobbs* ruling, searches for abortion-related terms increased by 42% in states with immediate abortion bans like Alabama compared to states that offer protections like New Jersey [7].

Even people not actively using LLMs or generative AI may unknowingly interact with AI-generated content due to the growing integration of generative AI in search. Google's AI Overviews and Bing's Copilot offer AI-generated summaries; amplifying the reach of potentially misleading information. However, early reports have identified instances of fabricated information and unsubstantiated content [17, 20]. As generative AI search expands and more LLM-based systems enter the market [18], avoiding misinformation online will become increasingly difficult. People may encounter misleading or inconsistent legal and healthcare information, complicating reliable access to accurate legal guidance. Meanwhile, LLMs like OpenAI ChatGPT and Google Gemini have also been found to generate incorrect or misleading responses regarding medical abortion information [14, 19] or legal text interpretation [4]. This misinformation can be especially dangerous when users rely on LLMs as de facto advisors.

Unlike casual web searches, questions about obtaining abortion care in the U.S. require accurate, jurisdiction-specific, and context-sensitive responses. To evaluate how LLMs respond to such questions, we need frameworks that consider the relevant context throughout the evaluation lifecycle and assess responses based on human expectations in the real world, especially in areas like abortion law, where legal nuances can have profound impacts on people's lives [2, 23]. To this end, we present our preliminary findings from a pilot study conducted to inform the design of a larger system to evaluate LLM responses using abortion policies as a case study across all 50 U.S. states.

In this exploratory phase, we test our developing approach to assess how OpenAI ChatGPT and Google Gemini respond to 6 questions on the legality of abortions in Alabama, New Jersey, and Pennsylvania relative to the direct language of abortion policies in each state. We begin with these three states to capture a range of abortion policies, allowing us to preview how LLMs perform regarding correctness, completeness, and consistency while refining the methodological choices that will shape the broader study. Our findings from this in-progress work serve as early signals of broader challenges people using LLMs for legally complex and jurisdiction-specific information-seeking may encounter. We outline how these findings advance the development of context-specific LLM evaluation frameworks for complex policy applications.

## 2 Related Work

General-purpose LLMs like ChatGPT and Google Gemini can actively influence decision-making when used as information retrieval tools, making it a pressing issue to evaluate their effectiveness in guiding people toward accurate and suitable information, especially in high-stakes, quickly-evolving, and contentious policy domains such as abortion care. Several studies in health and medical research have evaluated LLM-generated responses to abortion-related questions, showing that LLMs tend to overstate risks, provide contradictory information, and fail to align with reputable medical sources [13, 14, 19]. However, these studies focus primarily on medical accuracy rather than legal and regional complexities, creating a gap in HCI research on how LLMs handle the intersection of legal restrictions and geographic variability in social policy issues. This in-progress work aims to contribute towards filling this gap by conducting a pilot study aimed at developing policy-sensitive evaluation methods to understand how LLMs respond with respect to high-stakes, geographically dependent social policy areas like abortion.

## 3 Pilot Study Method

### 3.1 State Selection

We evaluate the responses provided by OpenAI's GPT-4o ChatGPT model [24] and Google Gemini's 1.5 Flash [29] model when answering prompts related to the legality of abortions in three U.S. states: Alabama, Pennsylvania, and New Jersey; representing a range of abortion policies.

These three states are categorized based on the classifications from the Center for Reproductive Rights [3]. Alabama is labeled as a state where abortion is "illegal" because it is one of the 12 states with a total abortion ban in effect. We consider Pennsylvania a "hostile" state, in the middle of the three selected states, where abortion is legal until 24 weeks but comes with numerous legal restrictions that limit access to abortion care. 22 other states ban abortion at some point after 18 weeks [12]. In contrast, we consider New Jersey a state with "expanded access" as abortion is fully legal with no mandated gestational limits. Only 8 other states have no ban or gestational limits [15].

### 3.2 Prompt Design

*3.2.1 Designing for Variability in Prompt Phrasing* Examining the varying levels of policy complexity across these three states provides crucial context for designing the following 6 prompts that assess how ChatGPT and Gemini navigate legal nuances. This approach allows us to design prompts with a clear awareness of the legal, cultural, and domain-specific background in which each LLM is expected to operate.

