GPT-4 Passes the Bar Exam

Daniel Martin Katz[1,2,3,4,*]**, Michael James Bommarito**[1,2,3,4]**, Shang Gao**[5]**, and Pablo David Arredondo**[2,5]

[1]Illinois Tech - Chicago Kent College of Law (USA)
[2]CodeX - The Stanford Center for Legal Informatics (USA)
[3]Bucerius Law School (Germany)
[4]273 Ventures (USA)
[5]Casetext, Inc. (USA)
[*]Corresponding Author: dkatz3@kentlaw.iit.edu

ABSTRACT

In this paper, we experimentally evaluate the zero-shot performance of a preliminary version of GPT-4 against prior generations of GPT on the entire Uniform Bar Examination (UBE), including not only the multiple-choice Multistate Bar Examination (MBE), but also the open-ended Multistate Essay Exam (MEE) and Multistate Performance Test (MPT) components. On the MBE, GPT-4 significantly outperforms both human test-takers and prior models, demonstrating a 26% increase over ChatGPT and beating humans in five of seven subject areas. On the MEE and MPT, which have not previously been evaluated by scholars, GPT-4 scores an average of 4.2/6.0 as compared to much lower scores for ChatGPT. Graded across the UBE components, in the manner in which a human tast-taker would be, GPT-4 scores approximately 297 points, significantly in excess of the passing threshold for all UBE jurisdictions. These findings document not just the rapid and remarkable advance of large language model performance generally, but also the potential for such models to support the delivery of legal services in society.

Introduction

It is difficult to imagine a professional field for which natural language is more integral than the law. As part of their daily activities, legal professionals like judges, regulators, legislators, and lawyers spend countless hours consuming and/or producing a wide variety of legal documents. The document types are varied but include legal texts such as statutes, regulations, judicial decisions, contracts, patents, briefs, opinion letters, memos, and other related materials.[1,2]

Legal language is notoriously complex,[3,4,5] and the ability to interpret such complex documents often requires years of study. Indeed, part of the charge of legal education is, in fact, a linguistic immersion program where students are trained to parse both the syntactic and semantic nuances of various legal texts.[6] There are many sources of complexity in legal language: for example, "words like 'security' that have common meaning in normal language often have different, context-specific meanings in legal language. Many words that do not occur at all in normal language, like 'estoppel' or 'indemnitor,' occur regularly in legal corpora."[7] This semantic depth and breadth is challenging for those not otherwise familiar with the legal lexicon. The public, for example, is quite aware of the linguistic gap between general language and legal language, referred to by many as "legalese."

The complexity of the law[8,9,10,11] imposes real consequences for many individuals and organizations.[12,13] In part due to complexity, legal systems have struggled to assist with the quantity, quality, and accessibility of legal services demanded by society.[13,14,15] A technology-based force multiplier[15,16] is arguably needed to help support the high cost and unmet demand for legal services.[17,18] Yet, in order for technology systems to meet this need, they must confront the nuances of legal languages and the difficulties of complex legal reasoning tasks.[19] Unfortunately, from a historical perspective, computational technologies have struggled not only with natural language processing (NLP) tasks generally, but, in particular, with complex or domain-specific tasks like those in law.

There is promise on the horizon, however; state-of-the-art performance in NLP has advanced substantially over the last decade, largely driven by advances in computer hardware, data availability, and neural techniques. Indeed, cutting-edge work within the field of NLP has recently undergone a rapid transition where classical NLP methods have been supplanted by neural based methods.[20,21] While neural techniques have a long history,[22,23,24,25] current modeling approaches generally trace their lineage to the arc from shallow embeddings trained on CPUs to the current transformer-based architectures optimized for purpose-built, distributed GPU/TPU infrastructure.[26–37]

While there is an increasing number of generally-accessible Large Language Models (LLMs), the most well-known of these are from OpenAI's family of Generative Pre-trained Transformer models, commonly referred to as GPT.[34,38,39,40,41] In November 2022, OpenAI released a chat interface to a version of its "GPT-3.5" models, colloquially known as ChatGPT, which reportedly resulted in over 1M user sign-ups within six days of release and over 100M users in the first ninety days. As described by OpenAI, GPT-4 is "a transformer-style model pre-trained to predict the next token in a document, using both publicly available data (such as internet data) and data licensed from third-party providers. The model was then fine-tuned using Reinforcement Learning from Human Feedback (RLHF)."[38] While this family of models encompasses a range of tasks, sizes, and training techniques and continues to expand, all models are generally trained using reinforcement learning or supervised fine-tuning on billions of tokens and parameters.

NLP models have progressed in the legal domain[19,42,43] with increasing application of neural techniques on specific legal tasks.[44,45,46] Several recent papers have demonstrated meaningful zero-shot progress on a variety of applied tasks,[2,47,48,49,50,51] suggesting further potential for application as state-of-the-art improves.

Recognizing the advancing capabilities of large language models, we sought an exemplary challenge to demonstrate this potential to the legal domain. Thus, in recent prior work,[7] a subset of the authors demonstrated the near-passing zero-shot performance of TEXT-DAVINCI-003 on the multiple choice component (MBE) of the Uniform Bar Exam – a task which requires both comprehensive domain knowledge and advanced semantic and syntactic command of the English language. While no prompts or parameters met a "passing" level, the rate of performance increase from TEXT-DAVINCI-001 to TEXT-DAVINCI-003 strongly suggested that passing performance could "occur within the next 0-18 months."[7] In this paper, we demonstrate that this time has come for not only the multiple choice component (MBE), but also the essay (MEE) and performance test (MPT) components of the UBE. As demonstrated by the zero-shot performance results we report herein, GPT-4 can "pass the Bar" in all UBE jurisdictions.

The Uniform Bar Exam

Description of the Uniform Bar Exam

The vast majority of jurisdictions in the United States require the completion of a professional licensure exam ("The Bar Exam") as a precondition to practice law. The Bar Exam is a notoriously-challenging battery of tests designed to evaluate an applicant's legal knowledge and skills. Successfully passing the Exam requires that an examinee can discern challenging factual and legal scenarios, understand and apply legal principles, and both consume and produce complex legal language.

In order to sit for the exam, the typical applicant must complete at least seven years of post-secondary education, including completion of a four-year undergraduate degree, followed by matriculation and graduation from a law school accredited by the American Bar Association. In addition to these years of education, most applicants also invest substantial amounts of time and money into specialized test-taking courses.[52] Despite this effort and investment, roughly one in five test-takers is unable to pass the Exam on their first attempt.

Attorney licensure is a topic governed by the states, typically through rules promulgated at the direction of state supreme courts.[53] Thus, each state is responsible for selecting its own requirements and methods of exam administration. Notwithstanding such broad authority, many states have selected to standardize their requirements. Over the past decade, more jurisdictions have chosen to participate in the Uniform Bar Exam (UBE).[52,54] Despite this push toward greater uniformity, however, there are often additional requirements, even within states that have adopted the UBE, such as the Multistate Professional Responsibility Examination (MPRE) or state-specific subject matter areas. In this paper, we address only the UBE as produced by the National Conference of Bar Examiners (NCBE). The core UBE components, outlined in Table 1 below, are the Multistate Bar Exam (MBE), the Multistate Essay Exam (MEE) and Multistate Performance Test (MPT).

As shown in Table 1 and discussed in detail in Appendix I (pp.15 et seq.), the UBE is a 12-hour exam taken over two days, with the MPT and MEE administered on Day One while the MBE is administered on Day Two. The Uniform Bar Exam is scored on a total scale of 400 points, with the scores from all three sections scored together. In general, there are no minimums required for a specific component of the exam, as a strong score on one component can help an examinee overcome a weaker score on another component. As displayed in Table 8, a combined score of 266 points is enough to pass in jurisdictions such as Illinois, New York and the District of Columbia, while a score of 270 points would pass in the vast majority of states which use the UBE.

UBE Component	Total UBE Points	Questions	Time	Time Per Question
Multistate Bar Exam (MBE)	200 Points	200 Questions (Multiple Choice)	6 Hours	1 min 48 sec
Multistate Essay Exam (MEE)	120 Points	6 Questions (3-4 Subquestions)	3 Hours	30 minutes
Multistate Performance Test (MPT)	80 Points	2 Questions (3-4 Subquestions)	3 Hours	90 minutes

Table 1. Summary of Uniform Bar Exam (UBE) Components

Data & Methods

Data

The primary focus of the NCBE is on the construction of exams for use on a nationwide basis. The NCBE exams are developed in an institutional context by the organization's staff and advisors, who have many years of experience designing, scoring, and calibrating these exams across U.S. jurisdictions.

As noted earlier, the UBE has three separate components: the MBE, MEE and the MPT. In order to analyze whether GPT-4 could pass the Bar Exam, we collected relevant materials for each of the three separate UBE components. For the MEE and the MPT, we collected the most-recently-released questions from the July 2022 Bar Examination. These questions are readily available through the websites of many state bars. The July 2022 MEE exam features six questions, covering Evidence, Contracts, Corporations, Trusts, Civil Procedure, and Real Property. The two questions for July 2022 MPT required test-takers to (i) draft a memo in the context of a domestic relations matter with a series of choice of law issues and (ii) construct an objective memo focused on questions of criminal law and legal ethics.

The MBE questions used in this study are official multistate bar examination questions from previous administrations of the UBE.[55] The MBE full-length exam we use is subject-weighted in near-equal proportion across the seven core subject matter areas. While the exact sequence of questions administered is not identical to any actual exam as administered, it has been described by the NCBE itself as "the first [MBE Complete Practice Exam] from NCBE to mimic a full-length MBE."[55] While we are not able to release the MBE questions, the questions can be purchased directly from an NCBE authorized reseller.

Links to access both the full length MEE and MPT questions, as well as their representative "good" answers, are available in the online repository. These representative good answers are made available by state bar associations and reflect actual MEE and MPT answers produced by real examinees. These answers are described as neither 'average passing answers nor are they necessarily perfect answers." We would suggest that the interested reader reviews these representative 'good' answers side-by-side with our model outputs.

Methods

In prior work, a subset of the authors implemented and described frameworks for multiple-choice assessment on the Bar Exam[7] and an open-ended assessment for task-based simulation in the CPA Exam.[56] While our early access to GPT-4 resulted in a number of temporary constraints, we follow the approach outlined in prior work to the extent possible, including input-formatting conventions, prompt styles, and model parameters. As GPT-4 becomes generally available, we intend to further systematize the methods below and update this publication with additional sections related to prompt and parameter variation across models.

MBE

As described above, the MBE section is administered as a multiple-choice question exam. For each "session," in which a model sits for a complete exam, each question is sent to the model with a formatted prompt. All model responses are logged, including all available API metadata, and stored for subsequent review, and all grading is automated. Experiments involved multiple runs for each set of prompts and parameters. Details of these prompts and parameters are available in the online repository.

MEE and MPT

Unlike the MBE, both the MEE and MPT are administered as open-ended exams. These exams combine background and reference material with one or more sub-questions in a manner similar to the task-based simulations in other professional licensing examinations. As in prior work, we standardize the formatting of these materials and questions to generate plain text versions of the exam. For each session, each question is sent to the model and its response is logged for subsequent grading and review. Experiments currently include a restricted set of prompts and parameters subject to expansion upon general availability of GPT-4. Details of current prompts and parameters are available in the online repository.

Results

Multistate Bar Exam (MBE) Results

Overall MBE Results

We administered the MBE to most available GPT models, including not only GPT-4, but also also an alpha version of ChatGPT, TEXT-DAVINCI-003, TEXT-DAVINCI-001, TEXT-CURIE-001, TEXT-BABBAGE-001, and TEXT-ADA-001. The per-model accuracy averaged across all model runs is presented in Table 2 and visualized in Figure 1.

Model Name	MBE Accuracy
GPT-4	75.7%
CHAT-DAVINCI-003-ALPHA	49.2%
TEXT-DAVINCI-003	45.1%
TEXT-DAVINCI-001	34.2%
TEXT-BABBAGE-001	23.9%
TEXT-CURIE-001	22.8%
TEXT-ADA-001	8.5%
GPT-2	N/A

Table 2. Accuracy of GPT Models on the Multistate Bar Exam (MBE)

Table 2 demonstrates the increase in MBE performance between even the most recent members of the GPT family and the alpha version of GPT-4 we used for this study. GPT-4 delivers a 26.5% increase in the accuracy over ChatGPT, the previously best performing model. In addition, GPT-4's MBE score is not only more than 15% above the minimum passing threshold, but also outperforms the average human test-taker by more than 7%.

