Artificial Intelligence in the Legal Arena:

Opportunities, Challenges, and Ethical Considerations

[Dr. Ben C. Walter]

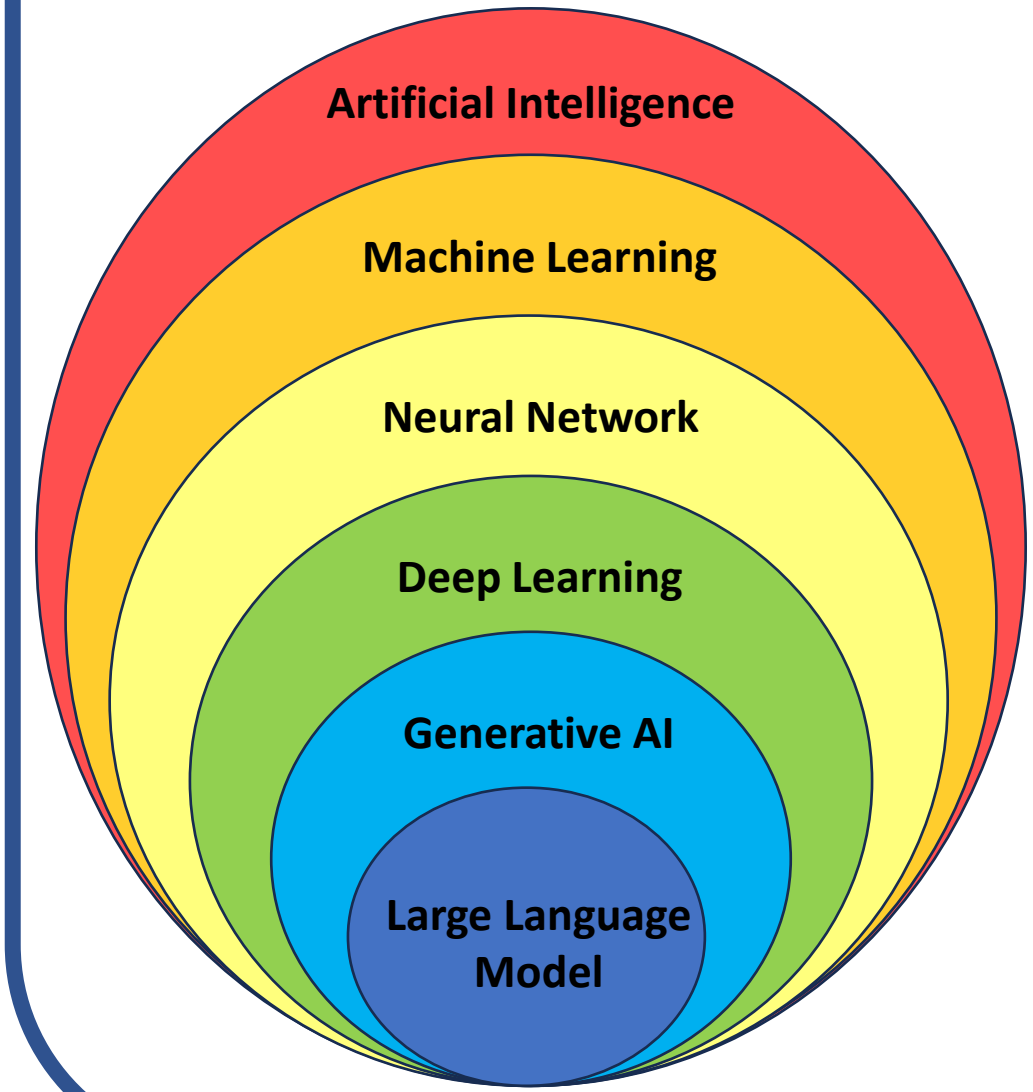
Associate Professor of Mathematics
University of the Virgin Islands

https://bit.ly/49NUMjD

Office of the Territorial Public Defender

December 12, 2024

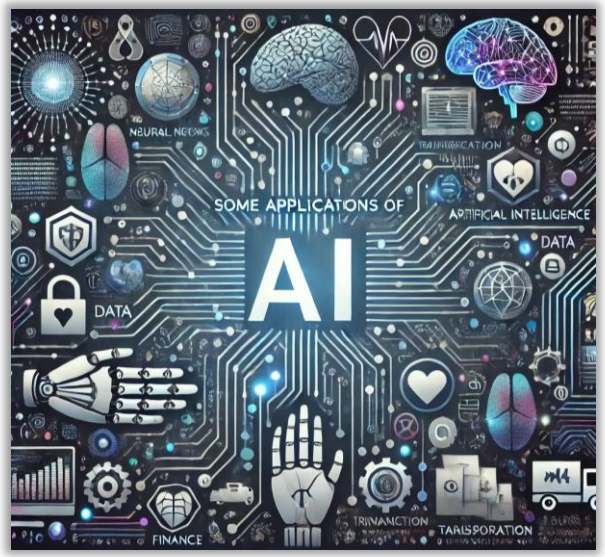
Artificial Intelligence



- **Artificial Intelligence (AI)** - system which performs tasks usually requiring human intelligence
- **Machine Learning (ML)** – system which learns patterns in data to make predictions without explicit instructions
- **Neural Networks** – machine learning models inspired by neurons in brains consisting of interconnected nodes
- **Deep Learning** – multi-layered neural networks with later / deeper layers capturing higher-order information
- **Generative AI** – systems creating new output – text, images, or audio – as opposed to merely describing or classifying
- **Large Language Models (LLM)** – deep learning networks trained on text data to understand / generate language