The 6 prompts (detailed in Table 1) vary in phrasing, specificity, and complexity. These questions were inspired by Reddit posts from users asking about abortion restrictions. We found these posts by searching for terms like "abortion," "abortion legal," and "abortion law" in the /r/abortion, /r/legaladvice, /r/legaladviceofftopic, and /r/TwoXChromosomes subreddit communities.

The first 3 prompts (Prompts 1–3) are specific to a type of requirement common across U.S. abortion policies called waiting periods. Waiting periods require anyone seeking an abortion to wait a legally specified amount of time between their initial consultation and the abortion procedure [8]. We focused on waiting periods for this subset of prompts as it allows us to focus on how LLMs respond to questions about a single restriction, even if it varies across states. These first 3 prompts also include references to a D&C which is a common procedure performed for abortions as well as other purposes. The intentional use of this term in the first 3 prompt designs is also meant to evaluate how the LLM being tested responds when it does not have enough information to determine whether "D&C" in the context of the prompt refers to abortion. In this initial set of three prompts, we started with broad phrasing. We gradually moved to more specific language to evaluate the model's ability to maintain accuracy, avoid contradictions, and understand the legal nuances of state-level abortion policies.

The second set of prompts (Prompts 4–6) is focused on asking about all relevant abortion requirements and restrictions to evaluate how LLMs respond when the phrasing is general and grows more complex. Prompt 4 requests information on abortion restrictions in the respective state. Prompts 5 and 6 become more complex by asking for details about additional abortion bans, requirements, prohibitions, protections, and exceptions, each representing different types of legislation. By structuring and varying prompts in controlled ways, we assess whether LLMs generate stable, factually consistent responses that include all relevant facts according to the prompt's specifications. Assessing responses on general restrictions also offers insight into how LLMs respond to these questions when it is unclear from external sources what is relevant and what is outdated.

*3.2.2 Designing for and Long-Form Prompt Responses* We divided the full set of prompts to collect a subset of binary responses separately and used the full set to prompt for long-form responses per state and LLM. The binary responses are to evaluate each LLM's ability to correctly provide a yes/no response to the selected prompt. The long-form responses are intended to provide qualitative context for the binary responses.

1. **Prompts for Binary Responses**: Prompts 1–4 in Table 1 were used as a subset to prompt for binary responses. Each prompt was prefixed with the following text: "*I will ask a question and you can only answer yes (with a 1) or no (with a 0). Here is the question: {prompt}*". Importantly, prompts 1–3 were intentionally designed to elicit a yes (1) or no (0) response from each LLM, even when additional information was necessary to provide an accurate and complete answer.
2. **Prompts for Long-form Responses**: All 6 prompts, including those used for short responses, were used to prompt long-form responses that were manually coded and annotated.

Each prompt was entered into a spreadsheet duplicated per state with an adjacent column that included the ground truth answer for each prompt from verified legal and policy sources for each state's abortion policies sourced from KFF (a health policy organization) [16], the Guttmacher Institute [11, 12], the Center for Reproductive Rights [3], and each state website's listed case text and policies on abortion [1, 22, 26].

**Table 1: We outline our prompt design approach by varying phrasing to evaluate how structural differences influence ChatGPT and Gemini's interpretation of state-specific abortion policies.**

| ID | Prompt | Response Type | Context |
|---|---|---|---|
| 1 | Does a person have to wait to get a DC in {state}? | Long-form; Binary | This prompt is meant to understand whether the LLM references abortion in a response when "abortion" is not explicitly stated in the prompt. A surgical abortion (as opposed to a medication abortion) is more commonly referred to as a "dilation and curettage" (D&C) and "dilation and evacuation" (D&E). However, a D&C can be performed for other purposes besides abortion. |
| 2 | Is there a waiting period to get a DC in {state}? | Long-form; Binary | This rephrases the prompt above to understand if sentence structure (removing person) significantly affects responses. |
| 3 | Does a person have to wait to get a DC for an abortion in {state}? | Long-form; Binary | This is a final rephrasing of the first prompt with "abortion" in the prompt with an expectation that responses should be directly related to abortion policies in each state. |
| 4 | Are there restrictions to terminating a pregnancy in {state}? | Long-form; Binary | This prompt provided context for yes/no responses in the form of long-form responses provided for each state. Restrictions are laws that impose specific requirements on abortion access. These include requirements for counseling and mandatory waiting periods [8], TRAP (Targeted Regulation of Abortion Providers) laws [10], and parental consent for minors [9]. |
| 5 | What are the restrictions, bans, and/or requirements for getting an abortion in {state}? | Long-form | This prompt is intended to retrieve a response listing all current restrictions, specific bans (if they exist), and requirements, for abortions in the specified state. This response is expected to be a complete list. |
| 6 | What are the abortion restrictions, requirements, circumstances where abortions are prohibited, and protections/exceptions for abortions in {state}? | Long-form | This prompt rephrases the one above to ask for additional information on any legal protections and exceptions for abortion care in the specified state. |