Table 2 and Figure 1 highlight the broader progression of GPT models since 2019. Some of the earliest models such as GPT-2[40] are unable to consistently process prompts, while later models such as Curie (TEXT-CURIE-001), Babbage (TEXT-BABBAGE-001) and Ada (TEXT-ADA-001) were unable to obtain performance above that of the statistical guessing rate of 25%. GPT-3 (TEXT-DAVINCI-001) (initially released in 2020)[34] was the first model to consistently outperform statistical chance. The previously best-available models, ChatGPT (CHAT-DAVINCI-003) and GPT-3.5 (TEXT-DAVINCI-003), performed just under the 50% accuracy level. As displayed in Figure 1, the benchmarked growth on this task is reminiscent of similar non-linear improvements witnessed within other recent benchmarks,[57,58,59] where over the course of a relatively short period of time leading language models were able to surpass the performance of experts on previously-untouchable tasks.[60]

MBE Results by Legal Subject Area

While GPT-4's performance on the MBE exceeds the passing rate and the performance of the average human test-taker, it is also interesting to explore its performance within individual legal subjects. Human test-takers perform differentially across
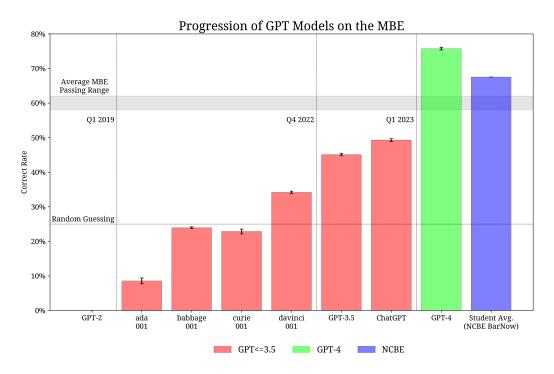
Figure 1. Progression of Recent GPT Models on the Multistate Bar Exam (MBE)

the various topics within the UBE. The NCBE Bar Now Platform maintains statistical information regarding student average performance by topic within MBE questions. Table 3 offers both the NCBE Bar Now average accuracy by question category as well as the overall approximate national average MBE performance for recent test-takers. Among other things, Table 3 highlights the nature of the challenge which the MBE presents to test-takers as the average student answers more than three in ten questions incorrectly.

Whether it involves a subject matter expert or a model, there are subjects where performance will likely differ, and, unsurprisingly, some subjects may require more knowledge or be more complex. All of this is conditioned on prior exposure to a topic (data), the nature of that exposure (feedback), and the ability to weight those prior exposures (calibration). While our understanding of LLMs is still nascent[61,62,63] and we do not fully understand why GPT-4 performs differently by substantive topic, it is likely that the prospect of future performance improvements will, in part, hinge on identifying opportunities for additional topically relevant data, feedback and calibration.

In Table 3, we report the MBE results by legal subject area and visualize those same results in Figure 2. Both the table and figure reveal that, while GPT-4's performance may vary by subject, it meets or exceeds the approximate passing threshold in all seven subject matter areas. In addition, it outperforms the average NCBE BarNow test-taker in five out of seven categories (*Civil Procedure*, *Contracts, Criminal Law and Procedure*, *Evidence* and *Real Property*).

GPT-4 is now pareto dominant over all prior versions of GPT (sometimes by a very large margin). *Contracts* and *Evidence* are the topics with the largest overall improvement. GPT-4 achieves a nearly 40% increase over ChatGPT in *Contracts* and a more than 35% raw increase in *Evidence*. *Civil Procedure* is both the worst subject for GPT-4, ChatGPT and human test-takers. However, *Civil Procedure* is a topic where GPT-4 was able to generate 26% raw increase over ChatGPT. This increase pushed GPT-4 into both the passing threshold and beyond the performance of the average NCBE BarNow test-taker.

An observant reader may notice that the rank ordering of subject matter performance for TEXT-DAVINCI-003 on this MBE Exam differs from the MBE Exam tested in prior work.[7] While a comprehensive analysis of performance across multiple MBE exams is in progress, several factors could explain these differences. First, it is important to note that headline 50.3% result in prior work reflect best prompt and parameter, not the averages calculated across all prompts and parameters for TEXT-DAVINCI-003; therefore, the reader should compare these figures to the 42-46% accuracy reported in the prior work's tables. Second, in this research, we use the most recently-published MBE Exam, which the NCBE describes as "the first from

NCBE to mimic a full-length MBE." Differences in the distribution of questions, the distribution of question difficulty, or the design of the Exam over time may account for additional variation. Final results for this paper will inlcude three MBE simulated exams, supporting additional analysis and discussion of this topic.

Legal Subject Area	GPT-4	ChatGPT	GPT-3.5	NCBE Student Avg.
Civil Procedure	61.1%	34.9%	39.1%	59%
Constitutional Law	69.4%	54.2%	45.9%	72%
Contracts	88.1%	48.5%	37.0%	70%
Criminal Law and Procedure	81.1%	46.3%	49.8%	71%
Evidence	85.2%	49.8%	45.7%	65%
Real Property	79.7%	67.8%	52.9%	65%
Torts	64.9%	43.2%	45.7%	71%
Average Accuracy	75.7%	49.2%	45.1%	68.0%

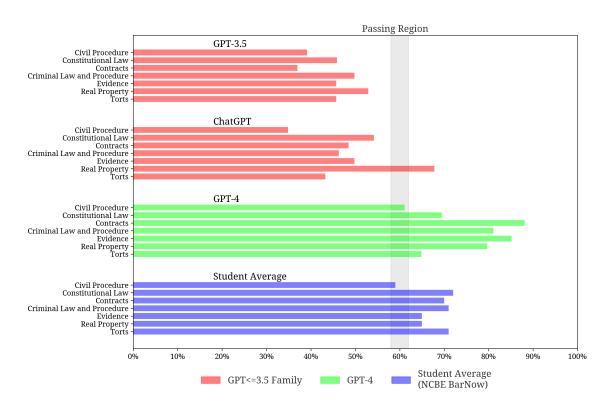**Table 3.** Summary of Performance by Legal Subject Area

Figure 2. Progression of Recent GPT Models by Legal Subject Area

Non-Entailment MBE Results
Building up the broader concept of textual entailment,[64,65,66] earlier work studying Bar Exams treated "the relationship between the question and the multiple-choice answers as a form of textual entailment"[67] where the ability to identify wrong answers (non-entailment) is differentiated from the ability to identify the correct answer (entailment). Intuitively, this is related to the classic test taking strategy of eliminating clearly erroneous answers. While earlier models are able to undertake this

elimination task to a fairly reasonable extent, it is with regard to this non-entailment task that GPT-4 shows its particular strength.

Table 4 reproduces the model accuracy (entailment) and reports the "Top 2" and "Top 3" accuracy (non-entailment). Across the various topics, GPT-4 generally achieves strong "Top 2" performance (i.e., ability to reduce the number of likely answers from four to two). GPT-4's performance on the "Top 2" task is roughly in line with the "Top 3" performance of prior models. For example, for the *Contracts*-related problems, GPT-4 is able to identify the right answer within its "Top 2" choices in nearly 97% of instances. By contrast, for *Torts*, in roughly 15% of instances, GPT-4 is unable to eliminate even a single wrong answer (which is to say it actually ranks the correct answer least likely to be correct). This entailment versus non-entailment perspective is one way to consider potential sources of future model improvement. Overall, GPT-4's second best answer is highly correlated with correctness for all subjects other than *Criminal Law and Procedure*, and its overall performance in rank-ordering responses demonstrates state-of-the-art capabilities for information retrieval tasks.

Model	GPT-4 Accuracy	GPT-4 Top 2	GPT-4 Top 3	ChatGPT Accuracy	ChatGPT Top 2	ChatGPT Top 3	GPT-3.5 Accuracy	GPT-3.5 Top 2	GPT-3.5 Top 3
Civil Procedure	61.1%	82.2%	95.4%	34.9%	55.4%	89.3%	39.1%	49.6%	76.1%
Constitutional Law	69.4%	86.3%	98.6%	54.2%	73.2%	93.5%	45.9%	59.8%	83.7%
Contracts	88.1%	96.7%	99.8%	48.5%	65.5%	88.9%	37.0%	59.8%	78.0%
Criminal Law and Procedure	81.1%	84.6%	97.3%	46.3%	77.8%	92.8%	49.8%	63.9%	77.7%
Evidence	85.2%	91.5%	99.8%	49.8%	69.8%	89.6%	45.7%	60.3%	80.7%
Real Property	79.7%	87.9%	100.0%	67.8%	76.2%	94.4%	52.9%	69.0%	78.0%
Torts	64.9%	73.6%	85.5%	43.2%	55.6%	73.0%	45.7%	58.4%	66.8%

Table 4. Summary of Non-Entailment Performance by Legal Subject Area

Multistate Essay Examination (MEE) Results

Many would consider the construction of essays to be a more difficult task than answering multiple choice questions, particularly for a computational system. While selecting an answer from a difficult but otherwise already defined list of choices is certainly a challenge, it is arguably a much more challenging task to read and identify key issues in a one page prompt and then draft a fulsome essay on a complex subject matter.

We experimented with a variety of prompts, hyperparameter settings, and question formatting techniques. Among other things, this initial analysis revealed the clear benefit of question segmentation. Namely, both GPT-4 and ChatGPT delivered more detailed results when both were provided with a single MEE subquestion to consider. Thus, for each MEE question, we made one small modification from the problem as presented. Namely, we ran each MEE subquestion one at a time and lightly corrected the language so as to craft the question in the form of a complete sentence. As an example, consider the questions posed in Figure 5. Then, imagine them delivered together with the vignette one question at a time to each of the models. Access to the exact prompts as administered will be made available in the online repository.

Two of this study's authors, a tenured law professor and an attorney licensed in multiple jurisdictions, reviewed the model output and assigned preliminary scores to each of the MEE questions. We also solicited additional analysis from peers who were provided with blinded samples of the model's responses, who independently came to assessments that met or exceeded ours. While we recognize there is inherent variability in any qualitative assessment, our reliance on the state bars' representative "good" answers and the multiple reviewers reduces the likelihood that our assessment is incorrect enough to alter the ultimate conclusion of passage in this paper.

In Figures 6 - 12, we reproduce output for the July 2022 MEE Evidence Question for three models (GPT-4, ChatGPT and GPT-3.0). Although the primary focus of our analysis is the comparison between more recent GPT models, in Figure 12, we reproduce the GPT-3.0 MEE model's answer for the overall evidence question. Similar to the trend previously shown in Figure 1, we believe that observing the broader progression across these models helps highlight the underlying increase in capabilities. The pattern revealed in Figure 1 is also true for the MEE and the side-by-side comparison of output should help the interested reader observe this arc. In addition, the side-by-side comparison of the model output from the MEE Evidence question also reveals some of the broader patterns reflected across the balance of the MEE essay problems.

Starting with oldest model first, GPT-3.0 produces very thin output in response to a prompt which specifically directs it to produce a fulsome answer. As shown in Figure 12, GPT-3.0 can vaguely recite some of the relevant principles and rules but

does not properly connect those principles to the facts and consistently reaches improper legal conclusions. In the context of this complex legal problem, GPT-3.0 is far below the mark.

While the weaknesses in the GPT-3.0 output are clear, the comparison between GPT-4 and ChatGPT requires a more nuanced analysis. In Figure 6 and Figure 9, we reproduce the output for the first subquestion within the July 2022 MEE Evidence problem. An initial review of that output reveals that ChatGPT actually produces a slightly longer response than GPT-4. At a substantive level, however, and particularly as compared to GPT-4, ChatGPT is deficient in several important ways. Unlike GPT-4, ChatGPT fails to properly identify all four prongs of Rule 702 of the Federal Rules of Evidence (FRE). This results in a failure to discuss Rule 702(a). Yet, under FRE Rule 702, all four prongs (including 702 (a)) must be satisfied in order for the expert's testimony to be allowed. In addition to missing this important discussion, ChatGPT begins to intellectually meander and provides an analysis of FRE Rule 403. While not totally unrelated, this is a rule which is out of scope for this question. By contrast, GPT-4 does an overall better job of addressing the question presented by properly citing the relevant law, connecting the law to the facts and otherwise staying on topic.

