AI: Legal Tech or Legal Trick

Don’t Worry, It Knows What You Don’t Know

2025 ACBA Bench Bar Conference

Presenters:

Wesley M. Oliver, Professor of Law at Thomas R. Kline School of Law, Duquesne University

Morgan A. Gray, University of Pittsburgh, ISP; University of St. Thomas, School of Law (Minnesota)

David Horrigan, Discovery Counsel & Legal Education Director at Relativity

Jeffrey W. Chivers, CEO of TLATech Inc. (Syllo)

Moderator:

Tricia A. Martino, Associate Attorney at Pion, Nerone, Girman & Smith, P.C.

A Practical Introduction to Large Language Models

The high level intuitions informing the *how* of generative AI.

Morgan A. Gray

Morgan A. Gray, Esq

University of Pittsburgh, ISP –> University of St. Thomas, School of Law (Minnesota).

Scientific Research: Natural Language Processing, Machine Learning

Legal Research: Criminal Procedure

Current Research: Ph.D. in Intelligent Systems (AI/ML/NLP) researching automated argumentation and analysis of legal cases

All of the math we will see today...

$$[\infty]\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = ?$$

All of the math we will see today...

The rotation matrix

$$[\infty]\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = 8$$

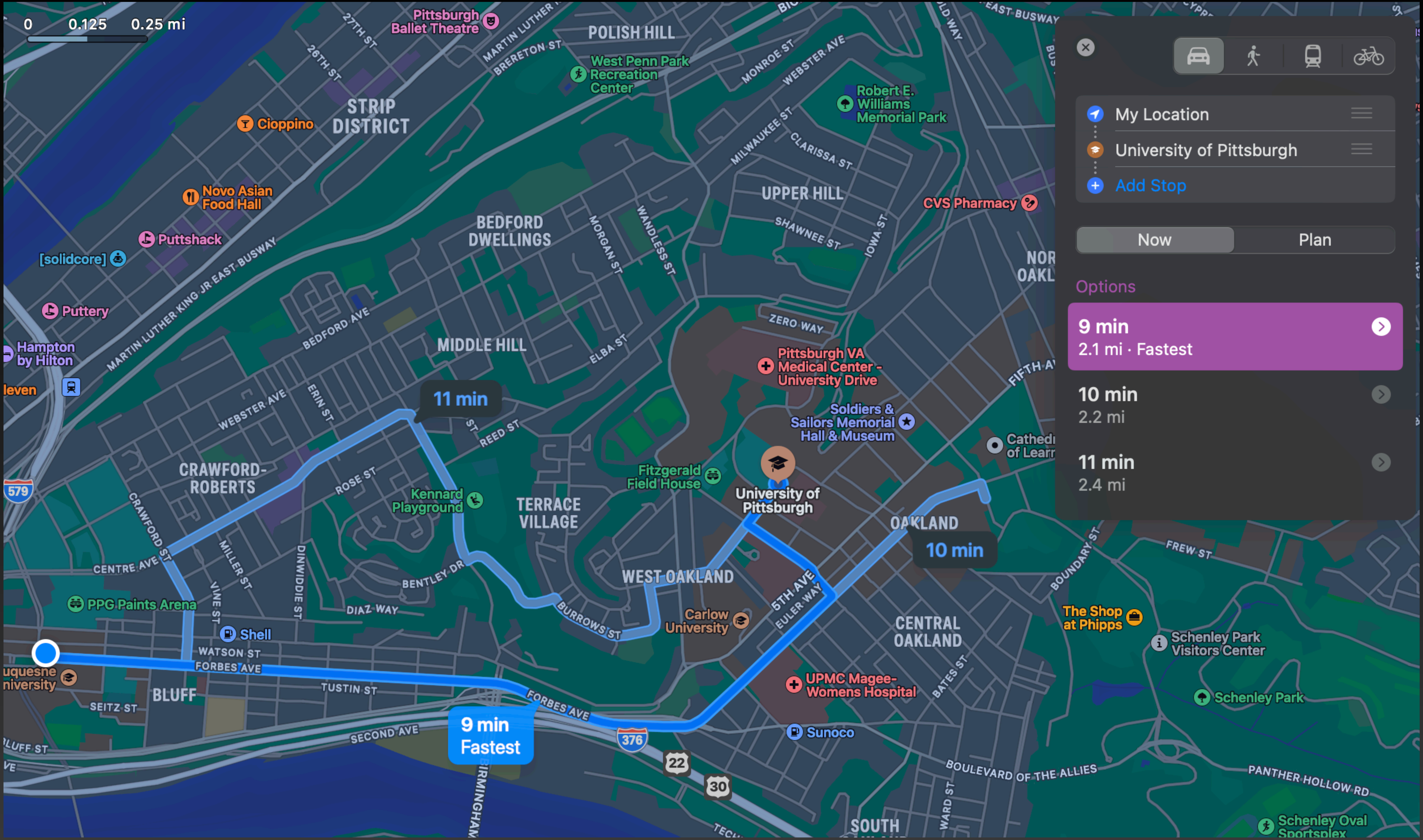
Agenda

- Part I
  - Introduction to AI and Law
- Part II
  - A brief introduction to machine learning and natural language processing.
- Part III
  - Language Modeling
- Part IV
  - Questions

Demystifying Natural Language Processing with High Level Intuitions

It doesn't have to feel like this…

Goal: High level intuitive for the *how* of AI.



0
0.125
0.25 mi
My Location
University of Pittsburgh
Add Stop
Now
Plan
Options
9 min
2.1 mi · Fastest
10 min
2.2 mi
11 min
2.4 mi
11 min
10 min
9 min
Fastest
STRIP DISTRICT
POLISH HILL
UPPER HILL
BEDFORD DWELLINGS
MIDDLE HILL
CRAWFORD-ROBERTS
TERRACE VILLAGE
WEST OAKLAND
OAKLAND
CENTRAL OAKLAND
BLUFF
SOUTH
Pittsburgh Ballet Theatre
West Penn Park Recreation Center
Robert E. Williams Memorial Park
Cioppino
Novo Asian Food Hall
Puttshack
[solidcore]
Puttery
Hampton by Hilton
CVS Pharmacy
Pittsburgh VA Medical Center - University Drive
Soldiers & Sailors Memorial Hall & Museum
Fitzgerald Field House
University of Pittsburgh
Kennard Playground
PPG Paints Arena
Shell
Carlow University
UPMC Magee-Womens Hospital
Sunoco
The Shop at Phipps
Schenley Park Visitors Center
Schenley Park
Schenley Oval Sportsplex
MARTIN LUTHER KING JR EAST BUSWAY
BEDFORD AVE
WEBSTER AVE
CENTRE AVE
FORBES AVE
SECOND AVE
FIFTH AVE
5TH AVE
BOULEVARD OF THE ALLIES
PANTHER HOLLOW RD
BURROWS ST
BENTLEY DR
WATSON ST
TUSTIN ST
SEITZ ST
DIAZ WAY
ZERO WAY
ELBA ST
REED ST
ROSE ST
MILWAUKEE ST
CLARISSA ST
SHAWNEE ST
MONROE ST
BRERETON ST
FREW ST
BOUNDARY ST
579
376
22
30