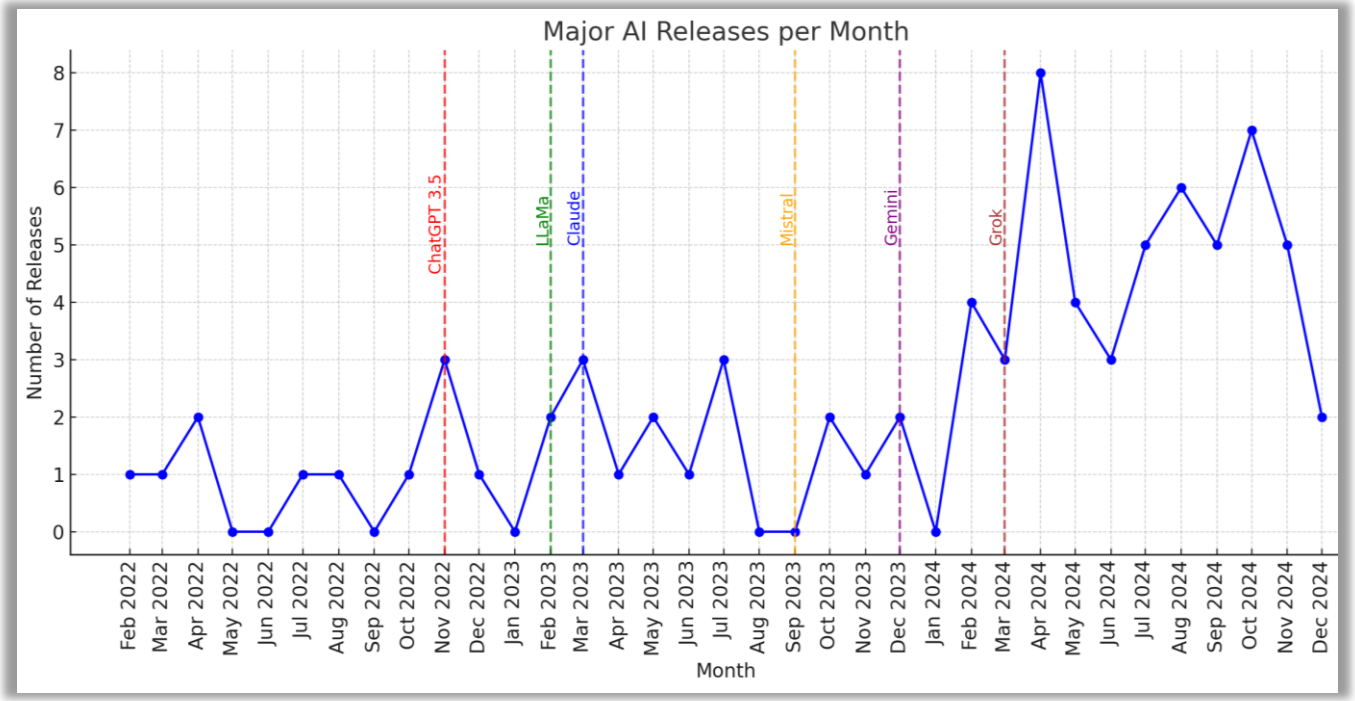
Some Applications of Artificial Intelligence

- Self-driving cars
- Smart home devices
- **Virtual assistants**
- Fraud detection
- Chemical research
- **Customer service**
- Weather forecasting
- Market prediction
- **Recommendation systems**
- Facial / object recognition
- Sentiment analysis