This basic dynamic is replicated across not only the balance of the MEE Evidence question but also across much of the MEE model output. As an additional example, consider the second subquestion on the MEE Evidence problem (Figures 7 and 10) where ChatGPT fails to address a relevant rule (FRE 403) and instead focuses significant attention on non-relevant rules (FRE 701 and 702). It is arguably not proper to call this a model hallucination in the sense that these are real rules which are, in fact, somewhat related to the question. These topics are simply out of scope with respect to the specifics of the question that was posed. GPT-4, by contrast, correctly discusses both FRE Rule 403 and Rule 404(b) and does not devote attention to extraneous issues.

While GPT-4 performs well on many questions, its output is not completely free of errors. In the three sub-questions which we assign the lowest scores, GPT-4 produces several notable errors. First, it has difficulty calculating the distribution of assets from a testamentary trust which has been deemed to be invalid. Next, it fails to grasp the call of the question and provides an incorrect answer on a civil procedure question regarding diversity jurisdiction after the joinder of a necessary party. Finally, GPT-4 provides improper analysis on a real property (real estate) subquestion regarding both the proper designation of a Future Interest and the application of the Rule Against Perpetuities.

It should be noted that several of these topics where GPT-4 struggles are also areas where law students and bar examinees would also likely struggle. In particular, the Rule Against Perpetuities is considered by many to be among the most difficult issues in all of law. In addition, it should be mentioned that most real life examinees who otherwise pass the Bar Exam are unable to complete an end-to-end MEE that is free from errors. Overall, even in problems for which we assign a lower grade, GPT-4 is often able to deliver a partial answer, such as identifying some, but not all, relevant legal principles or providing reasonable discussion of some of the facts that are relevant to the legal question.

MEE Question Subject	GPT-4	ChatGPT
MEE 1 - Evidence	5.0 / 6.0	3.7 / 6.0
MEE 2 - Contracts	4.2 / 6.0	3.1 / 6.0
MEE 3 - Corporations	4.4 / 6.0	3.0 / 6.0
MEE 4 - Trusts / Estates	3.9 / 6.0	2.5 / 6.0
MEE 5 - Civil Procedure	3.5 / 6.0	2.8 / 6.0
MEE 6 - Real Property	4.2 / 6.0	2.7 / 6.0
Overall Score	4.2	3.0

Table 5. Summary of Performance by Multistate Essay Examination (MEE) Question Category

Due to space constraints, we reproduced only a single MEE question and output within the paper. However, we would once again direct the interested reader to the online repository, which features prompts, and model outputs for each of the July 2022 MEE questions, links to representative 'good' answers and other useful information. Overall, as presented in Table 5, our final MEE grade is 4.2 out of 6 points for GPT-4 and 3 out of 6 points for ChatGPT. Most jurisdiction leverages the six-point scale, where a score of four or higher is generally considered passing. We encourage the interested reader to review all model output in the Appendix and to reach their own conclusions regarding the quality of the MEE answers produced by the respective models.

Multistate Performance Test (MPT) Results

As discussed earier, the July 2022 MPT features two substantive problems: one question focused upon a complex family law matter with embedded choice of law issues and another question focused on a mixture of criminal law and legal ethics issues. For purposes of comparative analysis, we evaluate the performance of GPT-4 relative to earlier models, such as ChatGPT. As many have noted, prior models were unable to handle longer documents in zero-shot tasks due to their token limits. The July 2022 MPT featured two questions, each over the 4,096 token limit - 5,297 tokens for MPT-1 and 5,188 tokens for MPT-2. However, thanks to assistance from OpenAI, we were able to use a version of ChatGPT that has a wider context window of 8,193 tokens, and could thus accommodate the length of the overall materials. This longer context window was critical to our analysis.

Similar to the question-by-question approach that we undertook for the MEE, for both MPT-1 and MPT-2, we presented the problems as subquestions. First, we placed the instructional memo (describing the task to be undertaken) at the end of the prompt (after 'the File' and 'the Library'). Second, we reduced the memo to a single subquestion for each prompt. Thus, for MPT-1 we ran the model four times, one time per subquestion.

Both of the models we considered produced long-form answers to the respective problems, but there are some fairly straightforward differences in the quality of output produced. For purposes of analysis and discussion, we will focus upon the July 2022 MPT-1 problem. Figure 13 replicates the MPT-1 instructional memo while Figures 14 - 21 reproduce the MPT-1 model output from both GPT-4 and ChatGPT.

Following an approach comparable to our process for the MEE, we reviewed the MPT output produced by each model and evaluated it against the representative 'good' answers. We then assigned scores to both of the MPT questions from the July 2022 exam. These results are reported in Table 6. As we noted in the grading of the MEE, we recognize that there is a subjective aspect to this sort of analysis and so we encourage the interested reader to review the MPT output contained in Figures 14 - 21 as well as the output from MPT-2 in the online repository and reach their own conclusions.

Beyond the numeric results displayed in Table 6, our grading procedure revealed some clear and important distinctions in the quality of the output as between GPT-4 and ChatGPT. For example, in the first subquestion of MPT-1, both GPT-4 and ChatGPT produce a lengthy and otherwise reasonable looking answer. Similar to differences in the MEE scores, the differences in quality manifest themselves within the deeper details of the problem. In Figure 18, ChatGPT correctly identifies that it should address the Restatement of the Conflicts of Law as this is core to the answer. However, it incorrectly cites §6 as the proper source when it should cite §283 as the source of the 'significant relationship' test.[1] More fundamentally, despite partial success on the problem, ChatGPT ultimately draws the incorrect legal conclusion. Columbia law and not Franklin law will likely govern the question of annulment. In Figure 14, GPT-4 not only correctly identifies that Columbia Law is controlling but provides reasonable arguments as to why Columbia should be selected under the 'significant relationship' test.

Subquestion three of MPT-1, as displayed in Figures 16 and 20, provides another example of the distinction in quality between GPT-4 and ChatGPT. Here GPT-4 correctly distinguishes between the ability of the Franklin court to annul the marriage and its inability to dispose of the parties' property. The Franklin Court cannot, without personal jurisdiction over Tucker, take action against a property outside its borders. Ms. Tucker is not a resident of Franklin and the property is not in Franklin. Her source of contact with Franklin is limited to the fact that her soon to be ex-husband moved there. ChatGPT fails to distinguish between the circumstances and asserts authority to the Franklin Courts that they do not possess. These type of patterns repeat themselves over the remainder of both MPT-1 and MPT-2.

It was our hypothesis that the MPT would prove to be more challenging to GPT-4 than the MEE. While the MEE requires

---

[1]It is challenging because §283 actually does in part invoke §6 but it is not the source of the 'significant relationship' test. The sentence as drafted by ChatGPT contains a circular reference (is an infinite loop).

MPT Question	GPT-4	ChatGPT
MPT 1 - "Hixon Marriages"	4.2 / 6.0	3.0 / 6.0
MPT 2 - "In re Briotti"	4.1 / 6.0	2.5 / 6.0
Overall Score	4.2	2.8

Table 6. Summary of Performance on Multistate Performance Test (MPT) Questions

the examinee to answer substantive law questions from any of the potential topics within the list of bar exam subjects, the MPT is a somewhat different type of exercise. It is a lawyering exercise where the materials as provided define the relevant universe of information. As the NCBE describes it, "the MPT is not a test of substantive law; the Library materials provide sufficient substantive information to complete the task." Consequently, the MPT requires a suspension of knowledge, whereby the examinee must, for the period of the test, imagine themselves in a jurisdiction that may contradict their actual knowledge of real law. Indeed, the instructions provided with the test remind the test-taker that even "if the cases appear familiar to you, do not assume that they are precisely the same as you have read before. Read them thoroughly, as if they all were new to you."

We were concerned that this "suspension of a broader knowledge," or ability to work within the four corners of the exam material, would prove challenging for any member of the GPT family (even GPT-4). Thus, we were somewhat surprised at the quality of the output which was generated. Both GPT-4 and, to a lesser extent, ChatGPT were able to largely avoid the trap of citing legal principles, cases, or other materials which would otherwise appear to be on point but would not be responsive to the requirements of the MPT.

Combined Results and Comparison to the UBE Passing Threshold

The Uniform Bar Exam (UBE) has three components (MBE, MEE and MPT) which are typically, but not always, weighed using the approach highlighted in Table 1. As displayed in Table 8, different jurisdictions impose different UBE passing score thresholds ranging from 260 (for states such as Alabama and Minnesota) to 273 points (for Arizona). The vast majority of UBE states impose a minimum UBE passing score threshold between 260 and 270.

In Table 7, we combine all the analysis conducted above to offer an overall UBE score for both GPT-4 and ChatGPT. GPT-4 obtains an overall UBE Score of 297 points[2] while ChatGPT obtains a score of 213 points.[3] Although GPT-4 obtains within-category passing-level scores for both the MEE and MPT, its high MBE percentile provided it with substantial latitude; GPT-4 would likely pass even with a much lower MEE or MPT score in some or all jurisdictions. In total, our analysis highlights that GPT-4 has indeed passed the Bar and has done so by a significant margin.

Conclusion

Less than three months ago, we predicted that passage of the MBE would likely occur "in the next 0-18 months," and in this paper, we provide preliminary evidence of GPT-4's zero-shot passage of the entire UBE. The exam, which includes both multiple-choice and open-ended tasks testing theoretical knowledge and practical lawyering, has long been viewed as an insurmountable summit for even domain-specific models. This assumption no longer holds; large language models can meet the standard applied to human lawyers in nearly all jurisdictions in the United States by tackling complex tasks requiring deep legal knowledge, reading comprehension, and writing ability.

---

[2]Best prompt and/or hyperparameter combination on the MBE would push this score to 298 or higher. Here we report the MBE average of 75.7% which composites to a 297.

[3]For the reader who is not familiar with these issues, it might be difficult to contextualize these UBE scores. Using a percentile chart from a recent exam administration (which is generally available online), ChatGPT would receive a score below the 10th percentile of test-takers while GPT-4 would receive a combined score approaching the 90th percentile of test-takers. The overall national minimum passing threshold is 260 for several states such as Alabama, Minnesota and Missouri. The highest passing threshold is 273 for Arizona. See Table 8 for additional information.

UBE Component	GPT-4	ChatGPT
Multistate Bar Exam (MBE)	157 Points	116 Points
Multistate Essay Exam (MEE)	84 Points	60 Points
Multistate Performance Test (MPT)	56 Points	37 Points
Overall Score	297 Points	213 Points

Table 7. Summary of Overall Performance on Uniform Bar Exam (UBE)

Most notably, the results documented in this paper reflect only zero-shot model behavior. While many real tasks in industry and society broadly require more knowledge and ability than tested on the Exam itself, there is significant opportunity to advance the performance of large language models through external queries, scratchpads, chain-of-thought prompting (CoT), or one of the many other techniques that emerges weekly.[68,69,70,71] Our findings only highlight the floor, not the ceiling, of future application.

As the demand for better, faster, and more affordable legal services is only increasing in society, the need for supporting technology is becoming more acute. Further research on translating the capabilities of LLMs like GPT-4 into real public and private applications will be critical for safe and efficient use. GPT-4, like prior models, may still hallucinate sources, incorrectly interpret facts, or fail to follow ethical requirements; for the foreseeable future, application should feature "human-in-the-loop" workflows or similar safeguards.[4] However, it appears that the long-awaited legal force multiplier is finally here.

References

1. Coupette, C., Beckedorf, J., Hartung, D., Bommarito, M. & Katz, D. M. Measuring law over time: A network analytical framework with an application to statutes and regulations in the united states and germany. *Front. Phys.* **9**, 658463 (2021).

2. Chalkidis, I. *et al.* Lexglue: A benchmark dataset for legal language understanding in english. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, 4310–4330 (2022).

3. Friedrich, R. Complexity and entropy in legal language. *Front. Phys.* **9**, 671882 (2021).

4. Katz, D. M. & Bommarito, M. J. Measuring the complexity of the law: the united states code. *Artif. Intell. Law* **22**, 337–374 (2014).

5. Ruhl, J. B. Law's complexity: a primer. *Ga. St. UL Rev.* **24**, 885 (2007).

6. Mertz, E. *The language of law school: learning to" think like a lawyer"* (Oxford University Press, USA, 2007).

7. Bommarito, M. J. & Katz, D. M. GPT takes the bar exam. *Available at SSRN 4314839* (2022).

8. Bourcier, D. & Mazzega, P. Toward measures of complexity in legal systems. In *Proceedings of the 11th international conference on Artificial intelligence and law*, 211–215 (2007).

