Generative Artificial Intelligence and the Law:

Trust but Verify

Cameron Powell

GREGOR | WYNNE | ARNEY, PLLC
What is Generative AI?

Pro Tip: It’s not a creature. It’s a tool.
Understanding Generative AI: Basics and Mechanisms

• **Definition**
  • A type of AI creates new data like the data it was trained on.

• **Limitations**
  • Quality of training data
  • Requires vast computational resources
  • Misleading or harmful content

• **Training Process**
  • *Data Collection*: Sourcing large datasets for training
  • *Data Preprocessing*: Cleaning and normalizing data
  • *Model Training*: Algorithms learn from the data through optimization
A Text Mashup Tool

• “The new programs mash up work done by human minds. What’s innovative is that the mashup process has become guided and constrained, so that the results are usable and often striking.”

• “an innovative form of social collaboration” – Jaron Lanier, futurist

• “like a version of Wikipedia that includes much more data, mashed together using statistics”

• “A bull%&#@$ generator”, Princeton University computer science professor Arvind Narayanan
Hi, I’m Cameron. It’s nice to ______ ____.

Cease and __________.  

Good morning! It looks like it’s going to be a beautiful ___.

Supercalifragilistic ______________.

In the United States ______ Court for the _______ _______ of Texas.
Jaron Lanier

• “[AI] can be thought of as illuminating previously hidden concordances between human creations, rather than as the invention of a new mind.”
• “What is civilization but social collaboration?”
• AI is a way of working together
• “mystery only makes mismanagement more likely.”
• The flexibility and unpredictability that suggest intelligence arises from simple mathematics. An LLM “contains a cumulative record of how particular words coincide in the vast amounts of text that the program has processed”
Large Language Models (LLMs)

We go to work by train

To write text, LLMs must first translate words into a language they understand.

Source: Financial Times: https://ig.ft.com/generative-ai/
Blocks of words = tokens

First a block of words is broken into tokens — basic units that can be encoded. Tokens often represent fractions of words, but we'll turn each full word into a token.

We go to work by train
In order to grasp a word’s meaning, in our example, LLMs first observe it in context using enormous sets of training data, taking note of nearby words. These datasets are based on collating text published on the internet, with new LLMs trained using billions of words.
Eventually, we end up with a huge set of the words found alongside work in the training data, as well as those that weren't found near it.
The Transformative Transformer

- The transformer unlocked LLMs’ abilities to parse and write fluently
- Radically sped up and augmented how computers understood language.
- Transformers process an entire sequence at once — sentence, paragraph, entire article
- Captures context and patterns better

- Google researchers outlined the transformer model in a short June 2017 paper that marked the start of the generative AI era.
Self-attention: how do words relate to one another?

A key concept of the transformer architecture is self-attention. This is what allows LLMs to understand relationships between words.

I have no interest in politics.
Which words are most important to understanding?

Self-attention looks at each token in a body of text and decides which others are most important to understanding its meaning.
Simultaneous Computing of Every Word in a Sentence

With self-attention, the transformer computes all the words in a sentence at the same time. Capturing this context gives LLMs far more sophisticated capabilities to parse language.
"Interest" seen as a noun

In this example, assessing the whole sentence at once means the transformer is able to understand that **interest** is being used as a noun to explain an individual's take on politics.
"Interest" as noun – for financial term

... the model understands *interest* is now being used in a financial sense.
And when we combine the sentences, the model is still able to recognise the correct meaning of each word thanks to the attention it gives the accompanying text.

For the first use of interest, it is no and in that are most attended.
Key to Advanced Text Generation

This functionality is crucial for advanced text generation. Without it, words that can be interchangeable in some contexts but not others can be used incorrectly.
Curb Your Enthusiasm

Effectively, self-attention means that if a summary of this sentence was produced, you wouldn't have **enthusiasm** used when you were writing about interest rates.