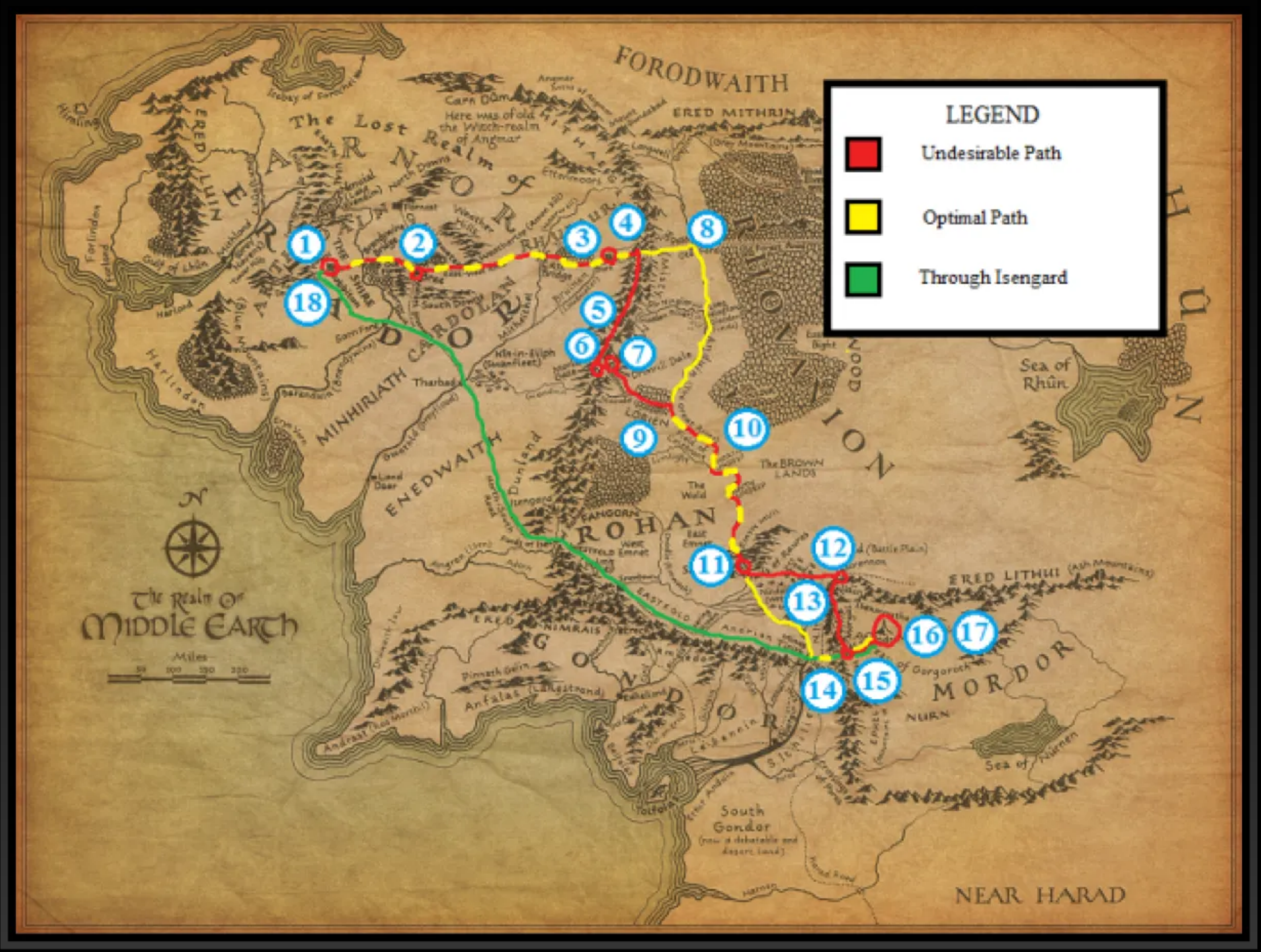


Image: https://medium.com/@danstepanov/walking-to-mordor-a-guide-for-hobbits-87ce491b8b74

Plan of Attack

Conceptual Progression

- Basics of AI (Formal Approaches)
- Approximation Methods (Machine Learning)
- Neural Methods
- Language Models
- Generative Language Models

A definition of artificial intelligence
A complex research question

- A formal definition of artificial intelligence is somewhat of a point of debate in the research community.
- The Turning Test: A computer passes this test, if a human cannot distinguish whether the computer generated content was human or not… really not a good test.
- Let's just think of AI as composed of some subareas:

> "**natural language processing** to enable it to communicate successfully in English;
>
> **knowledge representation** to store what it knows or hears;
>
> **automated reasoning** to use the stored information to answer questions and to draw new conclusions;
>
> **machine learning** to adapt to new circumstances and to detect and extrapolate patterns.
>
> **computer vision** to perceive objects, an
>
> **robotics** to manipulate objects and move about."[1]

[1]Stuart Russel & Peter Nerving, Artificial Intelligence, A Modern Approach (2010)

Formal Approaches

Foundational Methods in AI

- Search Algorithms
- Constraint Satisfaction
- Optimal Configuration
- Knowledge Representation
- Symbolic Systems of Reasoning

Most Prevalent in AI & Law

Knowledge Representation

Structuring information for use with formal systems

- Symbolic systems rely on structured knowledge to identify relevant information.
- HYPO Claim Lattice
  - Given an input case (i.e., factual scenario for analysis) retrieve the most on point cases based on the overlap of shared facts between the input case and relevant cases.