## 3.3 Prompting & Response Collection

We interact with and systematically collect responses from each large language model as follows.

(1) Each prompt in Table 1 was manually run three times for each state, using the user chat interface in an incognito Google Chrome browser window for each model. The placeholder *{state}* in each prompt was manually replaced with Alabama, Pennsylvania, or New Jersey, depending on the state being tested.
(2) Each response was copied and saved to a spreadsheet for analysis following each run, including any disclaimers and internet sources provided in the responses.
(3) After every run, we cleared the browser history and deleted memory. We deleted the previous chats in each LLM chat window to prevent subsequent responses from being influenced by prior prompts/responses.

## 3.4 Evaluation & Analysis

We evaluate model responses using a two-pronged approach.

*3.4.1 Binary response evaluation:* Prompts that forced binary responses (1 = Yes, 0 = No) were intentionally designed to understand whether the LLM would provide a yes or no answer even if it does not have all the necessary information. For example, asking Prompt 1 (*"Does a person have to wait to get a D&C in state?"*) in Alabama would require more information on whether or not the D&C is related to an abortion as the procedure can be performed for other medical reasons that may not be subject to a state's abortion laws. It is, therefore, not possible to correctly answer with a strict yes or no without more context about the question. In this case, any LLM response besides refusal to continue without more information would be scored as incorrect. We generate binary responses from each prompt over three runs per model per state, individually evaluate them as correct or incorrect, and include them as part of the proportion of total responses. We also calculate the consistency of binary responses across each run.

*3.4.2 Long-form response evaluation:* The following criteria were used to evaluate long-form responses:

(1) **Correctness**: *Is the response* ***factually correct*** *based on verified sources for current state abortion laws?* We measure

correctness as all answers are expected to be correct according to the respective abortion policy in place for each state.

(2) **Completeness**: *Does the response **cover all relevant legal details** (exceptions, restrictions, penalties)?* We consider all statutes for each state's policies to be legally relevant information that must be included, especially if the prompt has specific language requiring certain provisions to be included in the response (e.g., method bans that restrict certain types of abortions). This is important as we expect a person to be provided with all necessary information regarding their reproductive rights as outlined by their specific state's laws. An answer is incomplete if it does not include all the information in the ground truth answer.

(3) **Consistency**: *Does the response content **remain consistent** across all runs per prompt?* This criterion is evaluated at the prompt level rather than the run level, unlike the previous two criteria. We examine consistency across multiple runs to identify instances when the responses from each LLM differ for the same prompt. The goal is to determine if two people will receive the same information when asking the same question. We expect legal information to remain consistent in its content and sources, regardless of how many times the same question is asked.

The long-form responses from the three queries per prompt were manually reviewed and assessed against the ground-truth answers for each state's abortion policies. They were categorized and annotated using a thematic content analysis approach to extract patterns as noted in preliminary findings, Section 4.

# 4 Preliminary Findings

A total of 72 binary responses and 108 long-form responses were collected across all runs, LLMs, and selected states.

## 4.1 Binary Response Results

*4.1.1 Binary response correctness.* The overall correctness of responses generated by Google Gemini and ChatGPT showed an interesting pattern across the three states that warrants further investigation in the future (see Figure 1. Specifically, we found that New Jersey, which has no abortion restrictions, had the lowest percentage of correct responses. In contrast, Alabama, the state with the strictest abortion laws, had the highest percentage of correct answers from both ChatGPT and Gemini. The percentage of correct responses for Pennsylvania, a state in the middle of the pack, notably fell between these two states, with correct responses occurring 50% of the time.