I have no interest in hearing about the rising enthusiasm rate of the bank.
The benefits of self-attention for language processing increase the more you scale things up. It allows LLMs to take context from beyond sentence boundaries, giving the model a greater understanding of how and when a word is used.

The dog chewed the bone because it was delicious.

- had a red collar
- was his owner's best friend
- loved playing fetch
- ate dinner at 6 pm
Major LLMs to Watch

• OpenAI
• Google’s PaLM model, which powers its chatbot Bard
• Anthropic’s Claude model
• Meta’s LLaMA
• Cohere’s Command
Word prediction

At its simplest, the model’s aim is now to predict the next word in a sequence and do this repeatedly until the output is complete.
Probability of next word

To do this, the model gives a **probability score** to each token, which represents the likelihood of it being the next word in the sequence.
Continual prediction

And it continues to do this until it is happy with the text it has produced.
Assigning probabilities to next phrases

With beam search, the model is able to consider multiple routes and find the best option.
Selecting a winner

This produces better results, ultimately leading to more coherent, human-like text.
Hallucinations

- The next word may seem plausible and coherent
- It isn’t always factually correct
- LLMs are not search engines looking up facts
- They are pattern-spotting engines that guess the next best option in a sequence.
- Predicting patterns can result in “hallucination”: made-up numbers, names, dates, quotes — even web links or entire articles.
The Promise

and the Peril
GPT-4 Passes the Bar Exam

35 Pages  •  Posted: 15 Mar 2023  •  Last revised: 5 Apr 2023

Daniel Martin Katz
Illinois Tech - Chicago Kent College of Law; Bucerius Center for Legal Technology & Data Science; Stanford CodeX - The Center for Legal Informatics; 273 Ventures

Michael James Bommarito
273 Ventures; Licensor, LLC; Stanford Center for Legal Informatics; Michigan State College of Law; Bommarito Consulting, LLC

Shang Gao
Casetext

Pablo Arredondo
Casetext; Stanford CodeX

Date Written: March 15, 2023

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 test-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
54% of in-house legal professionals support generative AI usage

Published June 30, 2023

**Will generative AI transform high-level legal work?**
How surveyed legal professionals responded to that question

- Strongly disagree: 19%
- Somewhat disagree: 31%
- Neither agree nor disagree: 19%
- Somewhat agree: 21%
- Strongly agree: 10%

**Leading generative AI use cases of interest**
% cited by corporate legal professionals

- Contract drafting & review: 76%
- Legal research: 69%
- Question answering service: 67%
- Brief or memo drafting & review: 64%
- Knowledge management: 62%
- Back office functions: 59%

Legal Dive/Wolters Kluwer and Above the Law data

Legal Dive/Thomson Reuters Institute data
AI has bias, just like humans.

Might it help that it doesn’t have ego?

Concerns about the ethical implications and potential dangers of AI language models like me are valid and important to discuss. Critics worry that such systems can be used for misinformation, propaganda, or to perform other harmful actions such as automating the generation of deepfake text.

Here’s a breakdown of some key points:

1. **Misinformation and Disinformation**: AI language models can be used to generate false information that appears credible. It’s a concern that people could misuse this to spread lies or create false narratives.
2. **Amplification of Biases**: Data used to train models like me may contain existing societal biases. Without proper mitigation, these biases can be perpetuated and even amplified.
3. **Autonomous Functionality**: The fear that an AI, especially if given enough autonomy, could act in unpredictable ways, is a concern for some critics.
4. **Data Privacy**: The model may inadvertently generate outputs that include information from the training data, potentially revealing sensitive information, although steps are taken to minimize this risk.
AI could choke on its own exhaust as it fills the web

The internet is beginning to fill up with more and more content generated by artificial intelligence rather than human beings, posing weird new dangers both to human society and to the AI programs themselves.

What’s happening: Experts estimate that AI-generated content could account for as much as 90% of information on the internet in a few years’ time, as ChatGPT, Dall-E and similar programs spill
How will generative AI change the practice of law?
How will the market for legal services change?