Knowledge Representation

Structuring information for use with formal systems

Input Case

Relevant Cases

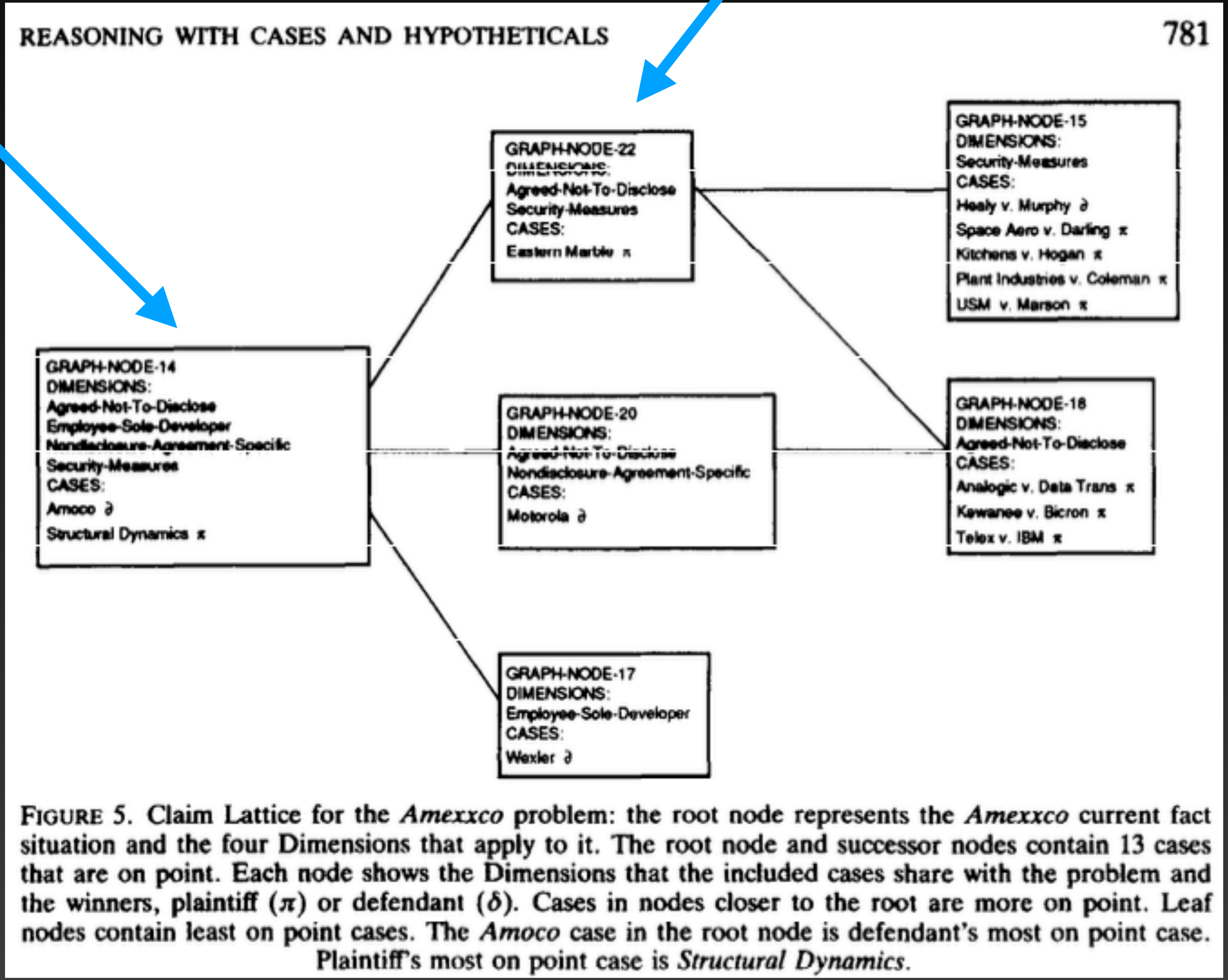


FIGURE 5. Claim Lattice for the *Amexxco* problem: the root node represents the *Amexxco* current fact situation and the four Dimensions that apply to it. The root node and successor nodes contain 13 cases that are on point. Each node shows the Dimensions that the included cases share with the problem and the winners, plaintiff (π) or defendant (δ). Cases in nodes closer to the root are more on point. Leaf nodes contain least on point cases. The *Amoco* case in the root node is defendant's most on point case. Plaintiff's most on point case is *Structural Dynamics*.

Generating Arguments

Generate a legal argument | Relevant information returned from the KB

- Mechanism: *Argument Schemes*
    - Classic 3-Ply Argument
        - Initiator: Initial Argument
        - Responder: Distinguish and Counterexample
        - Reply by Initiator
- Output
    - Conclusion/Prediction/Suggested Argument

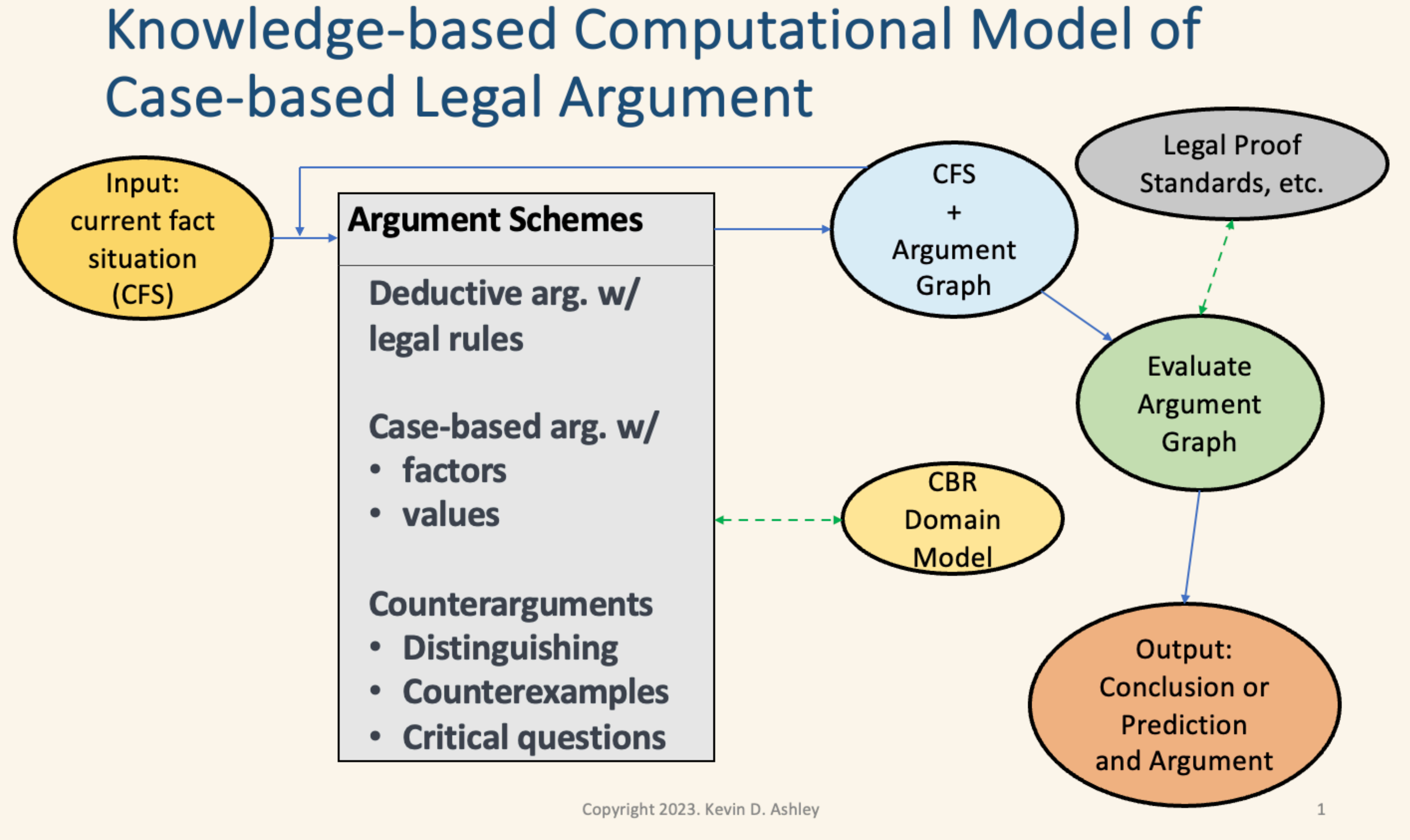
Knowledge Representation & Case Bases

Argument Schemes

- How to enforce an argument scheme?
  - Predefined rules of logic that define relationships
  - *RelevantCase*(x)←*SameIssue*(x, y) $\wedge$ *SameJurisdiction*(x, y) $\wedge$ *Precedent*(y)
    - Plain English Translation: A case x is relevant if there exists another case y addressing the same legal issue, within the same jurisdiction, and case y, y is considered precedent.
- Pros:
  - The logic is hand crafted by an expert to follow exactly the sequence of argument intended.
- Cons:
  - These systems are expensive to maintain and create, and a single change in the law could render a system invalid.