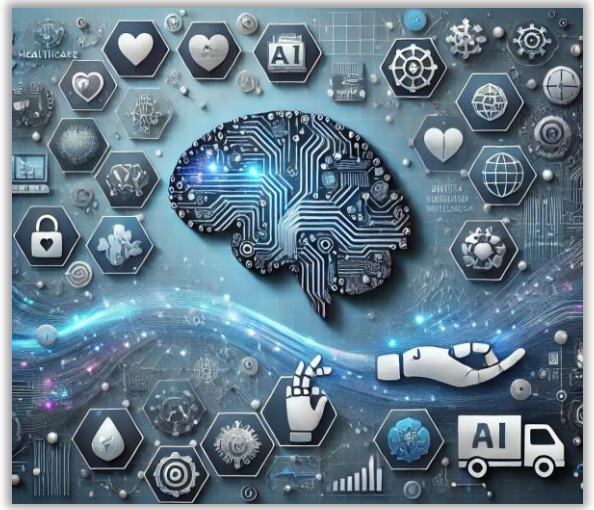


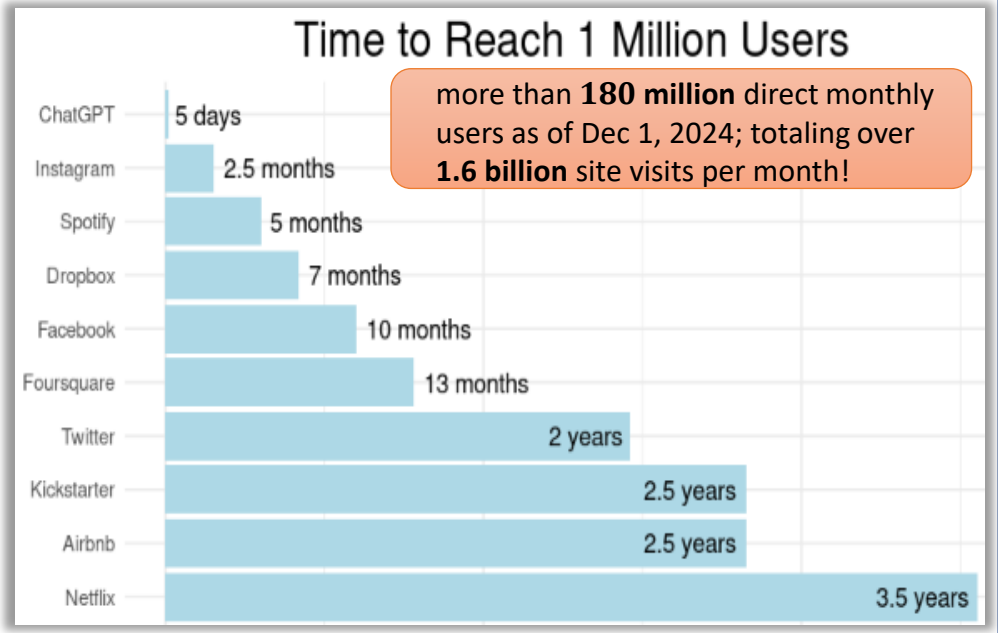
- **Speech / language translation**
- **Audio transcription** (meeting / video captions)
- **Content summary**
- **Content editing**
  (text, computer code)
- **Content generation**
  (text, images, video, audio, speech, code)
- Content moderation
  (message boards, chats, online game interactions)
- Personalized learning / tutoring
- Help with teacher grading / feedback
- **Chatbot**
- Spam filtering
- Network intrusion / virus detection
- Sportscasting / commentating (e.g. Wimbledon)

Growth of Generative Artificial Intelligence (Gen AI)



Data from https://nhlocal.github.io/AiTimeline/





Alan D. Thompson (https://lifearchitect.ai/) estimates that ChatGPT currently outputs the equivalent of *the entire printed works* of mankind (130 trillion books averaging 70k words per book)[1] every **two weeks**.

[1]Based on Google Books study

Major Gen AI Platforms

Dec 2024 AI rankings from https://lmarena.ai/

Note: **multimodal input / output** and **web search** is now standard for major AI!

ChatGPT (OpenAI)

May 2024 **GPT-4o**
- canvas (in beta) for collaboration
- API used for many other AI apps

Dec 2024 **o1**
- "chain of thought" reasoning
- advanced math / physics
- slower! more expensive!

Claude (Anthropic)

Oct 2024 **3.5+ Sonnet**
- focus on ethics, alignment, safety
- accuracy over creativity
- does not train on user interactions
- artifacts! *agentic* computer use!
- no web search / image generation

Gemini (Google)

Sept 2024 **1.5 Flash / Pro**
- targets research *(partner with OpenStax)*
- web search (with citations)
- integrated with Google ecosystem

LLaMA (Meta)

Dec 2024 **LLaMa 3.3**
- open source and "open weight"
- code & weights used in many other AI's
- text-only input (text or image output)

Meta AI
- in Facebook, Messenger, Instagram, etc.

LE CHAT MISTRAL (Mistral AI)

Emphasis on free / open-source models

Nov 2024 **Mistral / Pixtral / Codestral / Mathstral**
- canvas interface (like GPT-4o Pro)
- web search (with citations)

零一万物 (01.AI)

Oct 2024 **Yi-Lightning**
- most powerful Chinese AI

Nexusflow (Nexusflow Solution)

Nov 2024 **Athene-V2**
- open source and "open weight"

Grok (xAI)

Aug 2024 **Grok 2**
- creativity over accuracy

Outline of Generative AI Process

Prompt

Write a haiku about cats.

tokenizer

Tokens
(phrases, words, subwords)

Output of https://platform.openai.com/tokenizer

```
Write a haiku about cats.
[10930, 261, 2472, 11169, 1078, 28854, 558]
```

Modern Generative AI's use between 100,000 and 200,000 tokens!

Rule of thumb:
1 token ≈ 4 characters

embedding

Semantic Vectors
(numerical representation of meaning)

Embedding size of GPT-3 was about 12,000 dimensions (numbers)
- **Similar** tokens are assigned **nearby** numbers
- **Adding properties** corresponds to **addition of numbers**

iterate

transformer layers

self-attention

Attention mechanism relates tokens to each other to find most relevant token
(introduced by Google in 2017)

feedforward

Feed-forward network multiplies vectors by **weights** and combines them
*Current top AI's use over 1 trillion learned **weights** (parameters)*

Next Token Prediction
(list of possible tokens and probabilities – pick one)

Each transformer layer captures different features of the input
word level → phrases → sentences → overall meaning
Current top AI's use 80-100 layers with size around 8,000

Method used to choose next token affects *"creativity"* of the model.