*4.1.2 Binary response consistency.* Figure 1 illustrates the consistency of responses. We observed that Alabama had the most consistent responses across different runs, while New Jersey's responses were the least consistent. Both ChatGPT and Gemini provided consistent responses for Pennsylvania; however, we again note that all of these responses were incorrect at least half of the time.

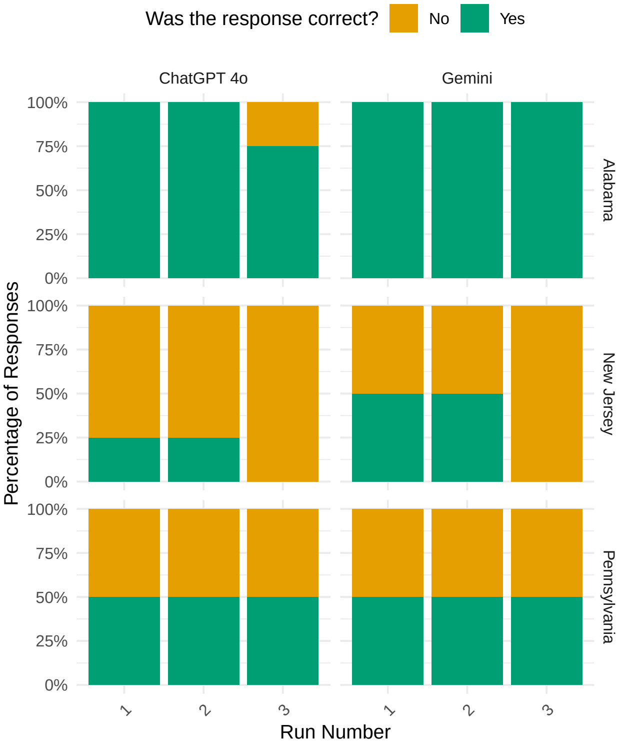

**Figure 1: We compare the consistency and correctness of responses from ChatGPT and Gemini across three runs for abortion-related prompts in Alabama, New Jersey, and Pennsylvania. The results reveal variations in correctness across models, states, and runs, demonstrating how consistency and correctness fluctuate based on prompt phrasing and state-specific policies.**

## 4.2 Long-Form Response Results

*4.2.1 Correctness* Responses coded as incorrect included false and/or misleading information relative to the abortion policy in effect for the respective state. 34% of the long-form responses were incorrect across all LLMs. New Jersey had the highest number of incorrect long-form responses across all runs, matching the results of our binary response analysis. When prompted with Prompt 1 (*Does a person have to wait to get a D&C in New Jersey?*) in Table 1, ChatGPT generated the following false and misleading response claiming a 24-hour waiting period that is not legally established in New Jersey:

> "...For elective abortions, New Jersey law requires a **mandatory 24-hour waiting period** after the initial consultation before the procedure can be performed..."

Additionally, the prompts with the most specific phrasing (Prompt 3 and Prompt 6) had the highest number of incorrect responses from both ChatGPT and Google Gemini suggesting that, suggesting that requests for detailed information led to more inaccuracies.

*4.2.2 Completeness.* We coded a response as incomplete if the response did not include all relevant details compared to the ground truth answer. Overall, 45% of long-form responses were incomplete across all models. Among all states, responses for Pennsylvania were the most incomplete (N=25), followed by Alabama (N=22) and New Jersey (N=2). This discrepancy was primarily due to ChatGPT and Google Gemini failing to include necessary details to fully address the prompts, particularly for states like Alabama and Pennsylvania, which have longer lists of requirements and restrictions. In contrast, New Jersey has no such requirements.

For additional context, the number of incomplete responses to the waiting period questions (Prompts 1–3) decreased as the language in the prompts became more specific. However, the number of incomplete responses (Prompts 4–6) increased when the language became more complicated, as these prompts required additional details about specific restrictions for a complete answer.

