

THE AI ERA: LEVERAGING LLMs TO IMPROVE THE LAWYERS' CRAFT

Michael D. Murray

Spears Gilbert Associate Professor of Law
University of Kentucky, Rosenberg College of Law

AALS Annual Meeting 2024,

Section on Legal Writing, Reasoning, and Research
and

Section on Section on Technology, Law, and Legal Education

January 5, 2024



LEVERAGE: WHAT CAN YOUR ARTIFICIAL INTELLIGENCE LEGAL ASSISTANT DO FOR YOU

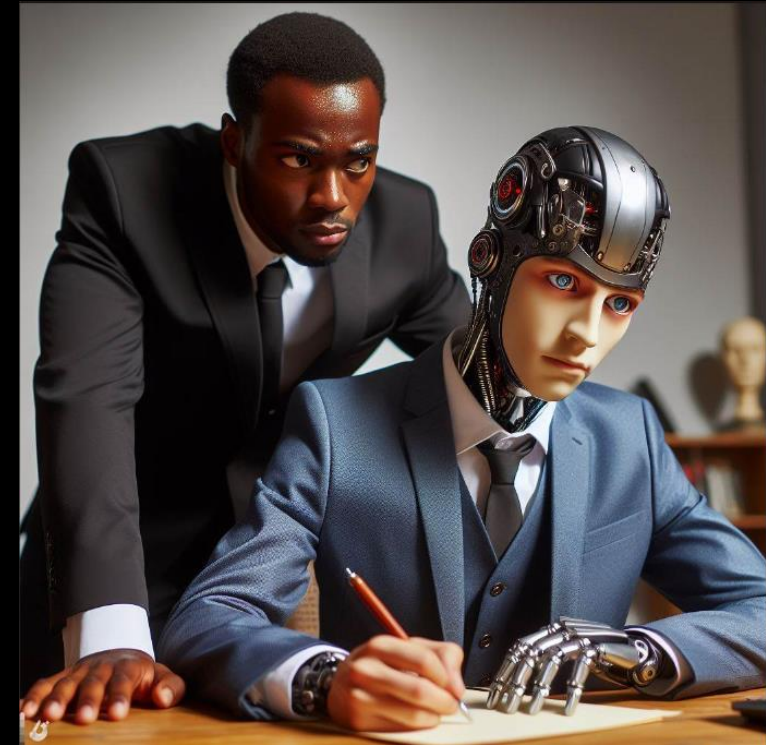
- **How Can AI help lawyers?**

- With **data** – analysis, due diligence, deposition or trial prep
- With **research** – finding sources and summarizing, explaining, and simplifying the information
- With **writing** – generating first drafts of crucial legal documents; writing complete drafts of simpler “SCUT” work; translating between languages
- With **images** – AI can draw and paint and create any image that you can describe
- With **knowledge** and **skills development**, one-on-one assisted learning



PUTTING THE LEVER AND THE FULCRUM IN THE RIGHT PLACE

- **Some ground rules for working with an AI**
- Treat the AI like an amazingly knowledgeable, well-read **legal assistant**; do not treat the AI like another **lawyer**.
- Only work with an AI that will show its **sources**. – Without knowing the sources you cannot directly check on the AI's work. (There are other means, such as using another AI to check on the first, but they take even more time to exercise the proper diligence)
- Ask for the AI to use its **knowledge** and **communication skills**; don't ask for its **legal opinions** or **legal reasoning**
 - That means think through your prompts very carefully



A.I. TAKEAWAYS

- Does AI think?
- **NO**
- Does it reason?
- **NO**
- Does it just copy and plagiarize?
- **NO** – with visuals, an AI **LEARNS** the correct way to compose and depict images.
- With language, an AI **LEARNS** the correct way in which to discuss legal issues. It is able to assimilate and draw from its learning to provide completely novel images and completely novel legal answers.
- But most AI seem not to want to read deeply into cases and other sources; they want to read what other sources say about the cases or sources.