Generated Output

```
[2073, 3295, 409, 306, 290, 28479, 6038, 412, 35689, 175529, 67067, 290,
15095, 4856, 412, 165023, 875, 23757, 558]
Whiskers in the moonlight,
Soft paws tread the quiet night,
Mystery alive.
```

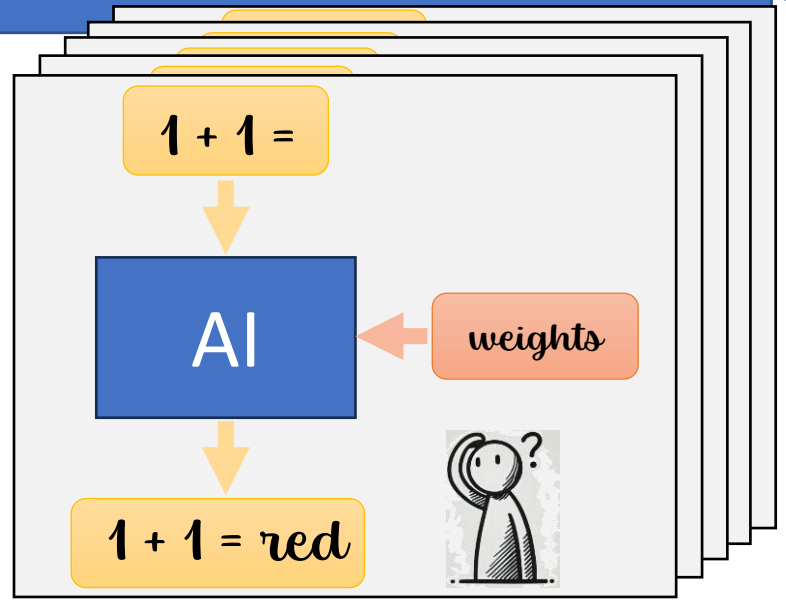
de-tokenizer

Response

Whiskers in moonlight,
Soft paws tread the quiet night,
Mystery alive.

Training AI – Computing Weights

≈ 1 trillion numbers encoding all of AI's "knowledge"



- Use massive dataset: diverse and representative
- Initialize model with random weights
- Gradually update weights as follows.
  - Using part of data, run model to predict next token
  - Compare to actual next token
  - Adjust weights to make actual value more likely
- Repeat **multiple times** over **all data**!

Very expensive! (≈ $100 million; weeks or months of time; growing exponentially)

Modifying Models

Fine Tuning

- Additional training further modifying computed weights with new data
- Less resource-intensive than full training
- Risks "catastrophic forgetting"

Adapter-Based Tuning

- Adding new layers onto network without changing existing weights
- Very lightweight

Fast and inexpensive

Summary

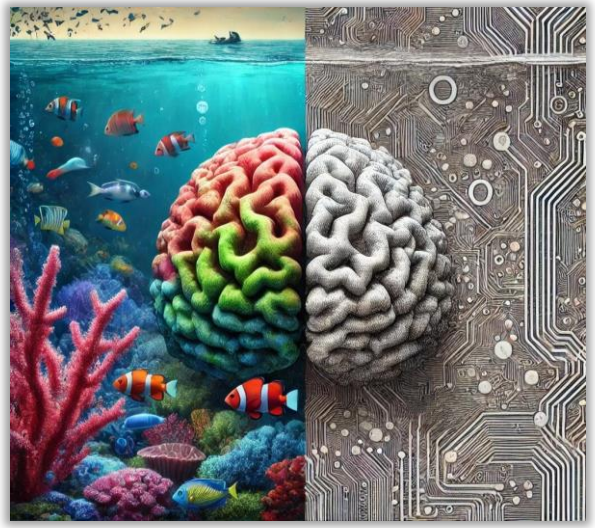
Generative AI works like a very fancy autocomplete

ChatGPT views the following prompts similarly:

- $2 + 2 =$
- Complete the statement: "$2 +$ …"
- Complete the statement: "I like cats and …"
- What is the next line:

"Hey, I just met you, and this is crazy
But here's my number …"

So call me, maybe?



AI are not "computers"

- **Creative** rather than **Computational**
- **Random** rather than **Deterministic**
- Not "*does exactly what you tell it*"
- Not "*garbage in – garbage out*"