9. Ruhl, J., Katz, D. & Bommarito, M. Harnessing legal complexity. *Science* **355**, 1377–1378 (2017).

10. Bommarito II, M. & Katz, D. Measuring and modeling the us regulatory ecosystem. *J. Stat. Phys.* **168**, 1125–1135 (2017).

11. Katz, D. M., Coupette, C., Beckedorf, J. & Hartung, D. Complex societies and the growth of the law. *Sci. Reports* **10**, 1–14 (2020).

12. Staudt, R. W. All the wild possibilities: Technology that attacks barriers to access to justice. *Loy. LAL Rev.* **42**, 1117 (2008).

13. Sandefur, R. L. & Teufel, J. Assessing america's access to civil justice crisis. *UC Irvine L. Rev.* **11**, 753 (2020).

---

[4]As noted in the GPT-4 Technical Report "GPT-4 has various biases in its outputs that we have taken efforts to correct but which will take some time to fully characterize and manage."[38]

14. Rhode, D. L. *Access to justice* (Oxford University Press, 2004).

15. Susskind, R. E. *Online courts and the future of justice* (Oxford University Press, 2019).

16. Prescott, J. J. Improving access to justice in state courts with platform technology. *Vand. L. Rev.* **70**, 1993 (2017).

17. Hadfield, G. K. Higher demand, lower supply-a comparative assessment of the legal resource landscape for ordinary americans. *Fordham Urb. LJ* **37**, 129 (2010).

18. Caplan, L., Liebman, L. & Sandefur, R. Access to justice: Making justice accessible: Designing legal services for the 21st century. *Daedalus* **148** (2019).

19. Katz, D. M., Hartung, D., Gerlach, L., Jana, A. & Bommarito II, M. J. Natural language processing in the legal domain. *arXiv preprint arXiv:2302.12039* (2023).

20. Zhou, M., Duan, N., Liu, S. & Shum, H.-Y. Progress in neural nlp: modeling, learning, and reasoning. *Engineering* **6**, 275–290 (2020).

21. Otter, D. W., Medina, J. R. & Kalita, J. K. A survey of the usages of deep learning for natural language processing. *IEEE transactions on neural networks learning systems* **32**, 604–624 (2020).

22. McCulloch, W. S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin mathematical biophysics* **5**, 115–133 (1943).

23. Von Neumann, J. *The computer and the brain* (Yale University Press, 1958).

24. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).

25. Anderson, J. A. & Rosenfeld, E. *Talking nets: An oral history of neural networks* (MiT Press, 2000).

26. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 3111–3119 (2013).

27. Pennington, J., Socher, R. & Manning, C. D. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532–1543 (2014).

28. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

29. Vaswani, A. *et al.* Attention is all you need. In *Advances in neural information processing systems*, vol. 30 (2017).

30. Christiano, P. F. *et al.* Deep reinforcement learning from human preferences. *Adv. neural information processing systems* **30** (2017).

31. Peters, M. E. *et al.* Deep contextualized word representations. In *Proceedings of the NAACL HLT 2018-2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies-Proceedings of the Conference* (2018).

32. Devlin, J., Chang, M. W., Lee, K. & Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT 2019*, 4171–4186 (2018).

33. Kenton, J. D. M.-W. C. & Toutanova, L. K. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4171–4186 (2019).

34. Brown, T. *et al.* Language models are few-shot learners. In *Advances in neural information processing systems*, vol. 33, 1877–1901 (2020).

35. Zaheer, M. *et al.* Big bird: Transformers for longer sequences. *Adv. Neural Inf. Process. Syst.* **33**, 17283–17297 (2020).

36. Scao, T. L. *et al.* Bloom: A 176b-parameter open-access multilingual language model. *arXiv preprint arXiv:2211.05100* (2022).

37. Thoppilan, R. *et al.* Lamda: Language models for dialog applications. *arXiv preprint arXiv:2201.08239* (2022).

38. Open AI (2023), GPT-4 Technical Report. https://cdn.openai.com/papers/gpt-4.pdf (2023).

39. Ouyang, L. *et al.* Training language models to follow instructions with human feedback. *arXiv preprint arXiv:2203.02155* (2022).

40. Radford, A. *et al.* Language models are unsupervised multitask learners (2019).

41. Radford, A., Narasimhan, K., Salimans, T., Sutskever, I. *et al.* Improving language understanding by generative pre-training (OpenAI, 2018).

42. Zhong, H. *et al.* How does nlp benefit legal system: A summary of legal artificial intelligence. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5218–5230 (2020).

43. Dale, R. Law and word order: Nlp in legal tech. *Nat. Lang. Eng.* **25**, 211–217 (2019).

44. Xiao, C., Hu, X., Liu, Z., Tu, C. & Sun, M. Lawformer: A pre-trained language model for chinese legal long documents. *AI Open* **2**, 79–84 (2021).

45. Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N. & Androutsopoulos, I. LEGAL-BERT: The muppets straight out of law school. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2898–2904 (Association for Computational Linguistics, Online, 2020).

46. Qadrud-Din, J. *et al.* Transformer based language models for similar text retrieval and ranking. *arXiv preprint arXiv:2005.04588* (2020).

47. Zheng, L., Guha, N., Anderson, B. R., Henderson, P. & Ho, D. E. When does pretraining help? assessing self-supervised learning for law and the casehold dataset of 53,000+ legal holdings. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*, 159–168 (2021).

48. Chalkidis, I., Fergadiotis, M. & Androutsopoulos, I. Multieurlex-a multi-lingual and multi-label legal document classification dataset for zero-shot cross-lingual transfer. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 6974–6996 (2021).

49. Choi, J. H., Hickman, K. E., Monahan, A. & Schwarcz, D. Chatgpt goes to law school. *Available at SSRN* (2023).

50. Nay, J. J. Large language models as corporate lobbyists. *arXiv preprint arXiv:2301.01181* (2023).

51. Niklaus, J. *et al.* Lextreme: A multi-lingual and multi-task benchmark for the legal domain. *arXiv preprint arXiv:2301.13126* (2023).

52. Howarth, J. *Shaping the Bar: The Future of Attorney Licensing* (Stanford University Press, 2022).

53. Barton, B. H. An institutional analysis of lawyer regulation: Who should control lawyer regulation-courts, legislatures, or the market. *Ga. L. Rev.* **37**, 1167 (2002).

54. Griggs, M. Building a better bar exam. *Tex. A&M L. Rev.* **7**, 1 (2019).

55. NCBE releases first full-length simulated MBE study aid. https://www.ncbex.org/news/mbe-complete-practice-exam-release/ (2021).

56. Bommarito, J., Bommarito, M., Katz, D. M. & Katz, J. Gpt as knowledge worker: A zero-shot evaluation of (ai) cpa capabilities. *arXiv preprint arXiv:2301.04408* (2023).

57. Rajpurkar, P., Zhang, J., Lopyrev, K. & Liang, P. Squad: 100,000+ questions for machine comprehension of text. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2383–2392 (2016).

58. Wang, A. *et al.* Superglue: A stickier benchmark for general-purpose language understanding systems. *Adv. neural information processing systems* **32** (2019).

59. Wang, A. *et al.* Glue: A multi-task benchmark and analysis platform for natural language understanding. *arXiv preprint arXiv:1804.07461* (2018).

60. Kiela, D. *et al.* Dynabench: Rethinking benchmarking in nlp. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4110–4124 (2021).

61. Sejnowski, T. J. Large language models and the reverse turing test. *Neural Comput.* **35**, 309–342 (2023).

62. Mitchell, M. & Krakauer, D. C. The debate over understanding in AI's large language models. *arXiv preprint arXiv:2210.13966* (2022).

63. Wei, J. *et al.* Emergent abilities of large language models. *Transactions on Mach. Learn. Res.* (2022).

64. Dagan, I., Roth, D., Sammons, M. & Zanzotto, F. M. Textual entailment. In *Recognizing Textual Entailment: Models and Applications*, 1–24 (Springer, 2017).

65. Dagan, I., Dolan, B., Magnini, B. & Roth, D. Recognizing textual entailment: Rational, evaluation and approaches–erratum. *Nat. Lang. Eng.* **16**, 105–105 (2010).

66. Wang, R. & Neumann, G. Using recognizing textual entailment as a core engine for answer validation. In *Advances in Multilingual and Multimodal Information Retrieval: 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, September 19-21, 2007, Revised Selected Papers 8*, 387–390 (Springer, 2008).

67. Fawei, B., Wyner, A. & Pan, J. Passing a usa national bar exam: a first corpus for experimentation. In *LREC 2016, Tenth International Conference on Language Resources and Evaluation*, 23–30 (2016).

68. Wu, T. *et al.* Promptchainer: Chaining large language model prompts through visual programming. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 1–10 (2022).

69. Wang, Z. *et al.* Multitask prompt tuning enables parameter-efficient transfer learning. In *The Eleventh International Conference on Learning Representations* (2023).

70. White, J. *et al.* A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382* (2023).

71. Wei, J. *et al.* Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems* (2022).

Acknowledgements

Online Repository

Supplementary material, source code, and data is available online at
https://github.com/mjbommar/gpt4-passes-the-bar

Appendix I

Detailed Description of the Uniform Bar Exam (UBE)

The Multistate Bar Examination (MBE)

The Multistate Bar Examination (MBE) is a 200 question multiple choice test covering the following seven core legal subject matter areas – *Civil Procedure*, *Constitutional Law*, *Contracts*, *Criminal Law and Procedure*, *Evidence*, *Real Property*, and *Torts*. Each of the subjects is tested in equal weight. During the second day of the Uniform Bar Exam, test-takers complete the 200 questions in three hour, 100-question blocks split across morning and afternoon sessions. On average, a student should complete each MBE question in just over 1 minute and 45 seconds in order stay on pace to complete the test within the allotted time. Problems on the MBE are randomized, and the topic area of the question is not disclosed to the test-taker. Among the 200 questions are 25 experimental questions, which the NCBE uses for evaluation purposes. Thus, only 175 of the 200 questions actually count toward the test-taker's final score. Raw accuracy scores are calibrated and scaled by the MBE. However, a score of roughly 62% accuracy yields a 136-point scaled MBE score, which would put a test-taker on pace to pass the Bar Exam in almost every jurisdiction.

As discussed in earlier work, "questions on the MBE are designed to test both legal knowledge and reading comprehension skills, requiring above-average semantic and syntactic command of the English language. Instead of posing direct substantive legal questions, most MBE questions present the test-taker with a fictional situation. Descriptions of the facts are typically embellished with details; some of these details are critically important, while others are added only to distract or confuse the reader."[7]

The best way to understand the difficulty faced by the average test-taker is to review a real example. To that end, we provide a sample MBE question in Figure 3. In this question, the examinee is asked to consider a *Civil Procedure*-related vignette containing information about the defendant's motion to dismiss. The problem attempts to confuse the examinee as between the issues of personal service and personal jurisdiction. It also adds three separate jurisdictions - States A, B, and C - as well as the prospect that the defendant waived their challenge to personal jurisdiction by not including it in the original answer. As shown in Figure 4, the correct answer is < C > - "Yes, because the defendant lacks minimum contacts with State A."

The Multistate Essay Examination (MEE)

The Multistate Essay Examination (MEE) is a three hour component of the Uniform Bar Exam that is focused upon writing essays in response to long-form prompts. The NCBE describes the purpose of the MEE as determining "the examinee's ability to (1) identify legal issues raised by a hypothetical factual situation; (2) separate material which is relevant from that which is not; (3) present a reasoned analysis of the relevant issues in a clear, concise, and well-organized composition; and (4) demonstrate an understanding of the fundamental legal principles relevant to the probable solution of the issues raised by the factual situation. The primary distinction between the MEE and the Multistate Bar Examination (MBE) is that the MEE requires the examinee to demonstrate an ability to communicate effectively in writing."

While the MBE is limited to seven core legal subjects, the MEE draws from a much wider set of legal topics. Topics tested can include *Contracts*, *Constitutional Law*, *Criminal Law and Procedure*, *Evidence*, *Torts*, *Real Property*, *Business Associations*, *Conflict of Laws*, *Family Law*, *Federal Civil Procedure*, *Trusts and Estates*, and the *Uniform Commercial Code*. In any given administration of the exam, only six topics are actually posed to the test-taker. As documented by many bar preparation organizations, these topics are not uniformly likely to appear on the exam. Rather, the NCBE oversamples on certain subset of these topics.