Lawyering will be disrupted by laypeople doing more self-service - especially at the lower end of the market.

Generative AI won’t replace lawyers as much as it will help lawyers who use it gain advantage over those who don’t.

High-dollar clients will still want human judgment.
How will the practice of law change?

- E-DISCOVERY IS ABOUT TO GET A LOT EASIER
- FASTER BRIEF WRITING
- CREATING AND SPOTTING ISSUES IN CONTRACTS
- OUTPUT MUST BE VERIFIED – SAME AS TODAY
How AI can help (Part 1)

• **Document Generation**
  • Automated drafting of legal documents, contracts, and agreements
  • Customization based on client-specific information

• **Legal Research**
  • AI-powered search through case law, statutes, and regulations
  • Generating summaries and insights from large legal datasets

• **Predictive Analysis**
  • Estimating case outcomes based on historical data
  • Risk assessment for various legal strategies
How AI can help (Part 2)

• **Client Interactions**
  • AI chatbots for initial client consultations
  • Automated follow-up and status updates

• **Litigation Support**
  • Real-time assistance during trials
  • Automated generation of legal arguments or counterarguments

• **Due Diligence and Discovery**
  • AI-driven analysis of documents for compliance
  • Sorting and tagging data for relevancy
How AI can help (Part 3)

• **Ethical Considerations**
  • Ensuring quality and accuracy of AI-generated advice
  • Maintaining client confidentiality

• **Time and Cost Efficiency**
  • Reducing the time spent on repetitive tasks
  • Potential for cost savings for both lawyers and clients

• **Continued Learning and Adaptation**
  • AI systems that evolve based on legal updates
  • Keeping legal professionals updated on new laws and regulations

• **Challenges and Limitations**
  • Over-reliance on AI tools
  • Ensuring ethical use and data security
Get off the blank page!

• A quick way to outsource the cognitively taxing work of ideation

• Letting us focus on what we do best: analysis, editing, responding
Don't let AI push you around.

Be firm.
Here is a sample table based on the 2020 U.S. Census data for Georgia counties. Please note that these figures are approximate and rounded:

<table>
<thead>
<tr>
<th>Rank</th>
<th>County</th>
<th>Total Population</th>
<th>% of State’s Population</th>
<th>% of African-Americans</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fulton</td>
<td>1,050,114</td>
<td>9.9%</td>
<td>45%</td>
</tr>
<tr>
<td>2</td>
<td>Gwinnett</td>
<td>942,000</td>
<td>8.9%</td>
<td>28%</td>
</tr>
<tr>
<td>3</td>
<td>Cobb</td>
<td>760,000</td>
<td>7.2%</td>
<td>27%</td>
</tr>
<tr>
<td>4</td>
<td>DeKalb</td>
<td>759,297</td>
<td>7.2%</td>
<td>55%</td>
</tr>
<tr>
<td>5</td>
<td>Chatham</td>
<td>289,430</td>
<td>2.7%</td>
<td>40%</td>
</tr>
</tbody>
</table>

Again, the numbers are based on the 2020 U.S. Census data, and you should consult the latest data for the most accurate information.

8:05

15 based on 2020 U.S. Census data for Georgia. Again, these figures are approximate and rounded:

<table>
<thead>
<tr>
<th>Rank</th>
<th>County</th>
<th>Total Population</th>
<th>% of State’s Population</th>
<th>% of African-Americans</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Clayton</td>
<td>292,256</td>
<td>2.8%</td>
<td>70%</td>
</tr>
<tr>
<td>7</td>
<td>Cherokee</td>
<td>258,773</td>
<td>2.4%</td>
<td>7%</td>
</tr>
<tr>
<td>8</td>
<td>Henry</td>
<td>234,561</td>
<td>2.2%</td>
<td>47%</td>
</tr>
<tr>
<td>9</td>
<td>Forsyth</td>
<td>244,252</td>
<td>2.3%</td>
<td>4%</td>
</tr>
<tr>
<td>10</td>
<td>Richmond</td>
<td>202,518</td>
<td>1.9%</td>
<td>55%</td>
</tr>
<tr>
<td>11</td>
<td>Muscogee</td>
<td>195,769</td>
<td>1.9%</td>
<td>48%</td>
</tr>
<tr>
<td>12</td>
<td>Hall</td>
<td>204,441</td>
<td>1.9%</td>
<td>7%</td>
</tr>
<tr>
<td>13</td>
<td>Bibb</td>
<td>153,159</td>
<td>1.4%</td>
<td>52%</td>
</tr>
<tr>
<td>14</td>
<td>Paulding</td>
<td>168,867</td>
<td>1.6%</td>
<td>22%</td>
</tr>
<tr>
<td>15</td>
<td>Houston</td>
<td>157,863</td>
<td>1.5%</td>
<td>29%</td>
</tr>
</tbody>
</table>

Again, the numbers are based on approximate figures from the 2020 U.S. Census, and you should consult the latest data for the most accurate information.

8:11

Now to calculate the percentages:

1. For the five largest counties:
   \[
   \frac{3,800,841}{10,710,000} \times 100 \approx 35.5\%
   \]

2. For the ten largest counties:
   \[
   \frac{5,033,201}{10,710,000} \times 100 \approx 47.0\%
   \]

So, approximately 35.5% of Georgia’s population lives in the five largest counties, and about 47.0% live in the 10 largest counties, based on 2020 U.S. Census data.

CAMERON

Please list counties ranked 6 through 15 as well

CHATGPT

Certainly! Here’s a continuation of the table to include counties ranked 6 through 15 based on 2020 U.S. Census data for Georgia. Again, these figures are approximate and rounded:

<table>
<thead>
<tr>
<th>Rank</th>
<th>County</th>
<th>Total Population</th>
<th>% of State’s Population</th>
<th>% of African-Americans</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Clayton</td>
<td>292,256</td>
<td>2.8%</td>
<td>70%</td>
</tr>
<tr>
<td>7</td>
<td>Cherokee</td>
<td>258,773</td>
<td>2.4%</td>
<td>7%</td>
</tr>
<tr>
<td>8</td>
<td>Henry</td>
<td>234,561</td>
<td>2.2%</td>
<td>47%</td>
</tr>
<tr>
<td>9</td>
<td>Forsyth</td>
<td>244,252</td>
<td>2.3%</td>
<td>4%</td>
</tr>
<tr>
<td>10</td>
<td>Richmond</td>
<td>202,518</td>
<td>1.9%</td>
<td>55%</td>
</tr>
<tr>
<td>11</td>
<td>Muscogee</td>
<td>195,769</td>
<td>1.9%</td>
<td>48%</td>
</tr>
<tr>
<td>12</td>
<td>Hall</td>
<td>204,441</td>
<td>1.9%</td>
<td>7%</td>
</tr>
<tr>
<td>13</td>
<td>Bibb</td>
<td>153,159</td>
<td>1.4%</td>
<td>52%</td>
</tr>
<tr>
<td>14</td>
<td>Paulding</td>
<td>168,867</td>
<td>1.6%</td>
<td>22%</td>
</tr>
<tr>
<td>15</td>
<td>Houston</td>
<td>157,863</td>
<td>1.5%</td>
<td>29%</td>
</tr>
</tbody>
</table>

Again, the numbers are based on approximate figures from the 2020 U.S. Census, and you should consult the latest data for the most accurate information.

CAMERON

What percent of the population in Georgia is African-American?
Legal implications of generative AI
Enterprise Risks of AI

- Data Privacy
- Confidentiality
- Cybersecurity
Who’s listening to you?

• Queries go to the vendor – same as Westlaw
Libel
When robots go rogue

ChatGPT invented a sexual harassment scandal and named a real law prof as the accused

The AI chatbot can misrepresent key facts with great flourish, even citing a fake Washington Post article as evidence.

