Artificial Intelligence in the Legal Arena:

Opportunities, Challenges, and Ethical Considerations

[Dr. Ben C. Walter]

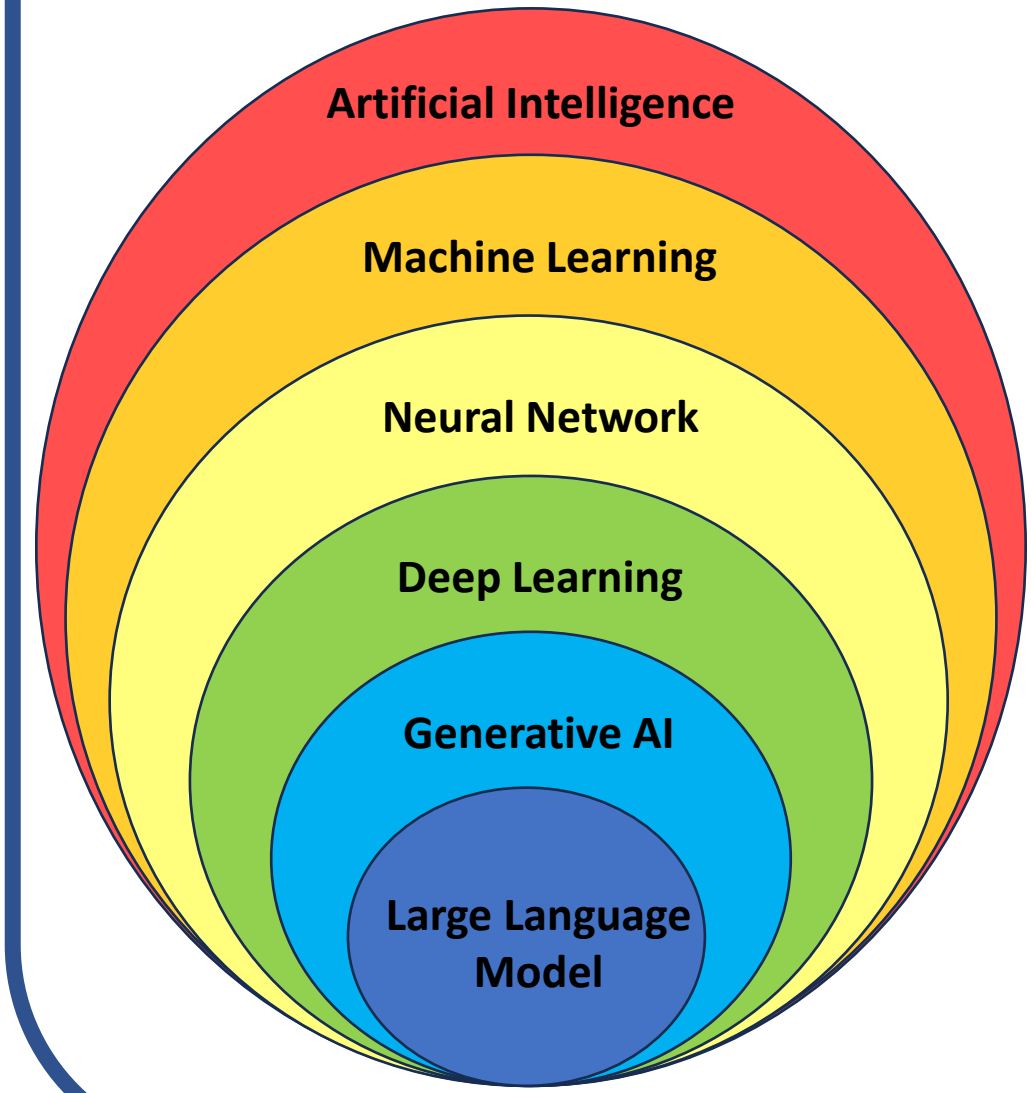
Associate Professor of Mathematics
University of the Virgin Islands