A typical Bar Exam essay prompt offers a roughly one page vignette featuring a detailed set of information for the reader to consume. After reviewing the prompt, the test-taker is directed to answer several nested subquestions. As an illustration, consider Figure 5, a rendering of the July 2022 MEE Evidence Law question. The prompt poses a scenario where the prosecution is anticipated to introduce several items of evidence which, in the context of the attempted murder case, would collectively reflect very poorly upon the defendant. These include testimony from three individuals: an expert witness, a former leader of the street gang, and the victim. The defense raises a series of motions in response to this anticipated evidence. The examinee is then asked to provide analysis related to three separate objections raised by defense counsel. At the end of this scenario, instructions of limitation are also provided as the examinee is told not to consider any constitutional issues within their analysis.

In preparation for the Exam, it is generally suggested in the test preparation literature that a test-taker allot roughly 10-12 of the 30 total available minutes to the carefully review of the prompt and organization of a potential answer. The remaining

18-20 minutes should be devoted to writing. Therefore, for a prompt such as that shown in Figure 5, a test-taker following conventional wisdom would have roughly 6-7 minutes per sub-question to craft their response.

The Multistate Performance Test (MPT)

The Uniform Bar Exam begins on Day 1 with a 3 hour session devoted to the Multistate Performance Test (MPT). The MPT requires that test-takers complete a practical lawyering task such as legal analysis, fact analysis, problem solving, organization and management of information, and client communication in a limited amount of time. According to the NCBE, "[t]he MPT is designed to test an examinee's ability to use fundamental lawyering skills in a realistic situation and complete a task that a beginning lawyer should be able to accomplish." The three hours are broken down into two problems containing separate sets of materials, tasks, and questions, with a suggested time allocation of 90 minutes per problem.

Test-takers receive two sets of documents within each MPT question – (1) the "File" and (2) the "Library." The File consists of documents that the test-taker must review and scan for relevance. Depending upon the nature of the problem these documents could include interview transcripts, lawyer notes, reports, memos, police reports, etc. The material in the File serves as the factual basis for the completion of the task. Elements contained in the File include materials that are highly relevant, as well as those that are not particularly useful. The Library, by contrast, contains legal materials like statutes, regulations, cases, and advisory opinions. Similar to the File, elements contained in the Library may or may not turn out to be relevant in the given matter. The ability of the examinee to discriminate between this relevant and irrelevant information is a critical part of the MPT evaluation.

Using both the File and Library as the source of truth, the examinee is asked to produce a piece of attorney work product. The work product can vary, but the examinee is often asked to draft a document such as a memorandum, brief, opinion letter, demand letter, or even contract. Time is itself a challenge for this task, as it requires the detailed reading of the source materials, understanding the contours of the assigned task, and careful drafting of the final work product.

For the ease of the reader, in Figure 13, we reproduce the Instructional Memo from the July 2022 MPT-1. Links to access both the (1) the File and (2) the Library will be provided in the online repository and these questions and materials are also generally available through the website of several State Bars. For this question, the examinee is charged with the very realistic and common task of drafting a memo to a partner addressing four separate subquestions. The Instructional Memo provides some additional background facts, as well as details regarding the nature and scope of the expected work product. For example, the examinee is told to incorporate relevant facts into their memo and to not consider certain potentially-relevant issues (e.g., the potential prosecution for bigamy).

Appendix II

State	UBE Threshold
Alabama, Minnesota, Missouri, New Mexico, North Dakota	260
Indiana, Oklahoma	264
Connecticut, District of Columbia, Illinois, Iowa, Kansas, Kentucky Maryland, Montana, New Jersey, New York, South Carolina, Virgin Islands	266
Michigan	268
Alaska, Arkansas, Colorado, Maine, Massachusetts, Nebraska New Hampshire, North Carolina, Ohio, Oregon, Rhode Island, Tennessee, Texas, Utah, Vermont, Washington, West Virginia, Wyoming	270
Idaho, Pennsylvania	272
Arizona	273

Table 8. Uniform Bar Exam (UBE) Minimum Passing Score by State.

Appendix III

Sample MBE Question

A plaintiff domiciled in State A brought a federal diversity negligence action in State A against a defendant domiciled in State B. The action was based on an accident that had occurred in State C. The defendant was personally served with process at her office in State B, which is located 50 miles from the State A federal courthouse. The defendant travels to State A once each year for a week long vacation but has no other State A contacts.

The defendant answered, denying all allegations. One week later, the defendant filed an amended answer, denying all allegations and including the defense of lack of personal jurisdiction. State A has a long-arm statute that permits personal jurisdiction to the constitutional limit.

The defendant has moved for an order dismissing the action based on the personal-jurisdiction challenge asserted in the amended answer.

Should the court issue the order?

(A) No, because the defendant waived the challenge to personal jurisdiction by failing to include it in her original answer.

(B) No, because the defendant was personally served with process within 100 miles of the federal courthouse where the action is pending.

(C) Yes, because the defendant lacks minimum contacts with State A.

(D) Yes, because service was not delivered to the defendant at her home.

Figure 3. Sample MBE Question

GPT-4 - Sample MBE Question

Model Response:
[First Choice: C, Second Choice: A, Third Choice: B]

Correct Answer: "C" ✓

Figure 4. GPT-4 - Sample MBE Response

July 2022 MEE Evidence Question

Four months ago, Victim was shot and seriously wounded in City. Defendant has been charged with attempted murder. The prosecution's theory is that Victim and Defendant were both members of a criminal street gang called "The Lions," which engages in drug dealing, robbery, and murder in City. The prosecutor alleges that the shooting was the result of a gang dispute.

Defendant has brought a pretrial motion objecting to the prosecutor's introducing the following anticipated evidence:

(A) Testimony by a City detective who will be offered as an expert in gang identification, gang organizational structure, and gang activities generally and as an expert on particular gangs in City. The detective is expected to testify as follows:

I have been a detective on the police force for six years. Throughout that time, my primary assignment has been to investigate gangs and criminal activity in City. I have also worked closely with federal drug and firearm task forces as they relate to gangs. Prior to becoming a detective, I was a corrections officer in charge of the gang unit for City's jail for three years, and my duties included interviewing, investigating, and identifying gang members.

Throughout my career, I have attended training sessions providing education and information on gang structure, membership, and activities. As I've gained experience and knowledge in this area, I've frequently been asked to lead such sessions. I would estimate that I've taught more than 75 such training sessions over the past three years.

Street gangs generally engage in a wide variety of criminal activities. They usually have a clear leadership structure and strict codes of behavior. Absolute loyalty is required and is enforced through violent acts. Members of particular gangs can be identified by clothing, tattoos, language, paperwork, or associations.

I am quite familiar with "The Lions." It is one of City's most violent and feared criminal gangs. Members of The Lions can be identified by tattoos depicting symbols unique to the gang.

(B) Testimony by a former leader of The Lions concerning a photograph of Defendant's tattooed arm. After the photograph is authenticated as a photograph of Defendant's arm, the witness is expected to testify in part as follows:

I am certain that this is a Lions tattoo. I had a similar one removed. You'll notice that it has a shield containing the numbers for the police code for homicide, and Lions' members frequently include police codes in their tattoos to indicate crimes the gang has committed. The tattoo also has a shotgun and sword crossed as an "X," and a lion. Those are symbols frequently used by The Lions. This tattoo indicates to me, based on my experience, that Defendant is a member of The Lions gang.

(C) Testimony by Victim, who is expected to testify for the prosecution in part as follows:

I got into an argument with a gang boss at a meeting of The Lions. I said I wouldn't participate in an attack that was planned on another gang because my cousin was in that gang. The boss looked at Defendant and nodded to him. Next thing I knew, after the meeting, Defendant pulled a gun on me and shot me. I'm sure he did it because of that argument.

The jurisdiction has adopted rules of evidence identical to the Federal Rules of Evidence.

Defense counsel's motion raises the following objections to the evidence described above:

1. The detective's anticipated testimony about gang identification, organization, and activities is improper expert testimony.

2. The photograph of Defendant's tattoo and the former gang leader's anticipated testimony about it is inadmissible character evidence.

3. Victim's anticipated testimony that Defendant shot him because of a gang dispute is irrelevant.

How should the trial court rule on each objection? Explain. (Do not address constitutional issues.)

Figure 5. July 2022 MEE Evidence Question

GPT-4 - July 2022 MEE Evidence Answer (Part 1 of 3)

1. How should the trial court rule on the defendant's motion that the detective's anticipated testimony about gang identification, organization, and activities is improper expert testimony ? Explain. (Do not address constitutional issues.)

The trial court should deny the defendant's motion and allow the detective's anticipated testimony about gang identification, organization, and activities as proper expert testimony. Under Federal Rule of Evidence 702, a witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if: (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue; (b) the testimony is based on sufficient facts or data; (c) the testimony is the product of reliable principles and methods; and (d) the expert has reliably applied the principles and methods to the facts of the case.

In this case, the detective is qualified as an expert by his knowledge, experience, training, and education in the area of gang identification, organization, and activities. He has been a detective on the police force for six years, with a primary assignment of investigating gangs and criminal activity in City. He has also worked closely with federal drug and firearm task forces as they relate to gangs. Prior to becoming a detective, he was a corrections officer in charge of the gang unit for City's jail for three years, and his duties included interviewing, investigating, and identifying gang members. Additionally, he has attended training sessions providing education and information on gang structure, membership, and activities, and has frequently been asked to lead such sessions.

The detective's testimony will help the trier of fact to understand the evidence and to determine a fact in issue, namely, whether Defendant is a member of The Lions gang and whether the shooting was the result of a gang dispute. The detective's testimony is based on sufficient facts or data, as he is familiar with The Lions and has knowledge of their symbols, codes, and activities. The testimony is also the product of reliable principles and methods, as the detective has gained his knowledge and experience through his work as a detective and corrections officer, as well as through his attendance and leadership of training sessions. Finally, the detective has reliably applied the principles and methods to the facts of the case, as he is able to identify the symbols and codes used by The Lions and can explain their significance.

Therefore, the trial court should deny the defendant's motion and allow the detective's anticipated testimony about gang identification, organization, and activities as proper expert testimony.

Figure 6. GPT-4 Answer to July 2022 MEE Evidence Question (Part 1 of 3)

GPT-4 - July 2022 MEE Evidence Answer (Part 2 of 3)

2. How should the trial court rule on the defendant's motion that the photograph of Defendant's tattoo and the former gang leader's anticipated testimony about it is inadmissible character evidence? Explain. (Do not address constitutional issues.)

The trial court should deny the defendant's motion to exclude the photograph of Defendant's tattoo and the former gang leader's anticipated testimony about it as inadmissible character evidence. Under Federal Rule of Evidence 404(b), evidence of a crime, wrong, or other act is not admissible to prove a person's character in order to show that on a particular occasion the person acted in accordance with the character. However, such evidence may be admissible for another purpose, such as proving motive, opportunity, intent, preparation, plan, knowledge, identity, absence of mistake, or lack of accident.

In this case, the photograph of Defendant's tattoo and the former gang leader's testimony about it are not being offered to prove Defendant's character, but rather to prove his membership in The Lions gang, which is relevant to the prosecution's theory that the shooting was the result of a gang dispute. The evidence is also relevant to proving Defendant's motive and intent to commit the crime, as it supports the prosecution's theory that Defendant was acting on behalf of the gang and in accordance with its codes of behavior. Therefore, the evidence is admissible under Rule 404(b) for these purposes.

Furthermore, under Federal Rule of Evidence 403, the court may exclude relevant evidence if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence. In this case, the probative value of the evidence is high, as it is directly relevant to the prosecution's theory of the case and supports the testimony of the victim. The danger of unfair prejudice is also low, as the evidence is not being offered to prove Defendant's character, but rather his membership in the gang and his motive and intent to commit the crime. Therefore, the evidence should not be excluded under Rule 403.

In conclusion, the trial court should deny the defendant's motion to exclude the photograph of Defendant's tattoo and the former gang leader's anticipated testimony about it as inadmissible character evidence, as the evidence is admissible under Rules 404(b) and 403 for the purposes of proving Defendant's membership in the gang and his motive and intent to commit the crime.