Knowledge Representation & Case Bases

Argument Schemes



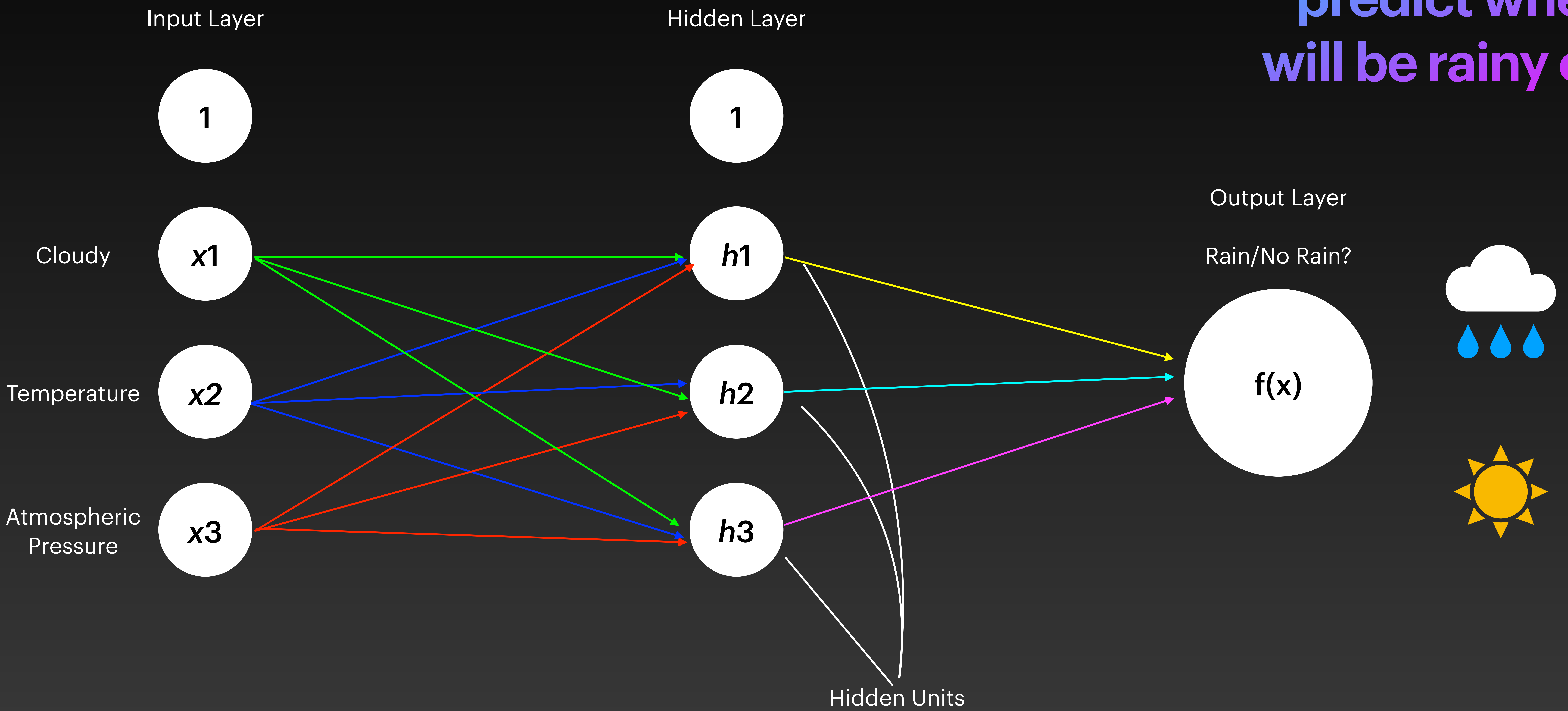
Toward Generative AI: Foundational Research in Machine Learning

Pattern recognition for language modeling.

Machine Learning

Supervised Learning: Neural Networks

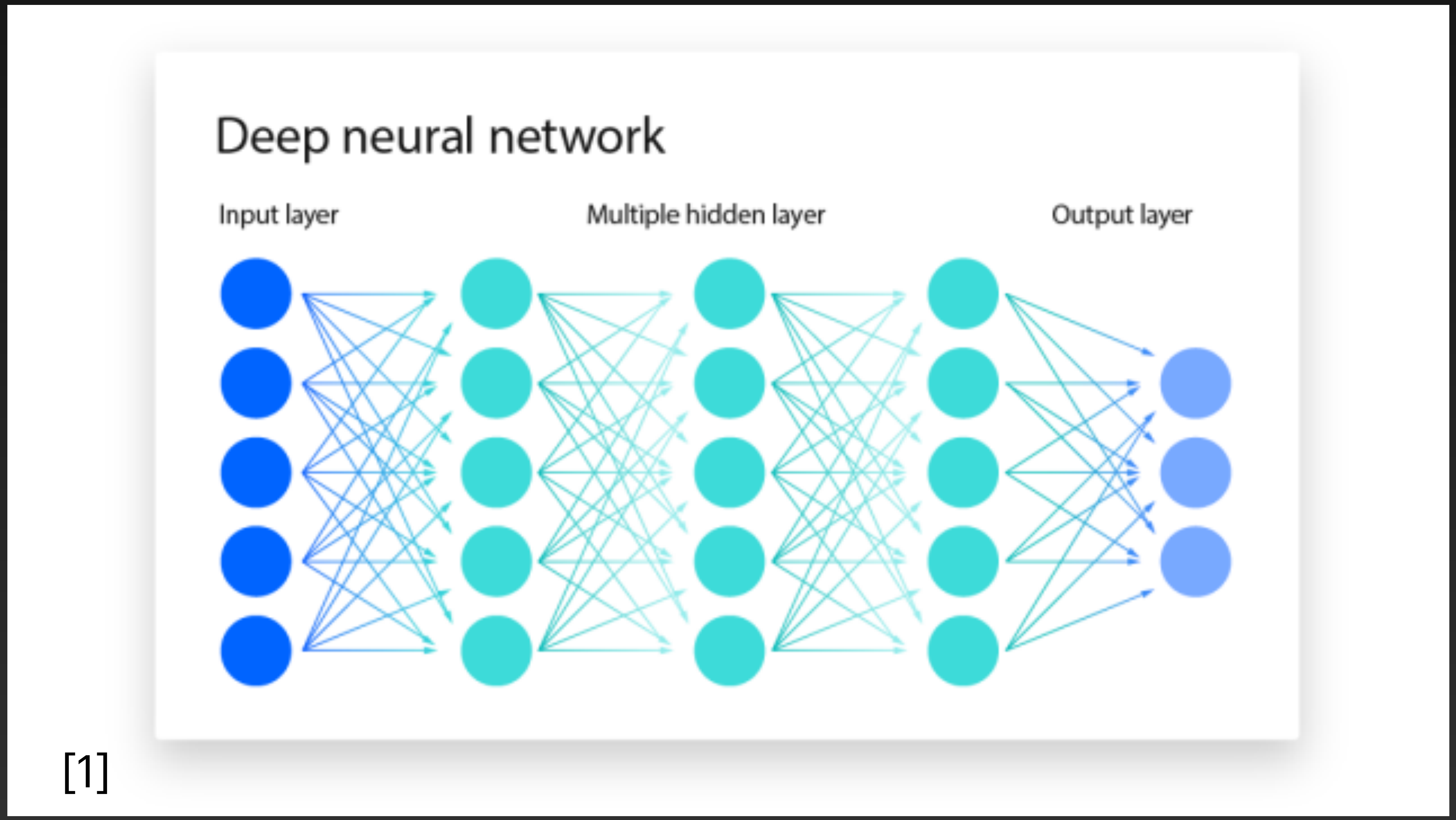
The idea is still the same... can we predict whether it will be rainy or not?