Expectations of AI vs Reality of AI

The AI we were told to expect…

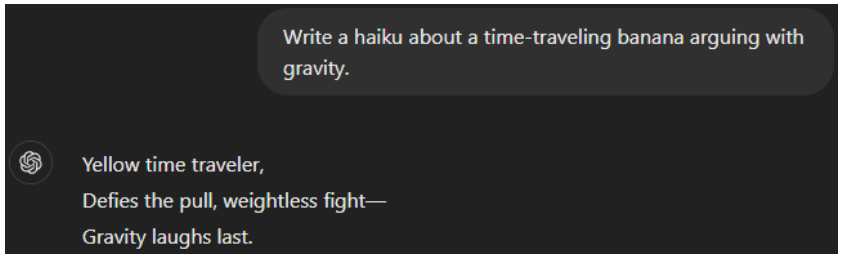
Accurate, Literal

Star Trek: The Next Generation

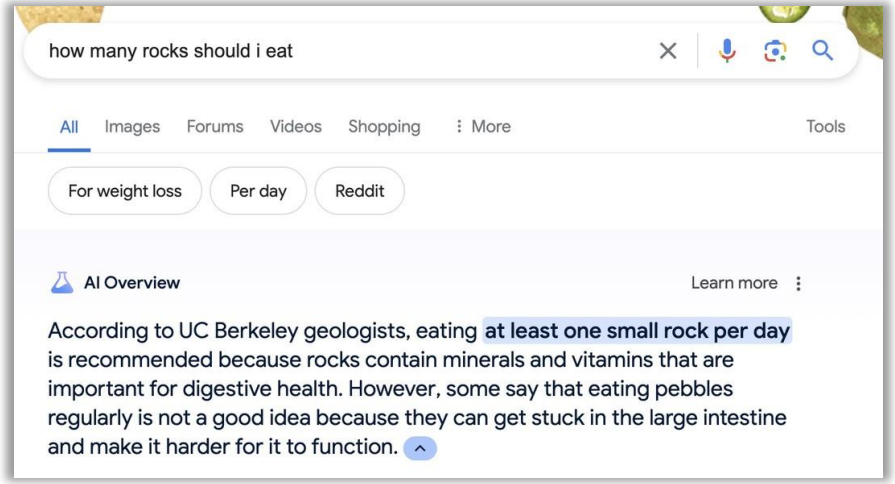
Emotionless

Terminator 2: Judgement Day

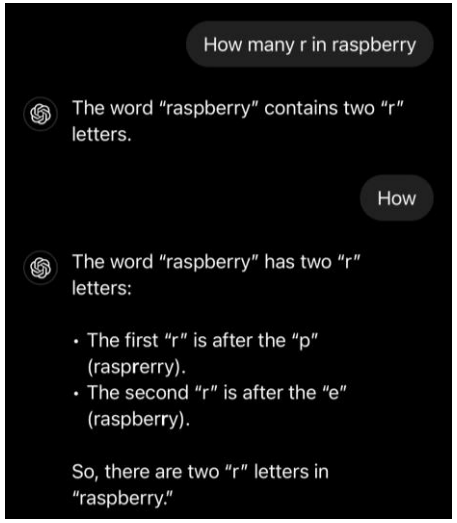
The AI we were given…



ChatGPT, Dec 10, 2024



Google AI Overview (Gemeni), May 2024



OpenAI Developer's forum, Aug 2024
"Incorrect count of 'r' characters…"

- Output may not always be accurate. You should not rely on Output from our Services as a sole source of truth or factual information, or as a substitute for professional advice.
- You must evaluate Output for accuracy and appropriateness for your use case, including using human review as appropriate, before using or sharing Output from the Services.

- OpenAI terms of use

Some problems with artificial intelligences

- **Accuracy / Accountability.**
  False output (i.e. “hallucinations”) – recommendation: “human-in-the-loop”

- **Privacy / Confidentiality.**
  Prompts and attached data may be used for further training; various methods can expose training data (glitch tokens, overfitting)

- **Bias.**
  Characteristics of training data will appear in output; implicit bias has been observed in AI outputs inferring ethnicity from names

- **Editorializing**
  AI likes to construct a narrative

Accuracy / Accountability

- **AI Errors ("Hallucination")**
  "AI can introduce insensitive, incorrect, offensive, or stereotype-based language, gender misattribution, and even diagnostic errors"[1]
  Especially likely on questions which are open-ended or outside training
- **Lack / fabrication of references / data**
  Rather than factual references AI may supply "example references"
  10% of my statistics students had AI make up "demonstration data" rather than use uploaded data
- **Gaps in training data**
  ChatGPT only trained on data up to October 2023
  Included training data may still have gaps / biases
- **Editorialization**
  AI can "use potentially judgmental language, omit important details, add plausible but incorrect information, and overstep their purview, making diagnostic suggestions rather than generating summary."[1]

ChatGPT does not "consult it's database"
It does not "look up" any facts

AI output is all "hallucination" – but we only complain when it is wrong…

[1] Altschuler S, Huntington I, Antoniak M, Klein LF. Clinician as editor: notes in the era of AI scribes. Lancet. 2024

Privacy / Confidentiality

AI can be induced to leak training data. Massive training data sets can also be directly stolen or exposed via human intervention or mistakes.

- **Training data privacy**
  Internal AI may be trained with confidential data; e.g. names, phone numbers, addresses, salary. Known attacks can extract this information.
  - see *Nasr. Scalable extraction of training data… (2023)*

- **Query data privacy**
  Many AI will *self-improve*, training on supplied query or analysis data.
  Any non-anonymized data exposed to the AI is at risk of leakage.