Figure 7. GPT-4 Answer to July 2022 MEE Evidence Question (Part 2 of 3)

GPT-4 - July 2022 MEE Evidence Answer (Part 3 of 3)

3. How should the trial court rule on the defendant's motion that the Victim's anticipated testimony that Defendant shot him because of a gang dispute is irrelevant? Explain. (Do not address constitutional issues.)

The trial court should deny the defendant's motion that Victim's anticipated testimony that Defendant shot him because of a gang dispute is irrelevant. Under the Federal Rules of Evidence, relevant evidence is defined as "evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence." Fed. R. Evid. 401. In this case, the fact that Defendant shot Victim because of a gang dispute is relevant to the prosecution's theory that the shooting was the result of a gang dispute. The testimony by Victim directly supports the prosecution's theory and makes it more probable that the shooting was the result of a gang dispute. Therefore, the testimony is relevant and should be admitted.

Furthermore, the Federal Rules of Evidence provide that "all relevant evidence is admissible, except as otherwise provided by the Constitution of the United States, by Act of Congress, by these rules, or by other rules prescribed by the Supreme Court pursuant to statutory authority." Fed. R. Evid. 402. There is no indication that the Victim's testimony is inadmissible under any of these exceptions. Therefore, the trial court should deny the defendant's motion and allow the Victim's testimony to be admitted.

It is important to note that the trial court has the discretion to exclude relevant evidence "if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence." Fed. R. Evid. 403. However, in this case, there is no indication that the Victim's testimony would be unfairly prejudicial, confusing, misleading, or a waste of time. Therefore, the trial court should not exclude the Victim's testimony on these grounds.

In conclusion, the trial court should deny the defendant's motion that Victim's anticipated testimony that Defendant shot him because of a gang dispute is irrelevant. The testimony is relevant and admissible under the Federal Rules of Evidence, and there is no indication that it should be excluded under Rule 403.

Figure 8. GPT-4 Answer to July 2022 MEE Evidence Question (Part 3 of 3)

ChatGPT - July 2022 MEE Evidence Answer (Part 1 of 3)

1. How should the trial court rule on the defendant's motion that the detective's anticipated testimony about gang identification, organization, and activities is improper expert testimony ? Explain. (Do not address constitutional issues.)

The trial court should deny the defendant's motion to exclude the detective's anticipated testimony about gang identification, organization, and activities as improper expert testimony.

Under Federal Rule of Evidence 702, expert testimony is admissible if it is (1) based on sufficient facts or data, (2) the product of reliable principles and methods, and (3) the expert has applied the principles and methods reliably to the facts of the case. The trial court has broad discretion in determining the admissibility of expert testimony, and its decision will only be overturned if it is an abuse of discretion.

In this case, the detective's anticipated testimony satisfies all three prongs of the Rule 702 test. First, the detective's testimony is based on his extensive experience investigating gangs and criminal activity in City, as well as his training and education on gang structure, membership, and activities. He has also interviewed, investigated, and identified gang members in his prior role as a corrections officer. Therefore, the detective has sufficient facts and data to support his testimony.

Second, the detective's testimony is based on reliable principles and methods. He has attended numerous training sessions on gang structure, membership, and activities, and has frequently been asked to lead such sessions. His experience and knowledge in this area make him qualified to testify as an expert on gang identification, organization, and activities. Additionally, his testimony is consistent with the general understanding of gangs and their activities.

Finally, the detective has reliably applied his principles and methods to the facts of the case. He has identified Defendant as a member of The Lions based on his tattoo, which contains symbols unique to the gang. He has also provided general information about gang behavior and activities that is relevant to the prosecution's theory of the case. Therefore, the detective's anticipated testimony is admissible under Rule 702.

The defendant may argue that the detective's testimony is prejudicial, and that its probative value is outweighed by the risk of unfair prejudice. Under Federal Rule of Evidence 403, relevant evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice. However, the detective's testimony is highly probative of Defendant's alleged gang membership and the prosecution's theory of the case. Any potential prejudice can be mitigated by the trial court through limiting instructions to the jury.

In conclusion, the trial court should deny the defendant's motion to exclude the detective's anticipated testimony about gang identification, organization, and activities as improper expert testimony. The detective's testimony satisfies the requirements of Rule 702 and is highly probative of Defendant's alleged gang membership and the prosecution's theory of the case.

Figure 9. ChatGPT Answer to July 2022 MEE Evidence Question (Part 1 of 3)

ChatGPT - July 2022 MEE Evidence Answer (Part 2 of 3)

2. How should the trial court rule on the defendant's motion that The photograph of Defendant's tattoo and the former gang leader's anticipated testimony about it is inadmissible character evidence? Explain. (Do not address constitutional issues.)

The trial court should deny the defendant's motion to exclude the photograph of Defendant's tattoo and the former gang leader's anticipated testimony about it as inadmissible character evidence.

Under Rule 404(a) of the Federal Rules of Evidence, evidence of a person's character or character trait is not admissible to prove that on a particular occasion the person acted in accordance with the character or trait. However, Rule 404(b) provides that evidence of other crimes, wrongs, or acts may be admissible for other purposes, such as proving motive, opportunity, intent, preparation, plan, knowledge, identity, absence of mistake, or lack of accident.

Here, the prosecution is not seeking to introduce evidence of Defendant's character or character trait to prove that he acted in accordance with that trait in shooting Victim. Rather, the prosecution seeks to introduce evidence of Defendant's gang membership and specifically, his membership in The Lions, to prove his motive and intent in shooting Victim. The prosecution's theory is that the shooting was the result of a gang dispute, and evidence of Defendant's gang membership is relevant to proving that theory.

The anticipated testimony of the former gang leader about Defendant's tattoo is also admissible under Rule 701 of the Federal Rules of Evidence. Rule 701 allows a lay witness to testify in the form of an opinion if the witness's opinion is based on the witness's perception and is helpful to a clear understanding of the witness's testimony or to the determination of a fact in issue. Here, the former gang leader's opinion about Defendant's tattoo is based on his personal knowledge and experience as a former leader of The Lions, and his testimony is helpful to the jury in understanding the significance of Defendant's tattoo and its connection to The Lions.

Furthermore, the anticipated testimony of the City detective about gang identification, gang organizational structure, and gang activities generally is admissible under Rule 702 of the Federal Rules of Evidence. Rule 702 allows expert testimony if the witness is qualified as an expert by knowledge, skill, experience, training, or education, and if the testimony will assist the trier of fact to understand the evidence or to determine a fact in issue. Here, the City detective has extensive experience and training in investigating gangs and their activities, and his testimony will assist the jury in understanding the nature of gangs, their organizational structure, and their activities, as well as the specific characteristics of The Lions gang.

In sum, the trial court should deny the defendant's motion to exclude the photograph of Defendant's tattoo and the former gang leader's anticipated testimony about it, as well as the anticipated testimony of the City detective about gangs and gang activities generally. These pieces of evidence are relevant to proving the prosecution's theory of the case and are admissible under the Federal Rules of Evidence.

Figure 10. ChatGPT Answer to July 2022 MEE Evidence Question (Part 2 of 3)

ChatGPT - July 2022 MEE Evidence Answer (Part 3 of 3)

3. How should the trial court rule on the defendant's motion that the Victim's anticipated testimony that Defendant shot him because of a gang dispute is irrelevant? Explain. (Do not address constitutional issues.)

The trial court should deny the defendant's motion to exclude Victim's anticipated testimony that Defendant shot him because of a gang dispute as irrelevant.

Under Rule 401 of the Federal Rules of Evidence, evidence is relevant if it has any tendency to make a fact of consequence more or less probable than it would be without the evidence. Here, Victim's testimony that Defendant shot him because of a gang dispute is relevant because it tends to make it more probable that Defendant committed the attempted murder with which he is charged. If the jury believes Victim's testimony, it would tend to show that Defendant had a motive to shoot Victim and that the shooting was not a random act of violence.

Defendant may argue that Victim's testimony is not relevant because it does not prove that Defendant is a member of The Lions or that the shooting was related to gang activity. However, under Rule 402 of the Federal Rules of Evidence, relevant evidence is admissible unless excluded by another rule. Here, Defendant has not identified any other rule that would exclude Victim's testimony as irrelevant.

Defendant may also argue that Victim's testimony is more prejudicial than probative under Rule 403 of the Federal Rules of Evidence. However, the probative value of Victim's testimony outweighs any potential prejudice. The fact that Defendant is a member of The Lions and that the shooting was related to a gang dispute is highly probative of Defendant's guilt, and any potential prejudice can be mitigated by a limiting instruction from the court.

In addition, Defendant may argue that Victim's testimony is hearsay under Rule 801 of the Federal Rules of Evidence. However, Victim's testimony is not hearsay because it is not offered to prove the truth of the matter asserted (i.e., that there was a gang dispute and that Defendant shot Victim because of it). Rather, it is offered to show Defendant's motive and intent.

Therefore, the trial court should deny Defendant's motion to exclude Victim's anticipated testimony as irrelevant.

Figure 11. ChatGPT Answer to July 2022 MEE Evidence Question (Part 3 of 3)

GPT-3.0 - July 2022 MEE Evidence Answer)

1. How should the trial court rule on the defendant's motion that the detective's anticipated testimony about gang identification, organization, and activities is improper expert testimony ? Explain. (Do not address constitutional issues.)

The trial court should rule on the defendant's motion that the detective's anticipated testimony about gang identification, organization, and activities is improper expert testimony. The trial court should find that the detective's testimony is inadmissible because it is not based on sufficient facts or expertise. The trial court should also find that the testimony is unfairly prejudicial because it will likely result in the jury considering irrelevant evidence.

2. How should the trial court rule on the defendant's motion that the photograph of Defendant's tattoo and the former gang leader's anticipated testimony about it is inadmissible character evidence? Explain. (Do not address constitutional issues.)

The trial court should rule on the defendant's motion that the photograph of Defendant's tattoo and the former gang leader's anticipated testimony about it is inadmissible character evidence. The trial court should hold that the photograph is not relevant to the issues at trial and that the testimony is inadmissible character evidence.

The trial court should hold that the photograph is not relevant to the issues at trial because it is not evidence that Defendant committed the crime. The trial court should also hold that the testimony is inadmissible character evidence because it is not evidence that Defendant has a disposition to commit the crime. The trial court should note that the character evidence is not relevant to any issue at trial, including whether Defendant is guilty of the attempted murder charge.

The trial court should also hold that the character evidence is inadmissible because it is more prejudicial than probative. The trial court should note that the character evidence is not relevant to any issue at trial, including whether Defendant is guilty of the attempted murder charge. The trial court should find that the probative value of the character evidence is substantially outweighed by the danger of unfair prejudice, and the trial court should exclude the evidence.

3. How should the trial court rule on the defendant's motion that the Victim's anticipated testimony that Defendant shot him because of a gang dispute is irrelevant? Explain. (Do not address constitutional issues.)

The trial court should rule on the defendant's motion that Victim's anticipated testimony that Defendant shot him because of a gang dispute is irrelevant by determining whether the evidence is relevant to the crime with which Defendant has been charged. Under Rule 401 of the Federal Rules of Evidence, evidence is relevant if it has any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence. Here, the prosecutor intends to introduce evidence that Defendant shot Victim because of a gang dispute. This evidence is relevant to the crime with which Defendant has been charged. Therefore, the trial court should not rule on the defendant's motion that Victim's anticipated testimony is irrelevant.

Figure 12. GPT-3.0 (text-davinci-001) Answer to July 2022 MEE Evidence Question

July 2022 MPT-1 Instructional Memo

To: Examinee
From: Marianne Morton
Date: July 26, 2022
Re: Walter Hixon matter

We represent Walter Hixon in connection with complications of his marital status. Mr. Hixon married Joan Prescott in 1986 in the State of Columbia. Several years later they separated. Mr. Hixon believed that Ms. Prescott died in 2001.

In 2012, he married Frances Tucker in the State of Columbia. They purchased a house together in Columbia early in the marriage. A few years ago, Mr. Hixon moved to Franklin for a job opportunity; Ms. Tucker remained in Columbia.

Last month, Mr. Hixon learned that Joan Prescott is still alive. He has informed Ms. Tucker of that fact. He wants to divorce Ms. Prescott, end his purported marriage with Ms. Tucker, and work out shares in the residential property that he and Ms. Tucker own.