https://bit.ly/49NUMjD

Office of the Territorial Public Defender

December 12, 2024

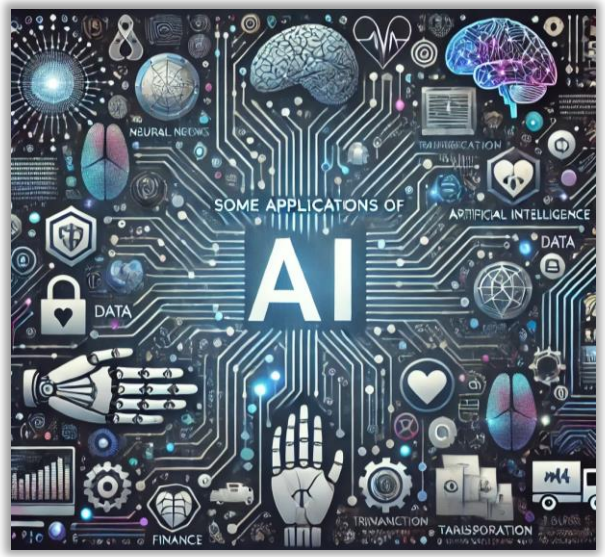
Artificial Intelligence Subcategories



- **Artificial Intelligence (AI)** - system which performs tasks usually requiring human intelligence
- **Machine Learning (ML)** – system which learns patterns to make predictions without explicit instructions
- **Neural Networks** – models inspired by neurons in brains consisting of interconnected nodes
- **Deep Learning** – multi-layered neural networks with later 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

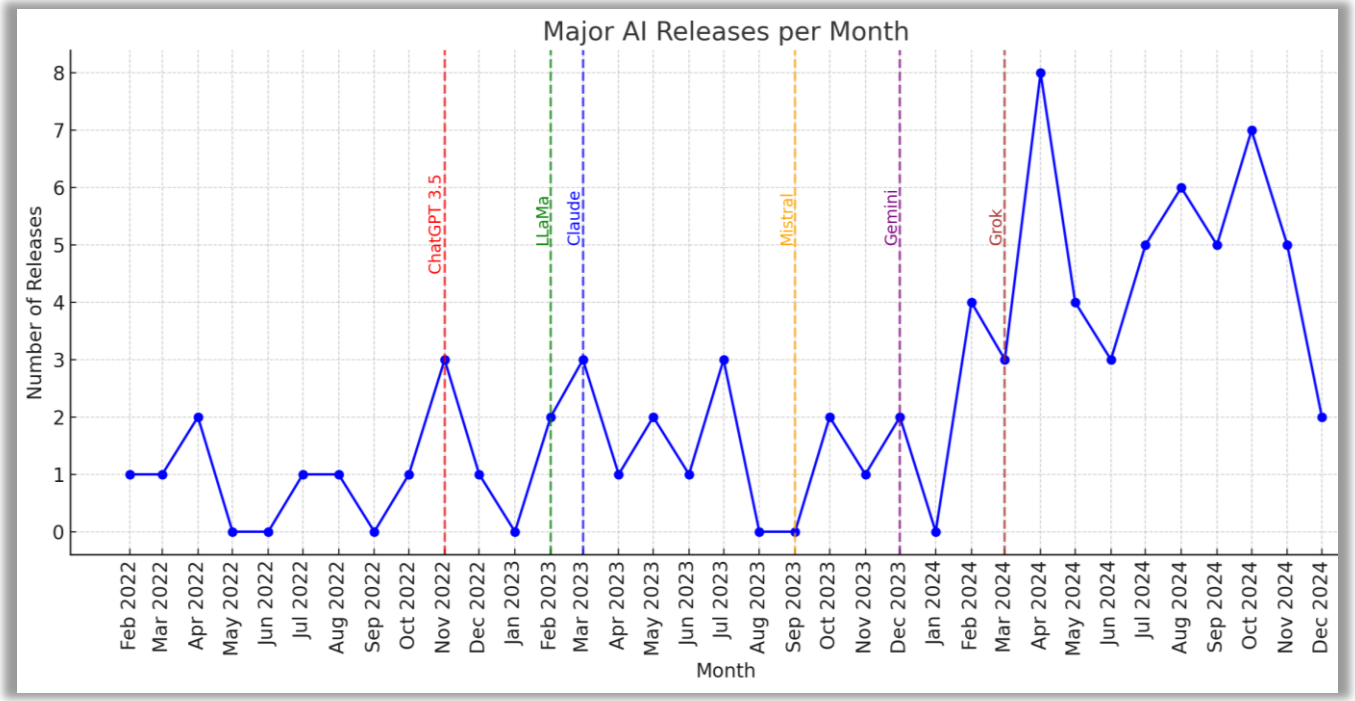
Some Applications of Artificial Intelligence