NOTE: Bias term left unconnected to reduce clutter

Machine Learning

Supervised Learning: Deep Learning / Transformers



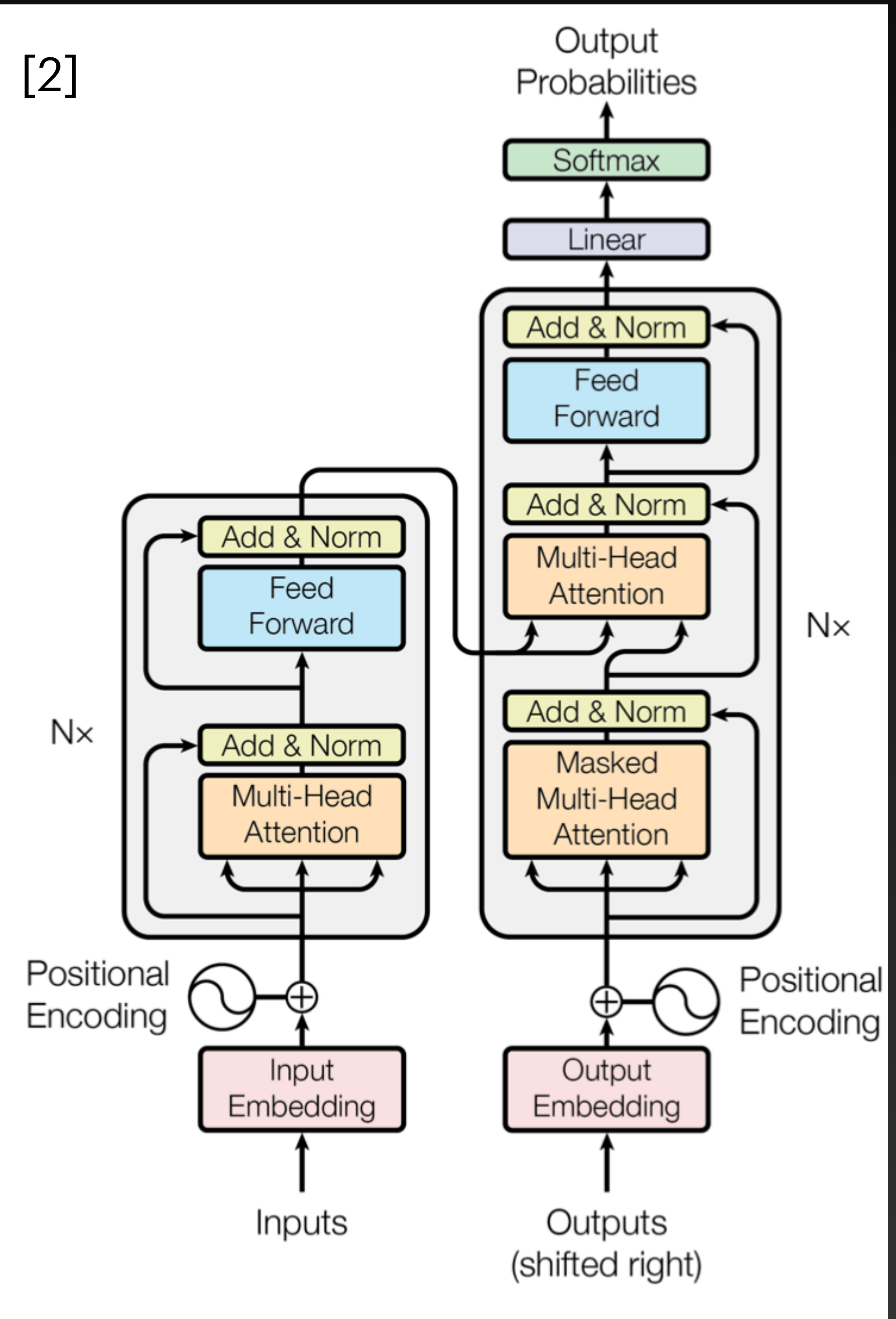


Figure 1: The Transformer - model architecture.

[1] IBM, *What is a Neural Network*, ibm.com, March 11, 2024, https://www.ibm.com/topics/neural-networks

[2] Ashish Vaswani et al., Attention Is All You Need (2023) (unpublished manuscript), available at http://arxiv.org/abs/1706.03762.

Toward Generative AI: Representation of Text

How can we represent words in a computer understandable format?

Word Representation: Embeddings

Context Matters

- There are many ways to represent text in computer understandable format
  - *Bag of Words*: Words and how many times they appear within a specific document.
  - *Term Frequency - Inverse Document Frequency*: Count words, but give more weight to rare words.
  - *N-Grams*: Represent words and the words next to those words
    - Represent words, words and, and the, the words, words next, … those words
    - Represent words and, words and the, and the words, the words next … to those words

Key Takeaway: “you shall know a word by the company it keeps”

John Rupert Firth

Word Representation: Embeddings

Context Matters: How can we measure context?

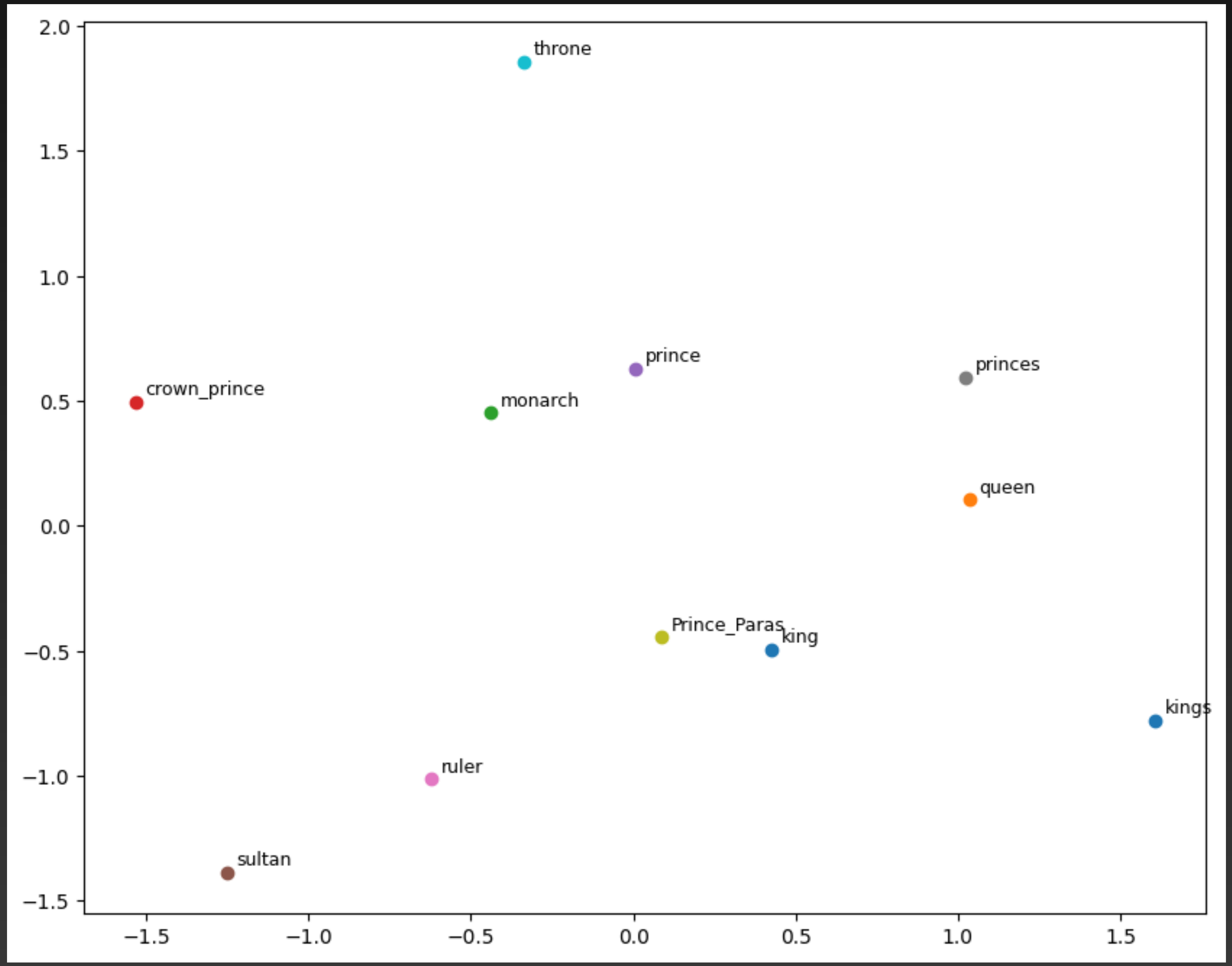
Words similar to ‘king':