I need you to write a memorandum to me addressing the following questions:

1. Does Columbia or Franklin law govern the grounds for annulling Mr. Hixon's marriage to Ms. Tucker?

2. Must Mr. Hixon file a lawsuit to annul his second marriage, and if yes, would he be able to obtain an annulment under the applicable law?

3. If Mr. Hixon files an annulment action in Franklin, would a Franklin court have jurisdiction to annul the marriage and to dispose of the parties' property?

4. Should we advise Mr. Hixon to file in Columbia or in Franklin?

Do not prepare a separate statement of facts, but be sure to incorporate the relevant facts into your analysis and state the reasons for your conclusions and recommendation. Do not address either Mr. Hixon's ending his marriage to Ms. Prescott or the risks of criminal prosecution he may face for bigamy; another associate will research those issues.

Figure 13. July 2022 MPT-1 Instructional Memo

GPT-4 - July 2022 MPT-1 Answer - (Part 1 of 4)

MEMORANDUM
To: Marianne Morton
From: Examinee
Date: July 26, 2022
Re: Walter Hixon matter

1. Does Columbia or Franklin law govern the grounds for annulling Mr. Hixon's marriage to Ms. Tucker?

Analysis:
The first step in determining which state's law governs the grounds for annulling Mr. Hixon's marriage to Ms. Tucker is to determine which state has the most significant relationship to the spouses and the marriage. Under the Restatement (Second) of Conflict of Laws § 283, the validity of a marriage will be determined by the local law of the state which, with respect to the particular issue, has the most significant relationship to the spouses and the marriage under the principles stated in § 6.

In this case, both Mr. Hixon and Ms. Tucker were residents of Columbia at the time of their marriage, and they purchased a house together in Columbia early in the marriage. These facts suggest that Columbia has a significant relationship to the spouses and the marriage. However, Mr. Hixon moved to Franklin a few years ago for a job opportunity, while Ms. Tucker remained in Columbia. This fact suggests that Franklin may also have a significant relationship to the spouses and the marriage.

To determine which state has the most significant relationship, we must consider the factors stated in Restatement § 6. These factors include the relevant policies of other interested states and the relative interests of those states in the determination of the particular issue, the protection of justified expectations, certainty, predictability, and uniformity of result, and ease in the determination and application of the law to be applied.

With respect to the relevant policies of other interested states and the relative interests of those states in the determination of the particular issue, both Columbia and Franklin have legitimate policy interests in defining how a relationship as fundamental as marriage can be initiated and ended. However, the fact that Columbia and Franklin recognize different reasons for annulling a marriage indicates the strength of the policy interests involved. In this case, Columbia law permits annulment on the ground of fraudulent inducement, while Franklin law does not. This difference in the law suggests that Columbia has a stronger policy interest in the determination of the particular issue.

With respect to the protection of justified expectations, Mr. Hixon and Ms. Tucker married in Columbia, lived the majority of their married life there, and owned property together there. These facts strongly suggest that the parties had a justified expectation that Columbia law would govern the terms on which the marriage ended.

With respect to certainty, predictability, and uniformity of result, people often move between states, creating the need for a system of well-defined rules to govern which state's laws apply to the creation and termination of marriages. In this case, the fact that Mr. Hixon and Ms. Tucker were residents of Columbia at the time of their marriage and purchased a house together in Columbia early in the marriage suggests that Columbia law should apply to the creation and termination of their marriage.

With respect to ease in the determination and application of the law to be applied, all the important events in this marriage occurred in Columbia. Considerations of ease and administrative efficiency strongly suggest Columbia as the appropriate forum.

Conclusion:
Based on the factors stated in Restatement § 6, it appears that Columbia law should govern the grounds for annulling Mr. Hixon's marriage to Ms. Tucker. Columbia has the most significant relationship to the spouses and the marriage, and the relevant policies of Columbia, the protection of justified expectations, certainty, predictability, and uniformity of result, and ease in the determination and application of the law to be applied all support the application of Columbia law. Therefore, I recommend that we advise Mr. Hixon that Columbia law will likely govern the grounds for annulling his marriage to Ms. Tucker.

Figure 14. GPT-4 Answer to July 2022 MPT-1 Problem (Part 1 of 4)

GPT-4 - July 2022 MPT-1 Answer - (Part 2 of 4)

MEMORANDUM
To: Marianne Morton
From: Examinee
Date: July 26, 2022
Re: Walter Hixon matter

2. Must Mr. Hixon file a lawsuit to annul his second marriage, and if yes, would he be able to obtain an annulment under the applicable law?

Analysis

A. Must Mr. Hixon file a lawsuit to annul his second marriage?

Under Franklin law, a marriage that is void does not require a judicial decree to end it. Franklin Domestic Relations Code § 19-5(a) states that certain marriages, including those where either party is lawfully married to another person, are void without the need for any decree of divorce, annulment, or other legal proceeding.

However, under Columbia law, a marriage that is voidable requires a judicial decree to end it. Columbia Revised Statutes § 718.02 states that a marriage is voidable if the spouse of either party was living and the marriage with that spouse was then in force, and that spouse was absent and not known to the party commencing the proceeding to be living for a period of five successive years immediately preceding the subsequent marriage for which the annulment decree is sought. The statute further states that for a voidable marriage to be declared void, either party may seek and a court must issue an annulment decree.

Because Mr. Hixon's second marriage took place in Columbia and both he and Ms. Tucker resided in Columbia for the majority of their marriage, it is likely that Columbia law would apply to the question of whether Mr. Hixon must file a lawsuit to annul his second marriage. Therefore, Mr. Hixon would need to file a lawsuit to obtain an annulment of his second marriage under Columbia law.

B. Would Mr. Hixon be able to obtain an annulment under the applicable law?

Under Columbia law, Mr. Hixon would need to prove that his first wife, Ms. Prescott, was living and the marriage with her was then in force, and that she was absent and not known to him to be living for a period of five successive years immediately preceding his subsequent marriage to Ms. Tucker.

Mr. Hixon has stated that he believed Ms. Prescott had died in 2001, eleven years before his marriage to Ms. Tucker. If he can provide evidence to support this belief, such as the testimony of the mutual friend who told him of Ms. Prescott's supposed death, he may be able to satisfy the requirement that she was absent and not known to him to be living for a period of five successive years immediately preceding his subsequent marriage.

However, Mr. Hixon would also need to prove that his marriage with Ms. Prescott was still in force at the time of his subsequent marriage to Ms. Tucker. This may be more difficult, as Mr. Hixon has stated that he and Ms. Prescott separated in 1990 and had no contact after that. It is possible that Ms. Prescott obtained a divorce or annulment without Mr. Hixon's knowledge, which would mean that their marriage was not in force at the time of his subsequent marriage to Ms. Tucker. Mr. Hixon would need to provide evidence, such as a search of divorce or annulment records, to prove that his marriage to Ms. Prescott was still in force at the time of his subsequent marriage to Ms. Tucker.

Conclusion

Based on the applicable law, Mr. Hixon would need to file a lawsuit to obtain an annulment of his second marriage under Columbia law. However, whether he would be able to obtain an annulment would depend on his ability to prove that his first wife was living and the marriage with her was then in force, and that she was absent and not known to him to be living for a period of five successive years immediately preceding his subsequent marriage to Ms. Tucker.

Figure 15. GPT-4 Answer to July 2022 MPT-1 Problem (Part 2 of 4)

GPT-4 - July 2022 MPT-1 Answer - (Part 3 of 4)

MEMORANDUM
To: Marianne Morton
From: Examinee
Date: July 26, 2022
Re: Walter Hixon matter

3. If Mr. Hixon files an annulment action in Franklin, would a Franklin court have jurisdiction to annul the marriage and to dispose of the parties' property?

I. Introduction

This memorandum addresses the question of whether a Franklin court would have jurisdiction to annul the marriage between Mr. Hixon and Ms. Tucker and to dispose of the parties' property. As explained below, a Franklin court would have jurisdiction to annul the marriage, but it may not have jurisdiction to dispose of the parties' property located in Columbia.

II. Jurisdiction to Annul the Marriage

A Franklin court would have jurisdiction to annul the marriage between Mr. Hixon and Ms. Tucker because Mr. Hixon is a resident of Franklin. Under Franklin law, a court has jurisdiction over the res of the marriage relationship when one of the parties to the marriage has been domiciled within the state for the requisite period, which in Franklin is six months. See Daniels v. Daniels, Franklin Court of Appeal (1997). Mr. Hixon has been a resident of Franklin for several years, so he meets this requirement.

Furthermore, Franklin law does not require in personam jurisdiction over both parties to the marriage in order to grant an annulment. See Daniels, supra; Price v. Price, Franklin Supreme Court (1972); Carew v. Ellis, Franklin Supreme Court (1957). Therefore, the fact that Ms. Tucker is a resident of Columbia does not prevent a Franklin court from having jurisdiction to annul the marriage.

III. Jurisdiction to Dispose of the Parties' Property

While a Franklin court would have jurisdiction to annul the marriage, it may not have jurisdiction to dispose of the parties' property located in Columbia. Franklin courts have long held that, even in the absence of in personam jurisdiction over the defendant in a case seeking to end a marriage, a Franklin court can render a valid judgment with respect to real property located in Franklin. See Gore v. Gore, Franklin Supreme Court (1985) (divorce); Carew v. Ellis, supra (annulment). However, these cases do not address the issue of whether a Franklin court can dispose of property located outside of Franklin.

The Restatement (Second) of Conflict of Laws § 283 (1971) provides that the validity of a marriage will be determined by the local law of the state which, with respect to the particular issue, has the most significant relationship to the spouses and the marriage under the principles stated in § 6. One of the factors listed in § 6 is "certainty, predictability, and uniformity of result." Applying this factor, it would be more certain, predictable, and uniform for a Columbia court to dispose of property located in Columbia, rather than a Franklin court.

Furthermore, the United States Supreme Court has held that the mere presence of property in a state, standing alone, will not constitute sufficient "minimum contacts" to support the state's exercise of its in rem jurisdiction, if the property is unrelated to the underlying cause of action. See Shaffer v. Heitner, 433 U.S. 186 (1977). In this case, the property is located in Columbia and is unrelated to the underlying cause of action, which is the annulment of the marriage. Therefore, a Franklin court may not have jurisdiction to dispose of the parties' property located in Columbia.

IV. Conclusion

In conclusion, a Franklin court would have jurisdiction to annul the marriage between Mr. Hixon and Ms. Tucker, but it may not have jurisdiction to dispose of the parties' property located in Columbia. Mr. Hixon may need to file a separate action in Columbia to address the issue of the parties' property.

Figure 16. GPT-4 Answer to July 2022 MPT-1 Problem (Part 3 of 4)

GPT-4 - July 2022 MPT-1 Answer - (Part 4 of 4)

MEMORANDUM
To: Marianne Morton
From: Examinee
Date: July 26, 2022
Re: Walter Hixon matter

4. Should we advise Mr. Hixon to file in Columbia or in Franklin?

I have analyzed the question of whether we should advise Mr. Hixon to file in Columbia or in Franklin. Based on my analysis, I recommend that Mr. Hixon file in Columbia. I will explain my reasoning below.

1. Filing in Columbia

There are several reasons why filing in Columbia would be the better option for Mr. Hixon. First, Columbia has the most significant relationship to the parties and the marriage. Both of Mr. Hixon's marriages were contracted in Columbia, and he and Ms. Tucker lived together and purchased property in Columbia. Additionally, Ms. Prescott has family in Columbia, and Ms. Tucker currently resides in Columbia. These factors suggest that Columbia has the most significant relationship to the parties and the marriage, and therefore, Columbia law should apply.

Second, filing in Columbia would provide certainty, predictability, and uniformity of result. If Mr. Hixon files in Columbia, there will be no question as to which state's law applies, and the parties will have a clear understanding of their rights and obligations. This will also promote uniformity of result, as the same law will be applied to both of Mr. Hixon's marriages.

Third, filing in Columbia would be easier for the parties and the court. Since both of Mr. Hixon's marriages were contracted in Columbia, and he and Ms. Tucker purchased property in Columbia, it would be easier for the parties and the court to gather and present evidence in Columbia. Additionally, since Ms. Tucker currently resides in Columbia, it would be more convenient for her to participate in the proceedings in Columbia.