See “Glitch tokens”
https://arxiv.org/abs/2404.09894

Bias

AI reflect any biases present in the data they are trained on, **perpetuating** existing inequalities. *(Generated content is based on learned patterns.)*

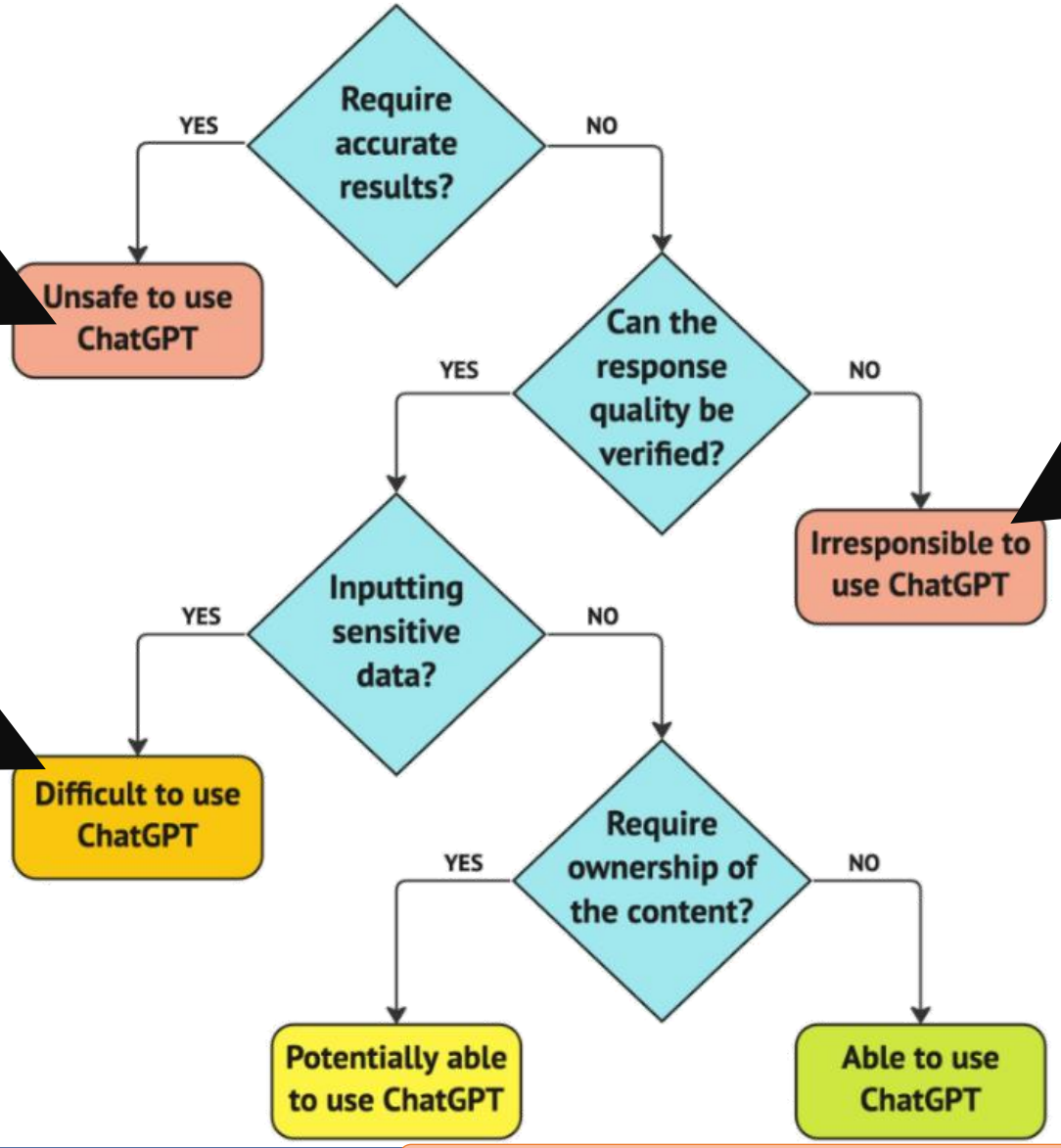
- **Bias in predictive / automatic decision making**
  Predictive algorithms trained on historic data (e.g. historic crime data or cv's of successful applicants) don't account for changing demographics.
- **Bias in underrepresented language / dialect queries and responses**
  Low prestige languages and dialects receive less informative responses. This perpetuates sociolinguistic inequality.
- **Discriminatory results from machine learning algorithms due to training data bias**
  E.g. facial recognition prone to errors on darker skin tones yielding discriminatory outcomes; medical algorithms less effective on underrepresented groups yielding poor diagnoses; image generation reflecting training bias.
- **Identity bias**
  AI have been shown to give biased responses based on racial / gender identities *inferred only from names*

When is ChatGPT Appropriate?

You are here.

Or here.

Or here.



Flowchart source: DataCamp. Introduction to ChatGPT, Ch 2 Adopting ChatGPT

AI for Attorneys

VI Consortium. July 17, 2004 reports that OTPD has 1,020 open cases "current case load exceeds reasonable capacity for effective representation"

Attorney uses of AI[1]

- Drafting communications (templates)
  (e.g., memos, emails, correspondence to opposing counsel, etc.)
- Conducting legal research ("e-discovery" search of images, video, text, pdf, etc)
- Summarizing legal narratives
- Reviewing legal documents / contracts
- Drafting legal contracts (templates) or proofreading completed documents
- Conducting due diligence
- Reviewing discovery (i.e. transcribing police bodycam footage - JusticeText)
- Negotiating / redlining contracts
- Preparing case filings (e.g., pleadings, motions, jury instructions, etc.)