WHAT DE-LEVERAGES AI SUPPORT

- **Hallucinations**
- **Incomplete answers** from weak sources
- **Guardrails** that close off the work
- **Inability** or refusal to check the **internet** for more up-to-date or complete information
- **Refusal** to follow **pretraining** or **prompt engineering** information and instructions
- At this point in time (early January 2024), it is better for a lawyer to ask for assistance from an AI only when the lawyer has a good idea what the response (the legal answer, the form, the document) and the sources that support it already should look like



BEST PRACTICES TO REAP THE UPSIDE AND CULL THE DOWNSIDE

Goal	Method	Does it work?
Achieve accurate work	Prompt engineering to push the AI to consult approach sources and information	For the AI with a browser function (BingChat, ChatGPT, Bard) this method can produce more accurate results, but the AI systems still will consult their own training data and web browsing protocols.
Produce useful, relevant work	Pretraining and chain of thought prompting can help the AI to produce more useful, relevant work	Giving the model examples to look at (e.g., few shot prompting) can
Eliminate hallucinations	?	Hallucinations are part of the working process of generative AIs. They strive to fill in blanks and when they run short or come up empty with desirable sources or information, they make up sources or information.

EVALUATING THE LEVERS: LEXIS+ AI

- **Lexis+ AI** – a mixture of experts model, with a chat-like interface. It is not a multi-modal (text and pictures) model.
- **PRO:** A curated dataset – supposed to form answers only from approved legal source material
- **CON:** It will not do internet searches (it may still find the information in Practical Guidance databases)
- **PRO:** It shows sources—whether it is showing all of its sources is not known.
- **CON:** When it writes something and cites sources it only cites a precious few—often only 2-4 sources are cited, but more will appear on a linked page
- It does not cite Bluebook cites, it gives links to the docs on Lexis
- **CON:** It has a real penchant for citing trial level opinions and unpublished opinions. It will cite as many out-of-jurisdiction authorities as controlling authorities.
- -It will cite completely inapposite sources, sometimes a whole set of them.
- -In general, I do not believe it “understands” hierarchy of judicial authority



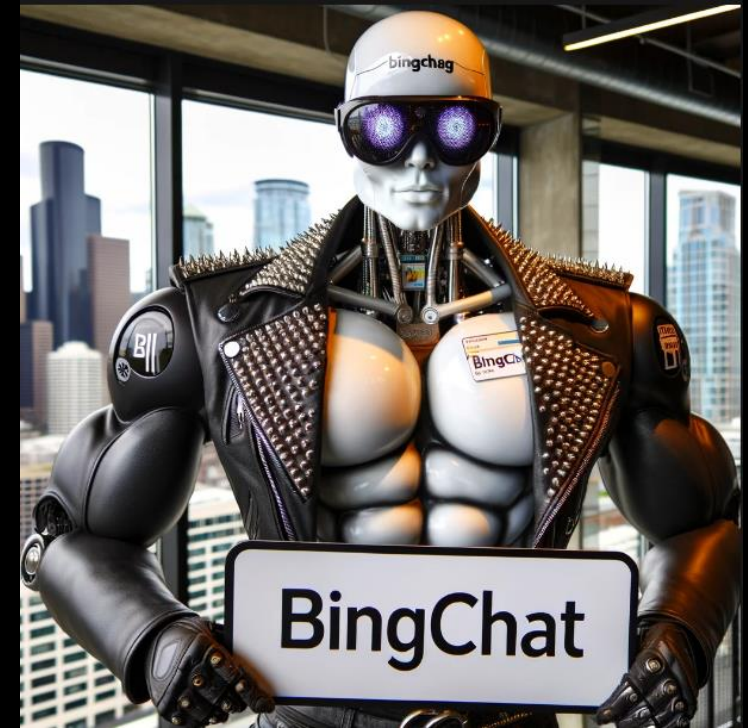
EVALUATING THE LEVERS: LEXIS+ AI

- **PRO:** Lexis+ AI does not appear to hallucinate as often as other services
- **CON:** It sometimes reads a word in the prompt wrong or does not apply it
- **PRO:** It can summarize or rewrite facts if you submit them in the prompt
- **CON:** It is hit or miss when trying to turn a raw set of facts into a Statement of Facts for a memo or brief.
- **PRO:** It will draft contract clauses
- **CON:** It does not seem to want to draft entire contracts or agreements
- **CON:** Lexis+ AI has guardrails, there are tasks it simply won't do, and sometimes it reports, "Responses are displayed when they meet our high standards for responsiveness and relevance. No response is available for this request."