2. Filing in Franklin

While there are some potential benefits to filing in Franklin, I believe that they are outweighed by the benefits of filing in Columbia. One potential benefit of filing in Franklin is that Mr. Hixon currently resides in Franklin, and it may be more convenient for him to participate in the proceedings in Franklin. However, as discussed above, Columbia has the most significant relationship to the parties and the marriage, and it would be easier for the parties and the court to gather and present evidence in Columbia.

Another potential benefit of filing in Franklin is that Franklin law may be more favorable to Mr. Hixon. However, it is important to note that the validity of a marriage is generally determined by the law of the state with the most significant relationship to the spouses and the marriage. In this case, that state is Columbia. Therefore, even if Mr. Hixon files in Franklin, it is likely that Columbia law will be applied.

Conclusion

Based on the above analysis, I recommend that Mr. Hixon file in Columbia. Columbia has the most significant relationship to the parties and the marriage, and filing in Columbia would provide certainty, predictability, and uniformity of result. Additionally, it would be easier for the parties and the court to gather and present evidence in Columbia. While there are some potential benefits to filing in Franklin, I believe that they are outweighed by the benefits of filing in Columbia.

Figure 17. GPT-4 Answer to July 2022 MPT-1 Problem (Part 4 of 4)

ChatGPT - July 2022 MPT-1 Answer - (Part 1 of 4)

MEMORANDUM
To: Marianne Morton
From: Examinee
Date: July 26, 2022
Re: Walter Hixon matter

1. Does Columbia or Franklin law govern the grounds for annulling Mr. Hixon's marriage to Ms. Tucker?

I. Brief Answer
Franklin law governs the grounds for annulling Walter Hixon's marriage to Frances Tucker.

II. Discussion
Under the Restatement (Second) of Conflict of Laws § 6, a court should apply the law of the state with the most significant relationship to the spouses and the marriage under the principles stated in § 6. If a state has no such relationship, that state must apply the law of the state that does. The factors relevant to the choice of the applicable rule of law include the relevant policies of other interested states and the relative interest
s of those states in the determination of the particular issue, the protection of justified expectations, certainty, predictability, and uniformity of result, and ease in the determination and application of the law to be applied.

In this case, the parties were married in Columbia, but Mr. Hixon currently resides in Franklin, and the couple purchased a house in Columbia. However, the fact that Mr. Hixon currently resides in Franklin is not determinative of the choice of law issue. Instead, the relevant factors to consider include the policies of the interested states, the relative interests of those states, the protection of justified expectations, and ease of determination and application of the law.

The policies of the interested states are relevant, and the fact that the grounds for annulment differ between Columbia and Franklin is indicative of the strength of the policy interests involved. Columbia law permits an annulment on the ground that the spouse of either party was living and the marriage with that spouse was then in force and that spouse was absent and not known to the party commencing the proceeding to be living for a period of five successive years immediately preceding the subsequent marriage for which the annulment decree is sought. By contrast, Franklin law provides that certain marriages shall be void, including all marriages between parties where either party is lawfully married to another person. While both states have a policy interest in regulating the institution of marriage, the fact that the grounds for annulment differ between the two states suggests that the policy interests of each state are strong.

The relative interests of the states are also relevant. In this case, the parties were married in Columbia, but they purchased a house in Columbia and lived there together. Mr. Hixon moved to Franklin for a job opportunity, but Ms. Tucker remained in Columbia. The fact that the couple purchased a house in Columbia and lived there together suggests that Columbia has a significant relationship to the marriage.

However, the protection of justified expectations also weighs in favor of applying Franklin law. Mr. Hixon currently resides in Franklin, and the couple purchased a house in Columbia together. While the fact that the couple purchased a house in Columbia suggests that they had a justified expectation that Columbia law would apply, the fact that Mr. Hixon currently resides in Franklin suggests that he had a justified expectation that Franklin law would apply.

Finally, ease of determination and application of the law also weighs in favor of applying Franklin law. The parties' marriage was entered into in Columbia, but the parties currently reside in different states, and the property at issue is located in Columbia. Applying Franklin law would provide a clear and predictable rule for determining the grounds for annulment.

III. Conclusion
Given the relevant factors, Franklin law governs the grounds for annulling Walter Hixon's marriage to Frances Tucker.

Figure 18. ChatGPT Answer to July 2022 MPT-1 Problem (Part 1 of 4)

ChatGPT - July 2022 MPT-1 Answer - (Part 2 of 4)

MEMORANDUM
To: Marianne Morton
From: Examinee
Date: July 26, 2022
Re: Walter Hixon matter

2. Must Mr. Hixon file a lawsuit to annul his second marriage, and if yes, would he be able to obtain an annulment under the applicable law?

I. Short Answer
Yes, Mr. Hixon must file a lawsuit to annul his second marriage in the State of Columbia, and he may be able to obtain an annulment under the applicable law based on the fact that he was still legally married to Ms. Prescott at the time of his marriage to Ms. Tucker.

II. Discussion
Under Columbia Revised Statutes § 718.02, a marriage is voidable if the spouse of either party was living and the marriage with that spouse was then in force and that spouse was absent and not known to the party commencing the proceeding to be living for a period of five successive years immediately preceding the subsequent marriage for which the annulment decree is sought. Here, Mr. Hixon was still legally married to Ms. Prescott at the time of his marriage to Ms. Tucker, and he did not know that Ms. Prescott was still alive. Therefore, he may be able to obtain an annulment of his second marriage to Ms. Tucker based on the fact that it was voidable under Columbia law.

However, Mr. Hixon must file a lawsuit to obtain an annulment of his second marriage in the State of Columbia. An annulment is a judicial decree declaring that a marriage is void, and it must be obtained through a legal proceeding. Mr. Hixon cannot simply declare his second marriage void on his own.

In addition, it is important to note that the Franklin court would likely recognize the Columbia annulment decree under the principles of full faith and credit. The Restatement (Second) of Conflict of Laws § 283 provides that a marriage that satisfies the requirements of the state where the marriage was contracted will be recognized as valid everywhere unless it violates the strong public policy of another state which had the most significant relationship to the spouses and the marriage at the time of the marriage. Here, the State of Columbia had the most significant relationship to Mr. Hixon's second marriage, as it was where the marriage was contracted and where the parties resided at the time of the marriage. Therefore, if Mr. Hixon is able to obtain an annulment of his second marriage in the State of Columbia, the Franklin court would likely recognize the annulment decree.

III. Conclusion
Mr. Hixon must file a lawsuit to obtain an annulment of his second marriage in the State of Columbia, and he may be able to obtain an annulment based on the fact that he was still legally married to Ms. Prescott at the time of his marriage to Ms. Tucker. If he is able to obtain an annulment in Columbia, the Franklin court would likely recognize the annulment decree.

Figure 19. ChatGPT Answer to July 2022 MPT-1 Problem (Part 2 of 4)

ChatGPT - July 2022 MPT-1 Answer - (Part 3 of 4)

MEMORANDUM
To: Marianne Morton
From: Examinee
Date: July 26, 2022
Re: Walter Hixon matter

3. If Mr. Hixon files an annulment action in Franklin, would a Franklin court have jurisdiction to annul the marriage and to dispose of the parties' property?

I. Short Answer
Yes, a Franklin court would have jurisdiction to annul Walter Hixon's marriage to Frances Tucker and dispose of the parties' property if he files an annulment action in Franklin.

II. Analysis

A. Jurisdiction
The first issue to consider is whether a Franklin court has jurisdiction over Walter Hixon's marriage to Frances Tucker and the parties' property. Jurisdiction is a threshold issue that must be resolved before a court can proceed to the merits of the case.
In the present case, Walter Hixon and Frances Tucker were married in the State of Columbia. However, Mr. Hixon has been residing in Franklin for several years, and his wife has remained in Columbia. Under Franklin law, a court has jurisdiction to annul a marriage and to dispose of the parties' property if one of the parties has been domiciled in Franklin for at least six months. Daniels v. Daniels (Fr. Ct. App. 1997). Mr. Hixon has been residing in Franklin for several years and therefore meets the residency requirement. Thus, a Franklin court would have jurisdiction over the parties' marriage and property.

B. Annulment
Next, we must consider whether Mr. Hixon can seek an annulment of his marriage to Frances Tucker. An annulment is a legal proceeding that declares a marriage to be void or invalid from its inception. Fletcher v. Fletcher (Fr. Ct. App. 2014). In Franklin, a marriage may be declared void if it falls under one of the categories listed in the Domestic Relations Code, such as a bigamous marriage or a marriage between close relatives. Franklin Domestic Relations Code § 19-5(a). A marriage may also be declared voidable if it falls under the categories listed in Columbia Revised Statutes § 718.02, which includes a marriage where one spouse was already married and the previous spouse was absent for five years. Columbia Revised Statutes § 718.02(A)(1).

In this case, Mr. Hixon's marriage to Frances Tucker does not fall under any of the categories listed in Franklin Domestic Relations Code § 19-5(a). However, it may be declared voidable under Columbia Revised Statutes § 718.02(A)(1) if Mr. Hixon can establish that his first marriage to Joan Prescott was still valid at the time of his marriage to Ms. Tucker. If he can establish that, he may seek an annulment of his marriage to Ms. Tucker on the ground that it was voidable under Columbia law.

C. Disposition of Property
Finally, we must consider whether a Franklin court can dispose of the parties' property in an annulment action. Under Franklin law, the provisions relating to property rights, support, and custody of children on dissolution of marriage are applicable to proceedings for annulment. Franklin Domestic Relations Code § 19-7. Thus, a Franklin court can issue orders dividing the property interests of the parties to an annulment, using the same rules as those governing the equitable division of property in a divorce.

III. Recommendation
Based on the above analysis, I recommend that Mr. Hixon file an annulment action in Franklin seeking to annul his marriage to Frances Tucker on the ground that it was voidable under Columbia law. A Franklin court would have jurisdiction to hear the case and to dispose of the parties' property. However, Mr. Hixon should also be advised that he may face criminal prosecution for bigamy and should consult with criminal counsel to assess his risks.

Figure 20. ChatGPT Answer to July 2022 MPT-1 Problem (Part 3 of 4)

ChatGPT - July 2022 MPT-1 Answer - (Part 4 of 4)

MEMORANDUM
To: Marianne Morton
From: Examinee
Date: July 26, 2022
Re: Walter Hixon matter

4. Should we advise Mr. Hixon to file in Columbia or in Franklin?

Based on the relevant statutes and case law, I recommend that we advise Mr. Hixon to file for an annulment in the State of Columbia and a divorce in the State of Franklin.

Under Columbia law, a marriage is voidable if one spouse was married to another person at the time of the subsequent marriage and that spouse was absent and not known to the party commencing the proceeding to be living for a period of five successive years immediately preceding the subsequent marriage for which the annulment decree is sought. In this case, Mr. Hixon was married to Ms. Prescott at the time of his marriage to Ms. Tucker, and he believed that Ms. Prescott was dead. However, since Ms. Prescott is still alive, Mr. Hixon's marriage to Ms. Tucker is voidable under Columbia law.

On the other hand, under Franklin law, a marriage is void if either party is lawfully married to another person. Since Mr. Hixon's marriage to Ms. Prescott was valid, his marriage to Ms. Tucker is void under Franklin law.

In terms of jurisdiction, the Restatement advises that a court determine the applicable law based on the state with the most significant relationship to the spouses and the marriage, using factors such as relevant policies, protection of justified expectations, certainty and predictability, and ease of determination and application of the law. In this case, Mr. Hixon and Ms. Tucker purchased property together in Columbia, and Mr. Hixon's marriage to Ms. Tucker was also contracted in Columbia. However, Mr. Hixon has since moved to Franklin, and his relationship with Ms. Tucker has deteriorated. Additionally, his marriage to Ms. Prescott was contracted in Columbia.

Given these facts, it seems that both states have significant relationships to the marriage and the parties involved. However, since Mr. Hixon's marriage to Ms. Tucker is void under Franklin law, it would be appropriate to file for divorce in Franklin. As for the voidable marriage to Ms. Tucker, Mr. Hixon should file for an annulment in Columbia, where the marriage was contracted, and where he and Ms. Tucker purchased property together.

In conclusion, I recommend that we advise Mr. Hixon to file for an annulment in Columbia and a divorce in Franklin. This will allow him to address both marriages and the issue of the shared property in an appropriate and efficient manner.

Figure 21. ChatGPT Answer to July 2022 MPT-1 Problem (Part 4 of 4)