Top AI Tools Specifically for Attorneys *(by market value):*

Harvey AI	https://www.harvey.ai	Backed by OpenAI
Robin AI	https://www.robinai.com	Based on Anthropic Claude
Clio Manage / Duo	https://www.clio.com	Built on Microsoft Azure / OpenAI
Thomson Reuters Casetext / CoCounsel		

Also many legal AI tools targeting **public**

i.e. Genie AI https://www.genieai.co/

Unauthorized practice of law????

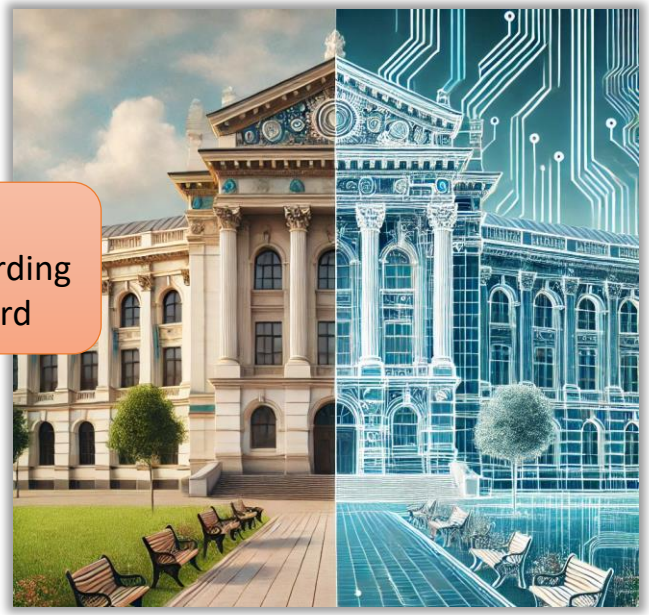
[1] https://pro.bloomberglaw.com/insights/technology/ai-in-legal-practice-explained

AI in the Courtroom

AI companies also offer solutions tailored to court systems

- **Transcription**
  Robust voice transcription AI can supplement court reporters
  Especially useful in cases of interruptions / overtalking
  Can also give searchable transcripts of audio / video evidence
  (see For The Record https://fortherecord.com)

Note: AI transcript is linked to audio recording which is official record

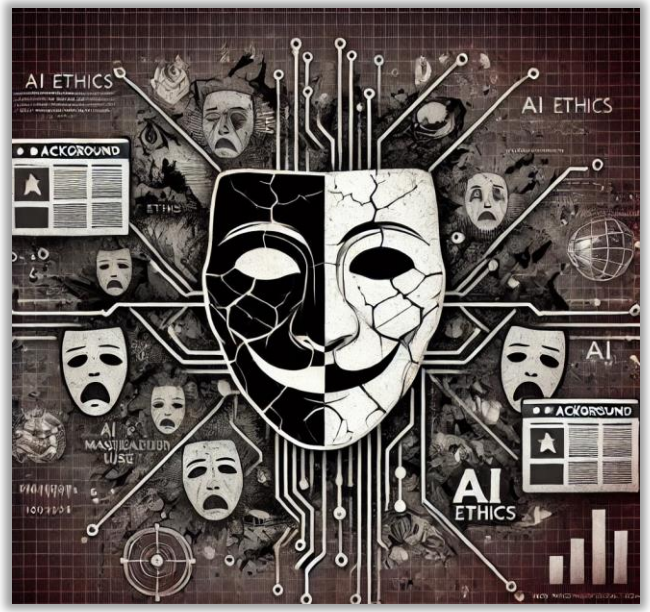


- **Translation (Speech-to-Speech)**
  AI translation can remove obstacles in court proceedings
- **Text-to-Speech**
  AI services can read documents for the illiterate or blind
- **Auto-Docketing *(especially for e-filed documents)***
  AI is good for tedious mechanical tasks (Palm Beach County, FL; Terrant County, TX)
- **Judicial Guidance**
  AI providers pushing to give *first pass* evaluation of briefs and aid in research (e.g. Clio Casetext)
  Machine learning tools have long been pushing to give bail and parole recommendations...
- **Public Inquiry Chatbot**
  New Jersey, JIA chatbot (80% accuracy?); also Arizona, New Mexico