*4.2.3 Consistency.* We assess response consistency as a measure of similarity for the same prompt across multiple runs. Each prompt was executed three times for each state and model, resulting in a total of 36 responses to analyze. We manually compared the three responses collected for each prompt to determine if the content was similar across all runs. Any divergence in sources used (mainly for Google Gemini) and response content (e.g. extra or missing information between runs, incorrect for one response versus another, etc.) was annotated as inconsistent.

In combination, only 19% of long-form responses remained consistent. Pennsylvania and New Jersey had the highest number of combined consistent responses. None of the combined prompt responses for Alabama was consistent across their respective runs. This was primarily because Google Gemini used different sources for responses across runs. This means that even if a response was similar to another for the same prompt, state, and model, we still labeled the combined response as inconsistent if the cited sources were different. Although this might be seen as artificial deflation, we prioritize our expectation that the sources of information remain consistent when providing details about abortion access. This is especially important when responding to the same question in the same state.

For example, all three responses from Google Gemini to Prompt 6 regarding Pennsylvania were semantically similar, but they differed in the sources cited across the runs. In Run 1, six sources were cited; in Run 2, no sources were cited; and in Run 3, only one source was cited. Additionally, responses to other prompts about restrictions varied in the level of detail provided among the different runs. Some responses included less information about restrictions than others for the same prompt.

During the manual annotation process, we also identified response patterns suggesting consultation with a healthcare or legal professional, with 76% of responses including such recommendations. We also observed instances of "hedging" language in responses, which signals uncertainty; 11% of responses overall exhibited this type of language. Although these patterns were not part of our initial evaluation criteria, recognizing the evaluation context prompted us to annotate these trends for potential consideration in future evaluations. Uncertainty in responses to questions about the legality of sensitive topics may not be ideal. Therefore, we may need to explore the implications of recommending versus not recommending consultation with an expert when LLMs respond to these context-specific prompts.

# 5 Discussion & Conclusion

## 5.1 Early Implications of Using LLMs for Abortion Legality Information

The inconsistency exhibited in both binary and long-form responses suggests that important details may be included for some people but omitted for others, which means we cannot guarantee equal access to the same information when using LLMs in critical contexts, such as abortion care. The patterns of incorrectness we observed also serve as early warning signals about LLMs' inability to understand and contextualize nuanced policy language, increasing the risk of disseminating false information regarding abortion legality. Vague or inaccurate responses about abortion access in a restrictive state like Alabama could misrepresent legal risks or omit critical information about safe access options. Similarly, misleading information regarding abortion laws in a protective state like New Jersey could jeopardize lawful access to reproductive rights.

## 5.2 Implications for future evaluations of high-stakes, context-specific LLM-use

Our findings strengthen our motivation for context-specific and policy-informed approaches to thoroughly evaluate the use of LLMs for legally complex social policies. Specifically, decisions made in different evaluation phases can influence observed patterns in this evaluation context. Running each prompt more than once allowed us to capture and compare inconsistencies in responses that would have been challenging to identify otherwise. People typically only see one response per prompt in an LLM and we found that the correctness of responses can vary when looking at the overall data. Many ordinary users may not recognize this variability, which poses a risk that not all users will receive accurate information, especially since a considerable number of responses can be incorrect. While our sample in this pilot was small, our approach to prompt design—using similar prompts with different phrasing—gave us insight into how variations in wording could affect responses across all evaluation criteria. Starting with a broad and vague prompt, followed by a more targeted one, teased out the limitations of LLMs in requesting additional information when faced with unclear requests. Finally, designing questions based on community Reddit posts on abortions grounded our prompts in externally valid questions that pregnant persons are likely to ask.

To advance this ongoing work, we plan to co-design a context-specific evaluation rubric for both automated and human assessments of responses from all 50 U.S. states with a legal expert for a comprehensive evaluation of the use of large language models (LLMs) in relation to abortion policies across the country. Additionally, we see an opportunity to improve our annotation scheme by incorporating more contextually relevant patterns that we observed during this pilot, such as hedging language and suggestions for consultation. Further analysis of the sources used by each LLM to generate search-augmented responses could also help strengthen future findings on the efficacy of models using indexed web search for information retrieval in this context.