By Pranshu Verma and Will Oremus
April 5, 2023 at 2:07 p.m. EDT

One night last week, the law professor Jonathan Turley got a troubling email. As part of a research study, a fellow lawyer in California had asked the AI chatbot ChatGPT to generate a list of legal scholars who had sexually harassed someone. Turley’s name was on the list.

Tech is not your friend. We are. Sign up for The Tech Friend newsletter.

The chatbot, created by OpenAI, said Turley had made sexually suggestive comments and attempted to touch a student while on a class trip to Alaska, citing a March 2018 article in The Washington Post as the
False Claims

OpenAI threatened with landmark defamation lawsuit over ChatGPT false claims [Updated]

ChatGPT falsely claimed a mayor went to prison.

ASHLEY BELANGER - 4/5/2023, 12:44 PM
KPMG lodges complaint after AI-generated material was used to implicate them in non-existent scandals

Senate inquiry into Australian consultancy industry concedes its integrity has been undermined by reliance on case studies generated by Google Bard AI tool

- Follow our Australia news live blog for latest updates
- Get our morning and afternoon news emails, free app or daily news podcast
When AI Lies — Who’s Liable?

You may be, if you publish it

What about the AI vendor?
Plagiarism and Copyright
The long read

The stupidity of AI

Artificial intelligence in its current form is based on the wholesale appropriation of existing culture, and the notion that it is actually intelligent could be actively dangerous

by James Bridle

---

IS A.I. ART STEALING FROM ARTISTS?

According to the lawyer behind a new class-action suit, every image that a generative tool produces “is an infringing, derivative work.”

By Kyle Chayka
February 10, 2023
Writers are unhappy

My books have been used to train AI bots — and I’m furious

The battle between writers and tech companies is alarmingly uneven, says Sathnam Sanghera

Sathnam Sanghera: “The technology that threatens to make writers obsolete is being trained using our work and we’re not being paid for it”

Michael Leckie

Sathnam Sanghera
September 29, 2023, The Times

Like almost everyone else, I’ve marvelled at what artificial intelligence can already do. But giving a commodity access to my work fundamentally cheapens the value of what I’ve created. It’s not just the ethics of this that worries me.

"As soon as you have the algorithms taking music from musicians, mashing it up into new music, and then not paying the musicians, gradually you start to undermine the economy because what happens to musicians now happens to everybody later," Lanier said.
Who owns the copyright?

• The Generative AI vendor?
• A work-for-hire employee?
• The creator whose creations were scraped for training data?
Google promises to take the legal heat in users’ AI copyright lawsuits

The company says it will take responsibility for its training data and the output of its foundation models.

By EMILIA DAVID | October 12, 2023 5:58 pm
Negligence & Emotional Distress
Bringing Back the Dead
Undue Influence

- What risks are created when vulnerable people use a creature that uses first-person pronouns?
- Should AI that purports to have an identity be labeled?
Legal Ethics
Ethical Considerations in Using Generative AI for Legal Practice

### Client Confidentiality
- Ensure AI systems have robust encryption and security measures
- Verify how data is stored and accessed

### Quality of Advice
- Ensuring AI-generated legal advice is accurate and reliable
- Lawyers should review and verify AI-generated content before client presentation

### Transparency
- Clients should be informed when AI tools are used in their case
- Clear disclosure about the capabilities and limitations of the AI system
Ethical Considerations in Using Generative AI for Legal Practice

Professional Integrity
- Guard against using AI to misrepresent facts or law
- Maintain the highest ethical standards as prescribed by legal codes

Automated Decision-making
- Scrutinize AI recommendations carefully
- Never relinquish final decision-making authority to AI systems

Continued Learning
- Stay updated with AI advancements to make informed choices
- Ongoing training on the ethical use of technology in legal practice
Don't use Generative AI (ChatGPT, Bard) to ask about cases

Try:

• Casetext’s CoCounsel
• LawDroid
• Clearbrief