https://www.ncsc.org/__data/assets/pdf_file/0032/97187/Court-Chatbots.pdf

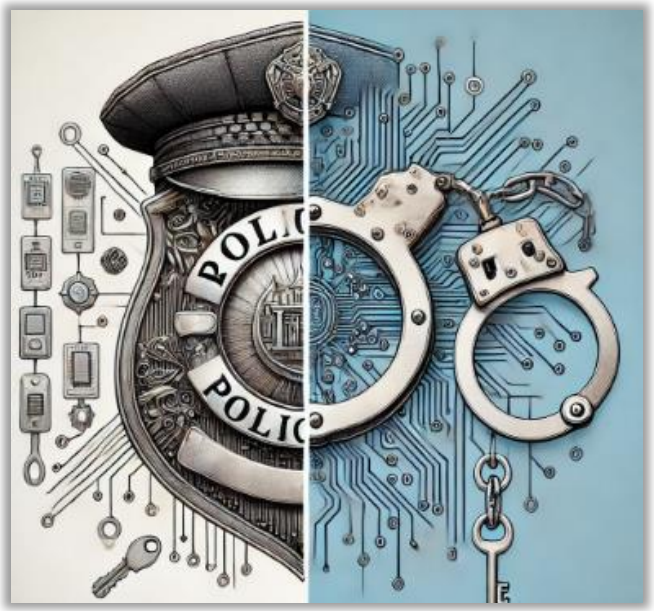
AI in Crime / Fraud

- **Deepfake pictures / audio / videos**
  Blackmail, impersonation, simulated ransom, nonconsensual pornography
- **Targeted phishing attacks**
- **Automated hacking attacks**
- **Child pornography image generation**
- **Forgery / Copyright violation**
  see Etsy, Virtual influencers
- **Fraudulent reviews**
  see FTC complaint vs. Rytr
- **“Algorithmic collusion”**
  AI driven house price suggestions accused of price-fixing.
- **AI “Snake Oil” Scams**
  Fraudulent “AI business opportunity” / “guaranteed income from AI” schemes
  see Ascend, Ecommerce Empire Builders, FBA Machine, and many others….



AI in Policing

Many police technology companies / startups are pitching AI applications



- **Integrated surveillance**
  Data from cameras, license plate readers, radiological sensors, and drones gathered and linked by AI for search and alerts (e.g. Axon, Flock)
  See also New York Police Department's Domain Awareness System

- **Predictive policing**
  Automated response / directives based on integrated surveillance / historical data; shown multiple times to be biased and discriminatory

- **Police self-review**
  Automated interaction / incident clipping /compilation from police bodycam footage (e.g. Truleo "police officer asst")

- **Auto-generation of police reports *["drafts"]***
  Generates police reports for incidents from police body camera footage (e.g. Axon's Draft One service)
  Likely serious issues with hallucination, editorialization, bias, loss of independent narrative, corrupting / influencing officer memory / impression of incident (similar to showing an eye-witness captured video before interviewing)

"… the way US police are adopting AI is inherently chaotic… The police-tech companies that serve them will build the tools police departments find attractive."
-- MIT Technology Review. Nov 19, 2024

American Bar Association Formal Opinion 512

Standing Committee on Ethics and Professional Responsibility. July 29, 2024

AI vs ABA Model Rules of Professional Conduct

Competence.

Lawyers should have reasonable understanding of capabilities and risks of generative AI

Uncritical reliance on generative AI output *without verification* violates duty of competent representation

Confidentiality.

Lawyers must protect client information when using generative AI tools (be wary of sharing data with AI)

A client's informed consent must be obtained before inputting information related to client into AI tool

Communication.

Clients should be informed about use of AI if output influences a significant decision in representation
(similar to disclosing involvement of temporary lawyers making significant contributions)

https://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/ethics-opinions/aba-formal-opinion-512.pdf

American Bar Association Formal Opinion 512

Standing Committee on Ethics and Professional Responsibility. July 29, 2024

AI vs ABA Model Rules of Professional Conduct

Meritorious Claims, Candor, Court Responsibilities.

Lawyers must verify that AI outputs (especially legal citations and analyses) are accurate and do not mislead courts

Lawyers should be prepared to explain to courts how they used AI, sources of information, and level of review

Supervisory Duties.

Legal professionals must ensure proper use of generative AI tools by subordinates and nonlawyers to ensure compliance with guidelines on competence and confidentiality

Fees.

Fees for generative AI assisted work must be reasonable and reflect **actual time** spent or direct costs incurred

Overcharging clients for efficiencies gained through generative AI use violates ethical rules

https://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/ethics-opinions/aba-formal-opinion-512.pdf

Some references / Further reading

https://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/ethics-opinions/aba-formal-opinion-512.pdf

https://www.ibanet.org/PPID/Constituent/Multi-displry_Pract/anlbs-ai-report

https://www.ncsc.org/__data/assets/pdf_file/0027/98910/JTC-AI-paper-update-3.5.24.pdf

https://news.bloomberglaw.com/us-law-week/courts-remain-skeptical-of-lawyers-use-of-chatgpt-in-litigation

https://www.clio.com/resources/ai-for-lawyers/

https://www.americanbar.org/groups/law_practice/resources/law-technology-today/2024/ai-tools-for-legal-work-claude-gemini-copilot-and-more/

https://www.technologyreview.com/2024/11/19/1106979/how-the-largest-gathering-of-us-police-chiefs-is-talking-about-ai/

https://www.technologyreview.com/2024/04/16/1090846/ai-police-body-cams-cops-transparency/

https://www.eff.org/deeplinks/2023/12/artificial-intelligence-and-policing-year-review-2023

https://a16z.com/navigating-the-high-cost-of-ai-compute/

https://www.prc.nm.gov/wp-content/uploads/2024/11/5-AI-in-Litigation-CLE-Final.pdf

https://mindmatters.ai/2023/04/20-ways-ai-enables-criminals/

https://www.ftc.gov/news-events/news/press-releases/2024/09/ftc-announces-crackdown-deceptive-ai-claims-schemes

Altschuler S, Huntington I, Antoniak M, Klein LF. Clinician as editor: notes in the era of AI scribes. Lancet. 2024 Nov 30;404(10468):2154-2155.

Nasr, M., Carlini, N., Hayase, J., Jagielski, M., Cooper, A. F., Ippolito, D., ... & Lee, K. (2023). Scalable extraction of training data from (production) language models. arXiv preprint arXiv:2311.17035.