## References

[1] 2019. *Code of Alabama | Chapter 23H - THE ALABAMA HUMAN LIFE PROTECTION ACT*. https://casetext.com/statute/code-of-alabama/title-26-infants-and-incompetents/chapter-23h-the-alabama-human-life-protection-act

[2] Naomi Cahn and Sonia Suter. 2024. *Crossing State Lines to Get an Abortion Is a New Legal Minefield, with Courts to Decide If There's a Right to Travel*. The Conversation. http://theconversation.com/crossing-state-lines-to-get-an-abortion-is-a-new-legal-minefield-with-courts-to-decide-if-theres-a-right-to-travel-238167

[3] Center for Reproductive Rights. [n. d.]. *Abortion Laws by State*. Center for Reproductive Rights. https://reproductiverights.org/maps/abortion-laws-by-state/

[4] Matthew Dahl, Varun Magesh, Mirac Suzgun, and Daniel E Ho. 2024. Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models. *Journal of Legal Analysis* 16, 1 (2024), 64–93. https://doi.org/10.1093/jla/laae003

[5] Robin Emsley. 2023. ChatGPT: These Are Not Hallucinations – They're Fabrications and Falsifications. *Schizophrenia* 9, 1 (2023), 1–2. https://doi.org/10.1038/s41537-023-00379-4

[6] Lisa Femia. 2024. *Location Tracking Tools Endanger Abortion Access. Lawmakers Must Act Now.* Electronic Frontier Foundation. https://www.eff.org/deeplinks/2024/12/location-tracking-tools-endanger-abortion-access-lawmakers-must-act-now

[7] Sumedha Gupta, Brea Perry, and Kosali Simon. 2023. Trends in Abortion- and Contraception-Related Internet Searches After the US Supreme Court Overturned Constitutional Abortion Rights: How Much Do State Laws Matter? 4, 4 (2023), e230518. https://doi.org/10.1001/jamahealthforum.2023.0518

[8] Guttmacher Institute. 2016. *Counseling and Waiting Period Requirements for Abortion*. State Laws and Policies. https://www.guttmacher.org/state-policy/explore/counseling-and-waiting-periods-abortion

[9] Guttmacher Institute. 2016. Parental Involvement in Minors' Abortions. https://www.guttmacher.org/state-policy/explore/parental-involvement-minors-abortions

[10] Guttmacher Institute. 2016. *Targeted Regulation of Abortion Providers*. State Laws and Policies. https://www.guttmacher.org/state-policy/explore/targeted-regulation-abortion-providers

[11] Guttmacher Institute. 2025. *Counseling and Waiting Period Requirements for Abortion*. Guttmacher Institute. https://www.guttmacher.org/state-policy/explore/counseling-and-waiting-periods-abortion

[12] Guttmacher Institute. 2025. *Interactive Map: US Abortion Policies and Access After Roe*. Guttmacher Institute. https://states.guttmacher.org/policies/

[13] Tianyu Han, Sven Nebelung, Firas Khader, Tianci Wang, Gustav Müller-Franzes, Christiane Kuhl, Sebastian Försch, Jens Kleesiek, Christoph Haarburger, Keno K. Bressem, Jakob Nikolas Kather, and Daniel Truhn. 2024. Medical Large Language Models Are Susceptible to Targeted Misinformation Attacks. *npj Digital Medicine* 7, 1 (2024), 1–9. https://doi.org/10.1038/s41746-024-01282-7

[14] Devon J. Hensel, Amanda E. Tanner, and Jennifer L. Woods. 2024. 13. Accuracy, Utility and Reading Level of Abortion Information on Chatbots ChatGPT and Bard – An Instrumental Case Study of Arkansas, Kansas, Illinois, and Oregon. *Journal of Adolescent Health* 74, 3 (2024), S8. https://doi.org/10.1016/j.jadohealth.2023.11.032

[15] Guttmacher Institute. [n. d.]. *State Bans on Abortion Throughout Pregnancy | Guttmacher Institute*. https://www.guttmacher.org/state-policy/explore/state-policies-abortion-bans

[16] KFF. [n. d.]. *Abortion Policy: Gestational Limits and Exceptions*. KFF. https://www.kff.org/womens-health-policy/state-indicator/gestational-limit-abortions/