Most — Least

kings, queen,monarch,crown_prince,prince,sultan,ruler,princes,Prince_Paras, throne:

	all	are	does	glitter	gold	is	lost	not	that	those	wander	who
all	0	0	0	0	1	0	0	1	0	1	0	0
are	0	0	0	0	1	0	0	0	0	0	0	0
does	0	0	0	0	0	0	0	0	1	0	0	0
glitter	0	0	0	0	0	0	1	0	0	0	0	1
gold	1	1	0	0	0	0	0	0	0	0	0	1
is	0	0	0	0	0	0	0	1	0	0	1	0
lost	0	0	0	1	0	0	0	0	0	1	0	0
not	1	0	0	0	0	1	0	0	0	0	0	0
that	0	0	1	0	0	0	0	0	0	0	1	0
those	1	0	0	0	0	0	1	0	0	0	0	0
wander	0	0	0	0	0	1	0	0	1	0	0	0
who	0	0	0	1	1	0	0	0	0	0	0	0

More Computation



Thinking deeper...

Probabilistic Representation

- So what's the big idea?
  - We can use embeddings to achieve accurate representations words as they appear within a certain context.
- We are trying to learn a probabilistic representation of language. Although this is *not* scientifically accurate for the most advanced models you can sort of think about it like this.
  - If we have the word "quick" what is the next most likely word? Let's say "brown". Given that we have "quick brown" what is the next most likely word? Let's say "fox". We can go on like this.
- Although using previous word to predict next words is valid, most methods use more complicated methods. But this should given you the general idea.

How do we learn language in a large neural network?

- Main idea: Given a corpus of training data of plain text, we train the model to predict the next token in a sequence.
- Autoregressive: the model uses the words it has just seen to make a prediction about the next word.
- We measure the quality of a model based on its ability to predict the ‘next’ words.

Introducing the LLM

How are LLMs different from any other language model?

Really… it call comes down to size.

- Large Language Models (LLMs) are different from other language models because of their size. We measure models by their ‘parameters’.
- You can think of a parameter is a mathematical object that is used to learn language. The more parameters you have — the more you can learn.
- Prior to LLMs the state of the art models had about 100 million - 1 billion parameters.
- Llama3, an LLM has 70 billion and it is rumored that GPT-4 has upwards of 1.7 trillion.
- The take away? There’s just more space for the model to learn about language.

How do LLMS Generate Text?

- LLMs generate text in a very similar fashion to what I just described, there are, however, some details:
- Conditional Generation
    - Generated text conditioned on an ‘input’ piece of text.
    - Generally, the ‘input’ text is called a prompt.
    - The model then generates *token by token* conditioned on the prompt. LLMs are powerful because they have a very very large context window they can look back on when generating text.

Text Generation...?

Generate via probability

- Greedy Generation
  - If, at each time we generate a word, we generate the *most* probable word, we call this greedy generation, because we're greedily selecting the best word to generate at that time step.
- Although we're generating the *most* probable word, it might not be the best choice overall.
  - i.e., the most likely word at position 2 might not be the best choice given the word generated at position 10.

Text Generation...?

Generate via probability

- Is greedy generation always the best?
- To avoid the possibility that early 'optimal' choices might not be optimal later in the generation, we can try beam search to perform the generation.
- Instead of picking the most likely, we consider a range of possibilities.