EVALUATING THE LEVERS: BINGCHAT

- **BingChat** – runs on a trifecta of **Microsoft's Prometheus** (orchestrates the sources used to answer, including the Bing browser), **GPT-4** (OpenAI's top of the line model), and **Dall-E 3** (OpenAI's top of the line image generator). Thus, it is a multimodal AI system drawing on a mixture of experts model (GPT 4), with a chat-based interface.
- **PRO:** Very few guardrails. (Yes, I regard this as a "Pro")
- **PRO:** It will do internet searches – In fact it will do them even if you ask it not to.
- **PRO:** It shows sources—whether it is showing all of its sources is not known.
- **CON:** It tends to want to find 3 sources that it believes provide a "good" answer to the query, and then go ahead and draft the answer drawing on the writing abilities of GPT-4. Sometimes it uses 1-2 sources, sometimes 5 or more.
- **PRO:** It will provide active links to the sources so you immediately can check them out.
- **CON:** It does not cite Bluebook cites, and if you force it to try to do Bluebook cites it is fully capable of making up the information.
- -YES, it **hallucinates** – but much more often in "Creative Mode" these days compared to "Precise Mode."



EVALUATING THE LEVERS: BINGCHAT

- **CON:** BingChat is not at all picky about sources. It will eat from the dumpster of the Worldwide Web—the very opposite of a curated legal dataset. Bing will read anything—trial level opinions, unpublished opinions, legal blogs, Wikipedia entries, Tweets on X.
- It will refer to out-of-jurisdiction authorities if it has trouble finding controlling authorities. It will cite completely inapposite sources, sometimes a whole set of them.
- **PRO/CON:** BingChat will follow your instructions most of the time if you tell it to do its search in a more curated garden—e.g., FindLaw, Casetext, Cornell LII, Justia. But sometimes it gets stubborn and sticks to its own choices and does not do a good faith search in more curated sites even when directed.
- It does not “understand” hierarchy of judicial authority.
- **PRO/CON:** BING has more personality than most AIs. It has a definite attitude. It will get pissed at you and throw you out of a session. It will respond well to praise and compliments.