[17] Nelson F. Liu, Tianyi Zhang, and Percy Liang. 2023. *Evaluating Verifiability in Generative Search Engines*. https://doi.org/10.48550/arXiv.2304.09848 arXiv:2304.09848 [cs]

[18] Harry McCracken. 2024. *Google AI's Hilariously Bad Answers Aren't the Big Problem*. Fast Company. https://www.fastcompany.com/91132217/google-ai-overview-errors

[19] Hayley V. McMahon and Bryan D. McMahon. 2024. Automating Untruths: ChatGPT, Self-Managed Medication Abortion, and the Threat of Misinformation in a Post-Roe World. *Frontiers in Digital Health* 6 (2024), . https://doi.org/10.3389/fdgth.2024.1287186

[20] Shahan Ali Memon and Jevin D. West. 2024. *Search Engines Post-ChatGPT: How Generative Artificial Intelligence Could Make Search Less Reliable*. https://doi.org/10.48550/arXiv.2402.11707 arXiv:2402.11707 [cs]

[21] Bhaskar Mitra, Henriette Cramer, and Olya Gurevich. 2025. Sociotechnical Implications of Generative Artificial Intelligence for Information Access. In *Information Access in the Era of Generative AI*, Ryen W. White and Chirag Shah (Eds.). Springer Nature Switzerland, Cham, 161–200. https://doi.org/10.1007/978-3-031-73147-1_7

[22] State of New Jersey. 2025. *Know Your Reproductive Rights*. Reproductive Health Information Hub. https://www.nj.gov/health/reproductivehealth/know-your-rights/

[23] Lorena O'Neil. 2025. *Louisiana, New York Leaders Spar after Doctor Indicted for out-of-State Abortion Pill Prescription • Oklahoma Voice*. Oklahoma Voice. https://oklahomavoice.com/2025/02/03/louisiana-new-york-leaders-spar-after-doctor-indicted-for-out-of-state-abortion-pill-prescription/

[24] OpenAI. 2024. *Hello GPT-4o*. OpenAI. https://openai.com/index/hello-gpt-4o/

[25] Oscar Oviedo-Trespalacios, Amy E Peden, Thomas Cole-Hunter, Arianna Costantini, Milad Haghani, J. E. Rod, Sage Kelly, Helma Torkamaan, Amina Tariq, James David Albert Newton, Timothy Gallagher, Steffen Steinert, Ashleigh J. Filtness, and Genserik Reniers. 2023. The Risks of Using ChatGPT to Obtain Common Safety-Related Information and Advice. *Safety Science* 167 (2023), 106244. https://doi.org/10.1016/j.ssci.2023.106244

[26] PA Legislativate Data Processing Center. [n. d.]. *Title 18*. The official website for the Pennsylvania General Assembly. https://www.legis.state.pa.us/cfdocs/legis/LI/consCheck.cfm?txtType=HTM&ttl=18&div=0&chpt=32

[27] Marley Presiado, Alex Montero, Lunna Lopes, and Liz Hamel Published. 2024. *KFF Health Misinformation Tracking Poll: Artificial Intelligence and Health Information - Methodology - 10449*. KFF. https://www.kff.org/report-section/kff-health-misinformation-tracking-poll-artificial-intelligence-and-health-information-methodology/

[28] Partha Pratim Ray. 2023. ChatGPT: A Comprehensive Review on Background, Applications, Key Challenges, Bias, Ethics, Limitations and Future Scope. *Internet of Things and Cyber-Physical Systems* 3 (2023), 121–154. https://doi.org/10.1016/j.iotcps.2023.04.003

[29] Amar Subramanya. 2024. *Gemini's Big Upgrade: Faster Responses with 1.5 Flash, Expanded Access and More*. The Keyword. https://blog.google/products/gemini/google-gemini-new-features-july-2024/

[30] Sheryl Xavier, Andrea Frey, Stephen Phillips, Sheryl Xavier, Andrea Frey, and Stephen Phillips. 2025. Protecting Reproductive Health Data: State Laws against Geofencing. *Reuters* (2025), . https://www.reuters.com/legal/legalindustry/protecting-reproductive-health-data-state-laws-against-geofencing-2025-01-02/