- 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

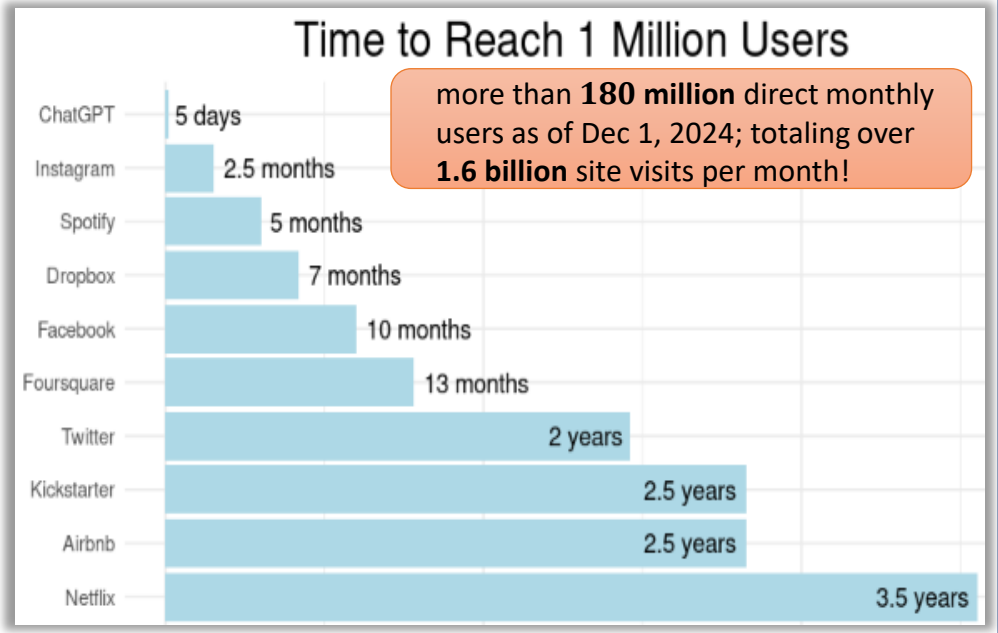


- **Speech / language translation**
- **Audio transcription** (meeting / video captions)
- **Content summary**
- **Content editing**
  (text, computer code)
- **Content generation**
  (text, images, video, audio, speech, code)
- Content moderation
  (message boards, chats, online game interactions)
- 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 in **all** 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

Rough Outline of Generative LLM AI

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

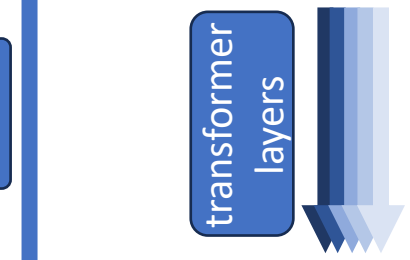
embedding

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.