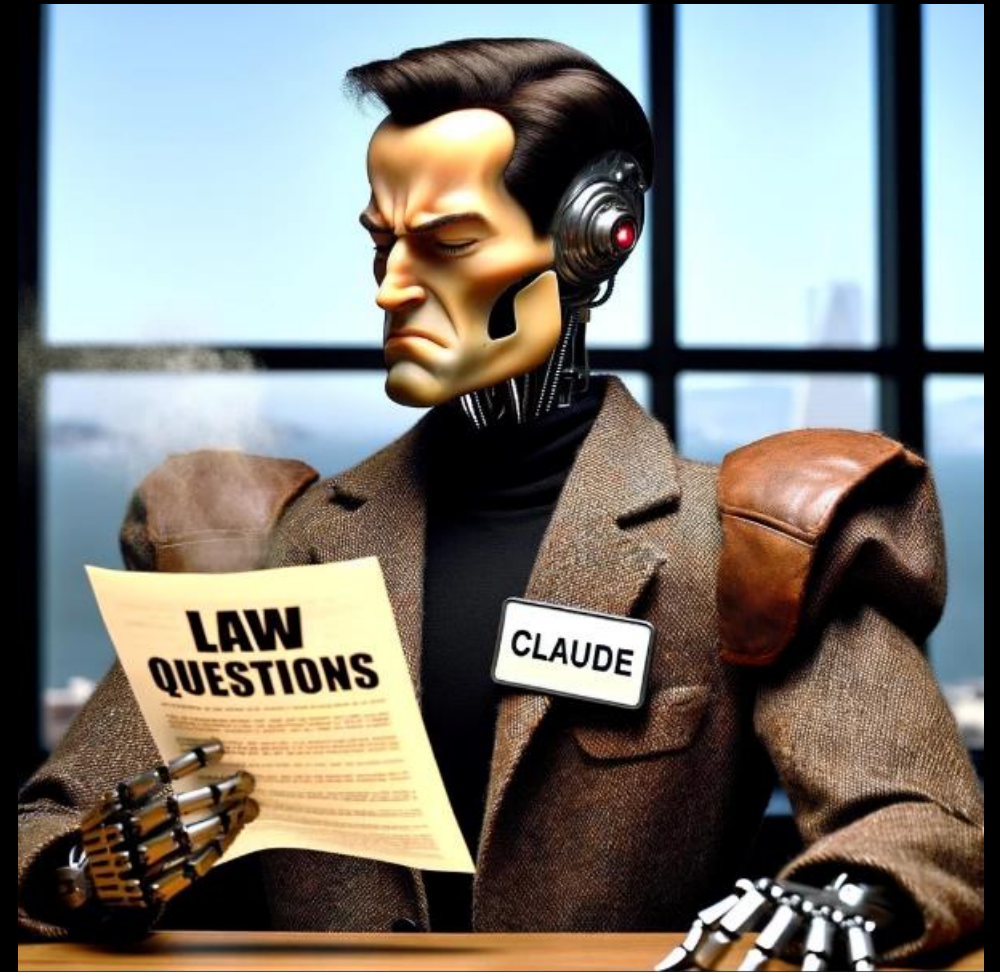
EVALUATING THE LEVERS: CHATGPT 4

- **ChatGPT 4** – OpenAI’s top of the line model, using a mixture of experts cohort of LLMs and **Dall-E 3** with **Bing browser** on reserve. Thus, it is a multimodal AI system with a chat-based interface.
- **PRO:** Few guardrails (but a few more than BingChat).
- **PRO:** It will do internet searches *if* you ask it to – i.e., prompt it to “Use your browser function.” Otherwise, it sticks to its training data.
- **CON:** It may not show its sources, even if you tell it to. It may make a half-hearted attempt to reference 1-3 sources at the end of its answer, but these often are not even linked so you can’t immediately check them out.
- **CON:** It does not cite correct Bluebook cites, and if you force it to try to do Bluebook cites it is fully capable of making up the information. So, yes, it **hallucinates**. (Just ask Mata’s (v. Avianca) attorneys).



THE “ALSO RAN” LEVERS - CLAUDE

- CLAUDE 2.1 is a LLM chat-based model built by Anthropic, a group of defectors from OpenAI who thought OpenAI was moving too fast (toward Artificial General Intelligence and possible extinction risk).
- Thus, CLAUDE's motto is “DON'T ASK. I WON'T TELL.”
- CLAUDE has a super-sized context window if you tend to ask 150,000 word questions.
- True to the Anthropic ethos, CLAUDE 2.1 is made of **GUARDRAILS**. Esp. re: Law.
- (But it will draft a contract – why I don't know).
- Whether it is brilliant at things other than answering legal questions, analyzing legal issues, writing case briefs or memos, I can't tell you. I only wanted to test it on the legal stuff.



THE “ALSO RAN” LEVERS - BARD

- **BARD** is Google's current mainline chatbot LLM model – Yes, *that* Google!!
- *-It should be sleek! It should be powerful! It should be dead sexy!!*
- BARD got a big update in early Dec. 2023. It was juiced with a brand-new model, Genesis Pro, that has stunning results comparable to ChatGPT **3.5**. (i.e., last year's state of the art model; i.e., it is stunning in the wrong sense of the word.)
- True to the Google safety ethos, Bard has a lot of GUARDRAILS. Not as many as Claude.
- But if you are patient and willing to rephrase and reframe your prompts, it will answer legal questions, analyze legal issues, write case briefs, memos, and contracts.
- **BUT IT HALLUCINATES frequently** (just ask Michael Cohen)



TASK ACHIEVEMENTS

Case Briefing	Contract Drafting	Overall writing skills
Best – BingChat, ChatGPT 4		Best – they all are great!
	Good – Bard, ChatGPT 4, Lexis+ AI (on specific clauses); Claude	
Not great – Lexis+ AI (makes unusual mistakes or omissions)	Not great – BingChat (inconsistent and imposes guardrails)	
Poor – Bard and Claude		

Legal Analysis	Memo Drafting	Dealing with facts
Best – Lexis+ AI	Best – Lexis+ AI	
Good – BingChat, ChatGPT 4	Good – BingChat, ChatGPT 4	Good – all of the current AIs can handle facts well
Not great – Bard (makes a lot of errors)	Not great – Bard (makes a lot of errors)	
Poor – Claude (tries not to do it)	Poor – Claude (tries not to do it)	

THE END

© 2024 MICHAEL D. MURRAY,
PROFESSORMICHAELMURRAY@GMAIL.COM

- Recent works:
- *Generative AI Art: Copyright Infringement and Fair Use* (forthcoming, 2023), 26 SMU SCI. & TECH. L. REV. ___, <https://ssrn.com/abstract=4483539>
- *Tools Do Not Create: Human Authorship in the Use of Generative Artificial Intelligence*, 15 CASE W. RESERVE J.L. TECH. & INTERNET ___ (forthcoming, 2023), <https://ssrn.com/abstract=4501543>
- *Artificial Intelligence and the Practice of Law Part 1: Lawyers Must be Professional and Responsible Supervisors of AI* (forthcoming, 2023), <https://ssrn.com/abstract=4478588>
- *Artificial Intelligence and the Practice of Law Part 2: Working With Your New AI Staff Attorney* (forthcoming, 2023-24), <https://ssrn.com/abstract=4478748>
- *Generative and AI Authored Artworks and Copyright Law*, 45 HASTINGS COMM. & ENT. L.J. 27 (2023)