de-tokenizer

Response

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

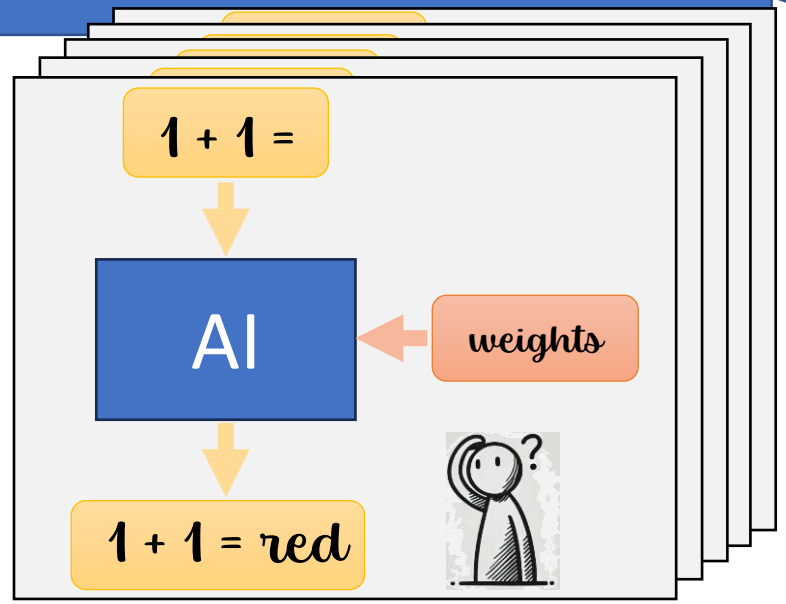
Generated Output

```
[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.
```

Training AI – Computing Weights

- Use massive dataset: diverse and representative
- Initialize model with random weights
- Gradually update weights
  - Run model to predict next token in data
  - Compare to actual value
  - 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 Training

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

Summary

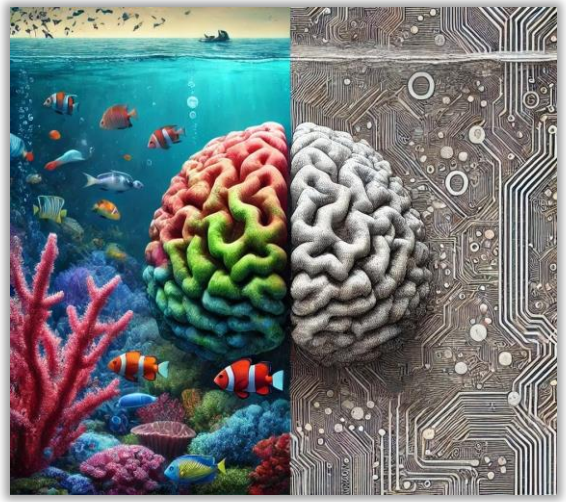
Generative AI works like a <u>very fancy</u> autocomplete

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?



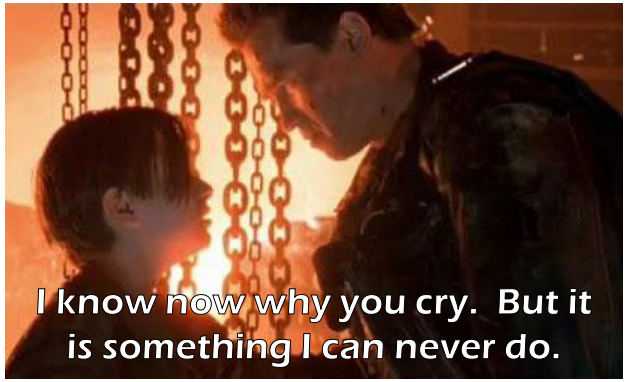
AI are not "computers"

- Not computational
- Not deterministic
- Not "garbage in – garbage out"
- Not "does exactly what you tell it"

Expectations of AI vs Reality of AI

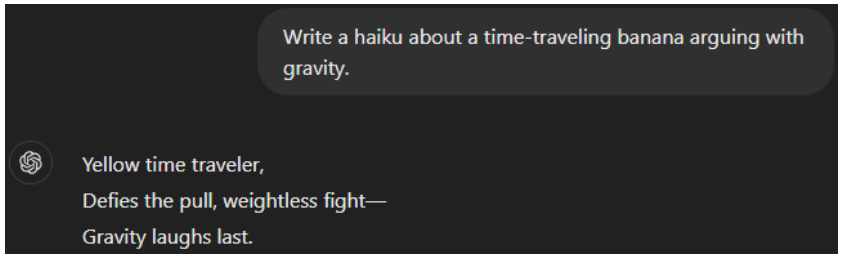
The AI we were told to expect…

Star Trek: The Next Generation

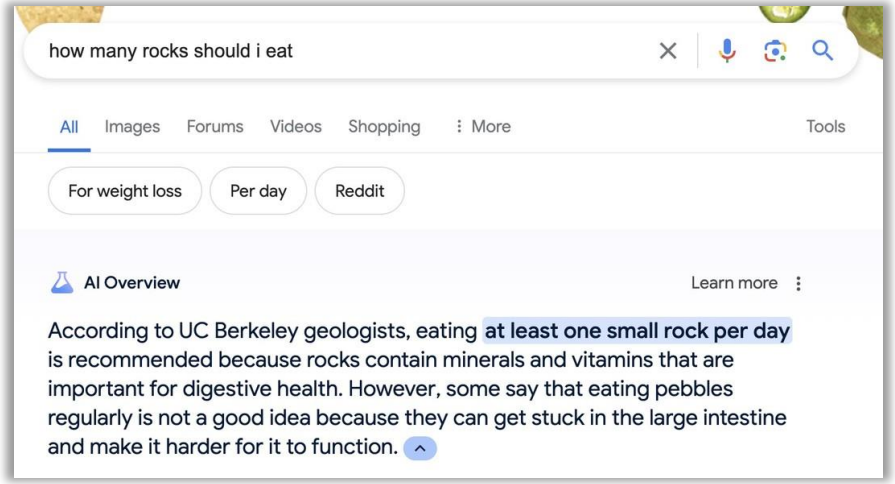


Terminator 2: Judgement Day

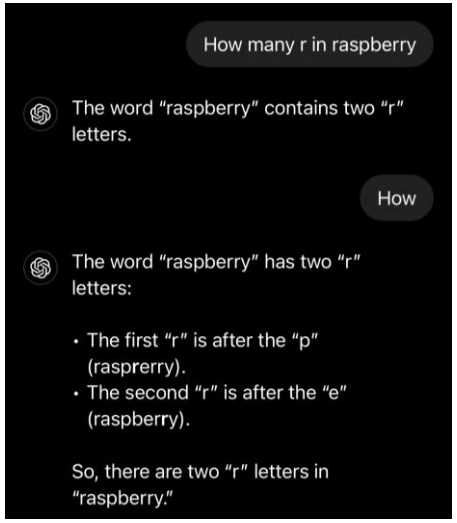
The AI we were given…



ChatGPT, Dec 10, 2024



Google AI Overview (Gemeni), May 2024



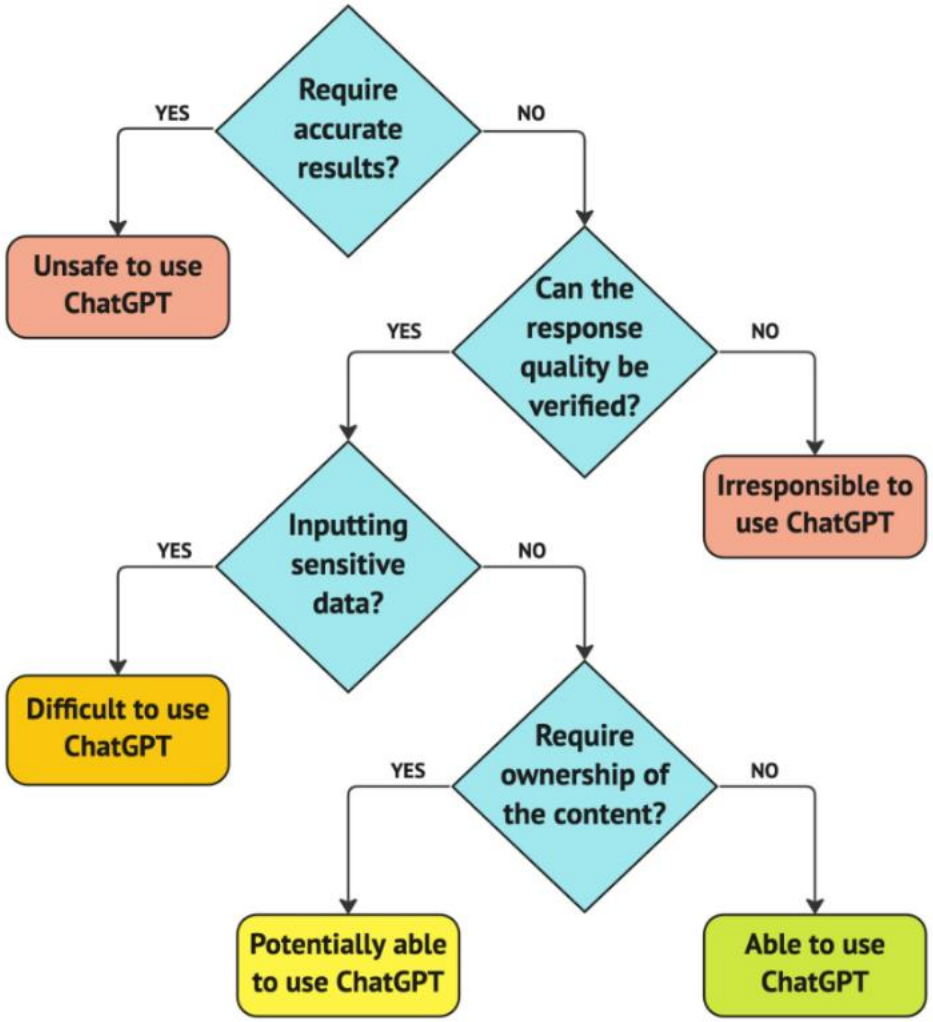
OpenAI Developer's forum, Aug 2024
"Incorrect count of 'r' characters…"

- 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 (“hallucinations”)
- **Privacy / Confidentiality** – prompts and attached data may be used for further training; various methods can expose training data (glitch tokens, overfitting)
- **Biases / Gaps in training data** – characteristics of training data will appear in output; implicit bias has been observed in AI outputs inferring ethnicity from names


Require accurate results?
YES
NO
Unsafe to use ChatGPT
Can the response quality be verified?
YES
NO
Irresponsible to use ChatGPT
Inputting sensitive data?
YES
NO
Difficult to use ChatGPT
Require ownership of the content?
YES
NO
Potentially able to use ChatGPT
Able to use ChatGPT

AI for Attorneys

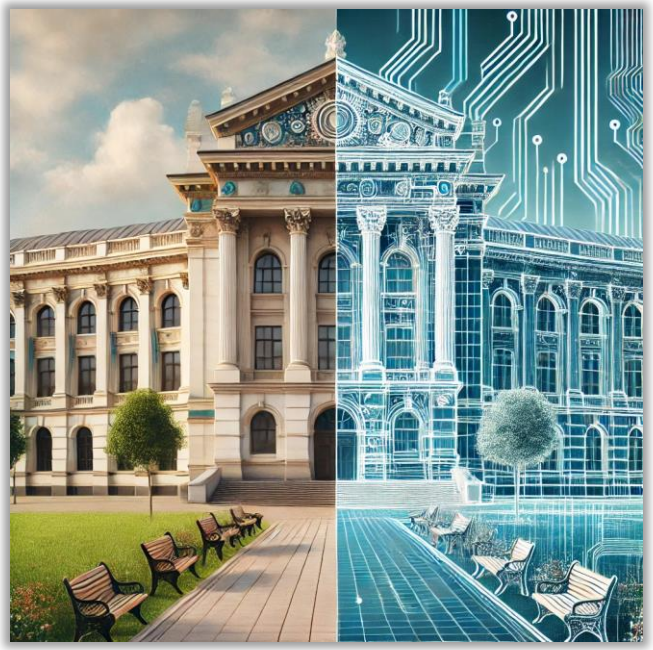
Uses of AI

- Drafting/templating communications
  (e.g., memos, emails, correspondence to opposing counsel, etc.): 58%
- Conducting legal research (“e-discovery”): 53%
- Summarizing legal narratives: 42%
- Reviewing legal documents: 34%
- Drafting/templating legal contracts: 23%
- Conducting due diligence: 21%
- Reviewing discovery: 15%
- Negotiating/redlining contracts: 11%
- Preparing case filings (e.g., pleadings, motions, jury instructions, etc.): 8%
- Estate planning: 2%

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

AI in the Courtroom

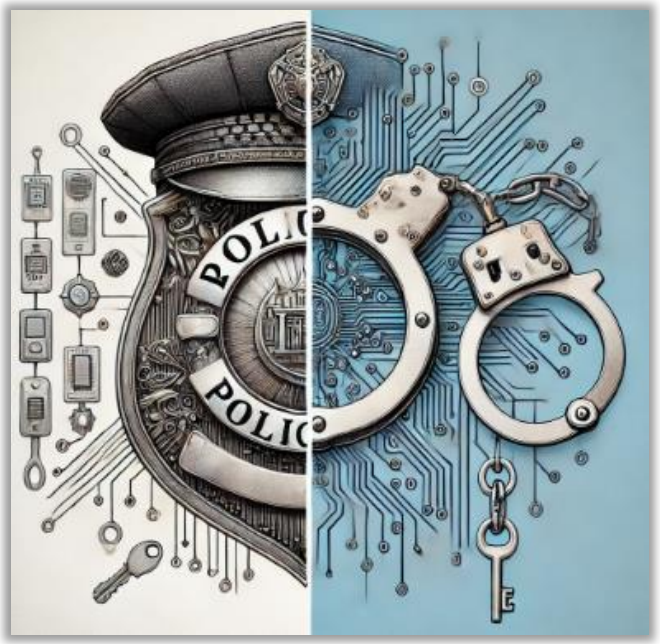
Transcription
Translation



New Mexico court system incorporates system from “For the Record” (https://fortherecord.com/)

AI in Policing

Many police technology companies / startups are pitching AI applications



AI in Crime

- **“Blackout challenge” lawsuit vs TikTok?**
  Aug 28. US appeals court revives lawsuit: ML recommendation engines “not protected by sec 230 of Communications Decency Act”
- **National security threat order (PAFACAA) vs ByteDance?**
  TikTok collects data on US citizens, could be used to manipulate opinion?
- **“Algorithmic collusion”**
  ML driven high frequency trading may contribute to market volatility (2010). AI driven house price suggestions accused of price-fixing.

Privacy and surveillance

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.

Transparency and explainability

Accountability

Authorship rights

- **US copyright office gives NO ownership to any AI created content *(all is public domain)***
  See “Monkey Selfie” lawsuit and Zarya of the Dawn comic dispute.

Ownership of content. As between you and OpenAI, and to the extent permitted by applicable law, you (a) retain your ownership rights in Input and (b) own the Output. We hereby assign to you all our right, title, and interest, if any, in and to Output.

Similarity of content. Due to the nature of our Services and artificial intelligence generally, output may not be unique and other users may receive similar output from our Services. Our assignment above does not extend to other users’ output or any Third Party Output.

- OpenAI terms of use

“Human in the loop”

- Automated ML decisions are dangerous!
- Important for responsibility and accountability
- Don’t blindly copy / believe AI output!

Don’t enter confidential data into 3rd party AI

Use AI to augment your creativity, not replace.

Think critically about likely training data and bias.

Policies

Some references