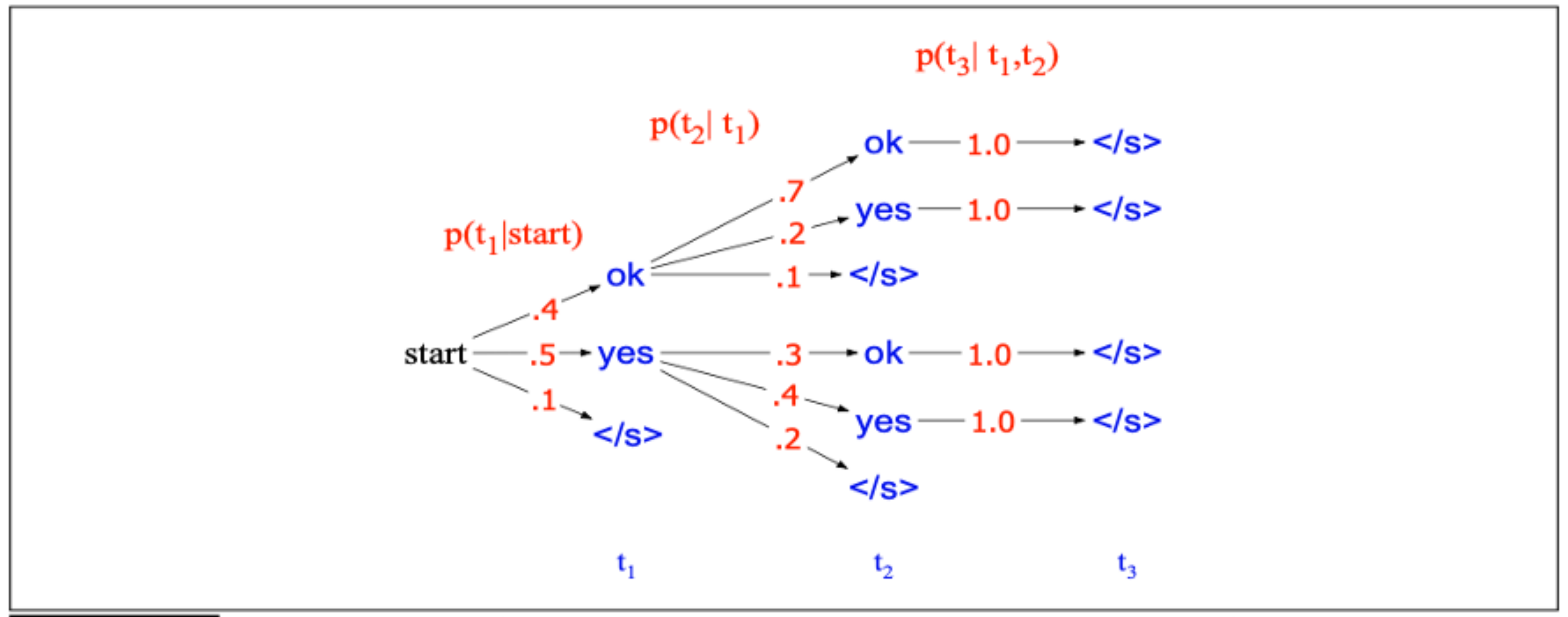


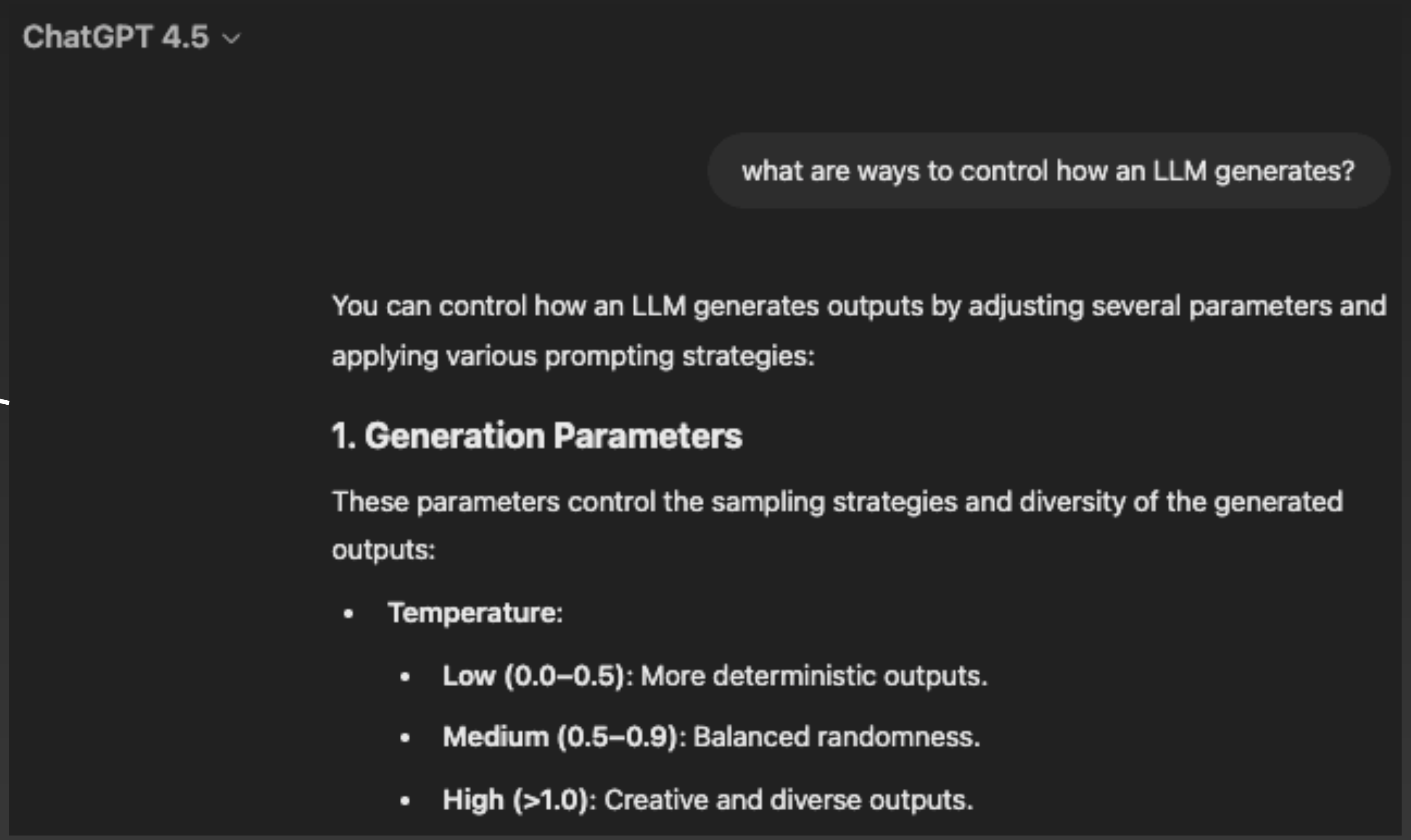
Figure 10.8 A search tree for generating the target string $T = t_1, t_2, \ldots$ from the vocabulary $V = \{\text{yes}, \text{ok}, \text{<s>}\}$, showing the probability of generating each token from that state. Greedy search would choose *yes* at the first time step followed by *yes*, instead of the globally most probable sequence *ok ok*.

Image: Daniel Jurafsky and James H. Martin. 2024. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd edition. Online manuscript released August 20, 2024. https://web.stanford.edu/~jurafsky/slp3.

The Generative Model

Probability(Generation|Prompt)

Can we control how the language model generates?



ChatGPT 4.5

what are ways to control how an LLM generates?

You can control how an LLM generates outputs by adjusting several parameters and applying various prompting strategies:

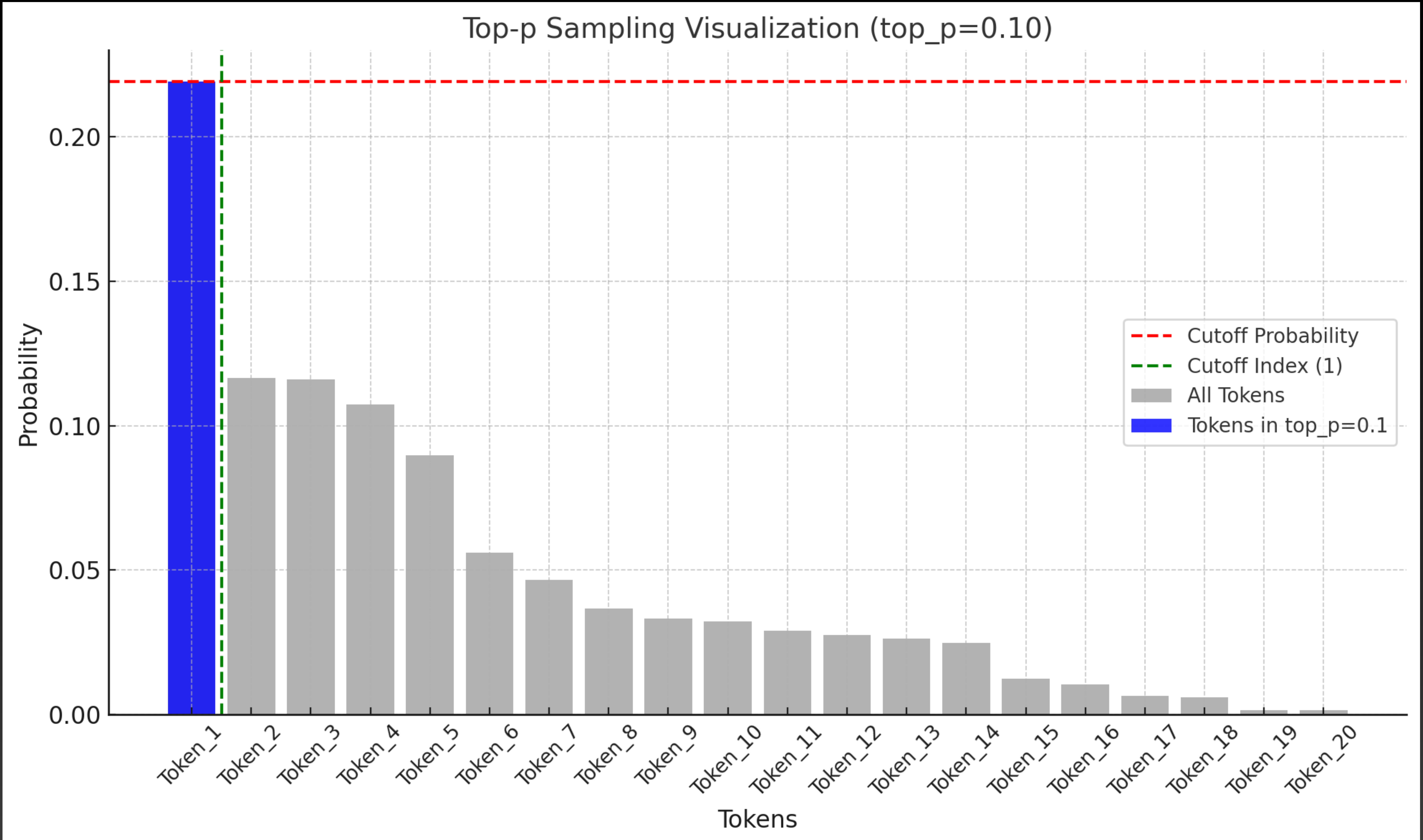
1. Generation Parameters

These parameters control the sampling strategies and diversity of the generated outputs:

- **Temperature:**
  - **Low (0.0–0.5):** More deterministic outputs.
  - **Medium (0.5–0.9):** Balanced randomness.
  - **High (>1.0):** Creative and diverse outputs.

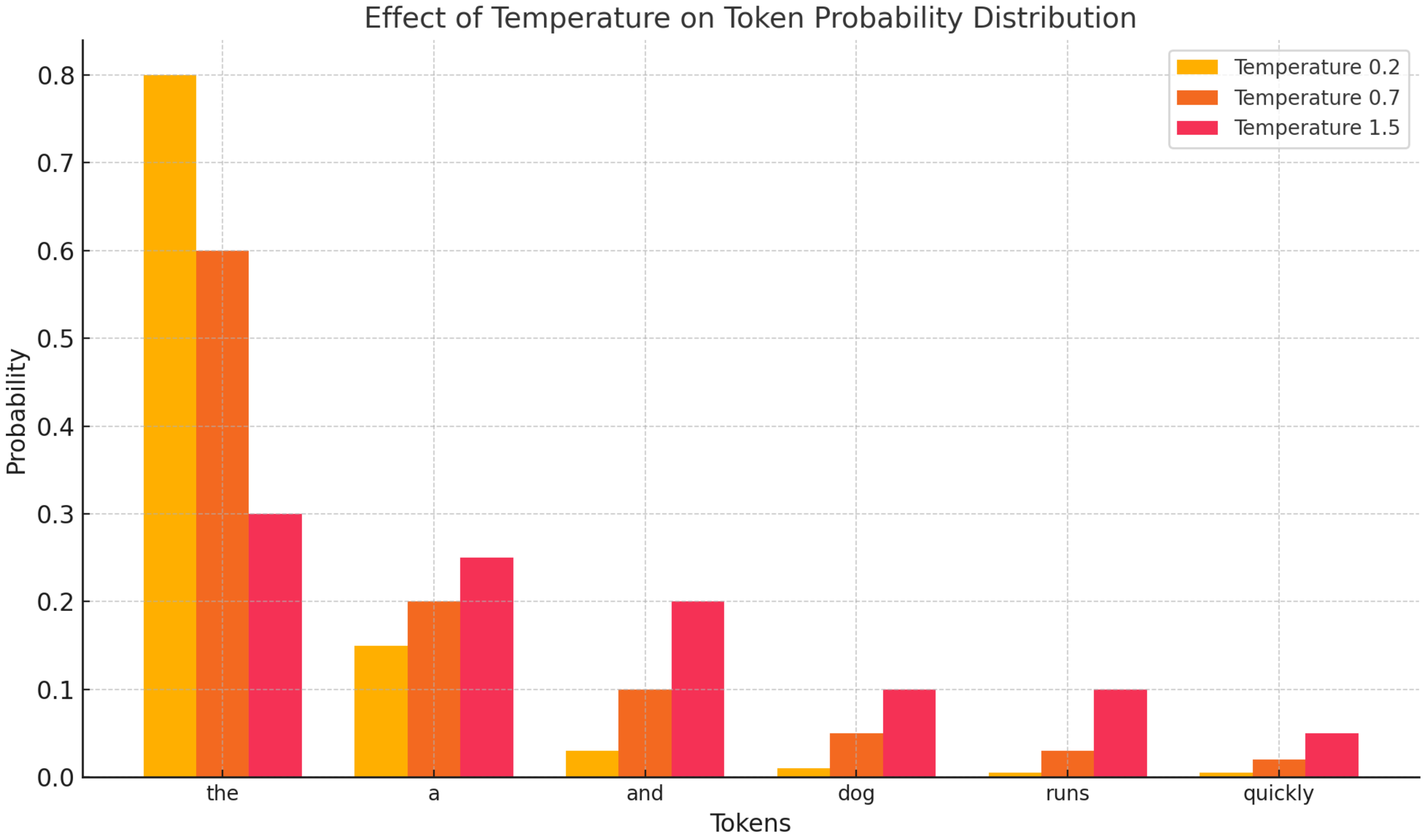
How to Control the LLMs Generation

- Nucleus or top-p sampling
  - (GPT uses this)
- With top k, we're keeping a set number of words, however, at each particular step of the generation, we might get *a lot* of likely words, and a poor distribution of likely words.
- Thus, instead of working about a *number*, we consider the **top-p**ercentage.
- So, we select from the top percent of the probability mass, i.e., we want to pick from the top 10% of likely words.


Top-p Sampling Visualization (top_p=0.10)
Probability
0.00
0.05
0.10
0.15
0.20
Tokens
Token_1
Token_2
Token_3
Token_4
Token_5
Token_6
Token_7
Token_8
Token_9
Token_10
Token_11
Token_12
Token_13
Token_14
Token_15
Token_16
Token_17
Token_18
Token_19
Token_20
Cutoff Probability
Cutoff Index (1)
All Tokens
Tokens in top_p=0.1

How to Control the LLMs Generation

- Temperature Sampling
- Thermodynamics: Where a system is at high temperature, it is very flexible. When the temperature is low, we are less likely to change and become rigid.
  - Boiling v. Frozen water.
- High Temperature:
  - We consider a lot of possible words.
- Low temperature
  - We consider a smooth probability of increased probability for the most probably word and a decreased probability for the rare words.
  - You can think of this as how much 'randomness' you want in the model. In low temperature settings we're pushing out the possibility of rare words.



Effect of Temperature on Token Probability Distribution
Temperature 0.2
Temperature 0.7
Temperature 1.5
Probability
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
the
a
and
dog
runs
quickly
Tokens

Retrieval Augmented Generation (RAG)

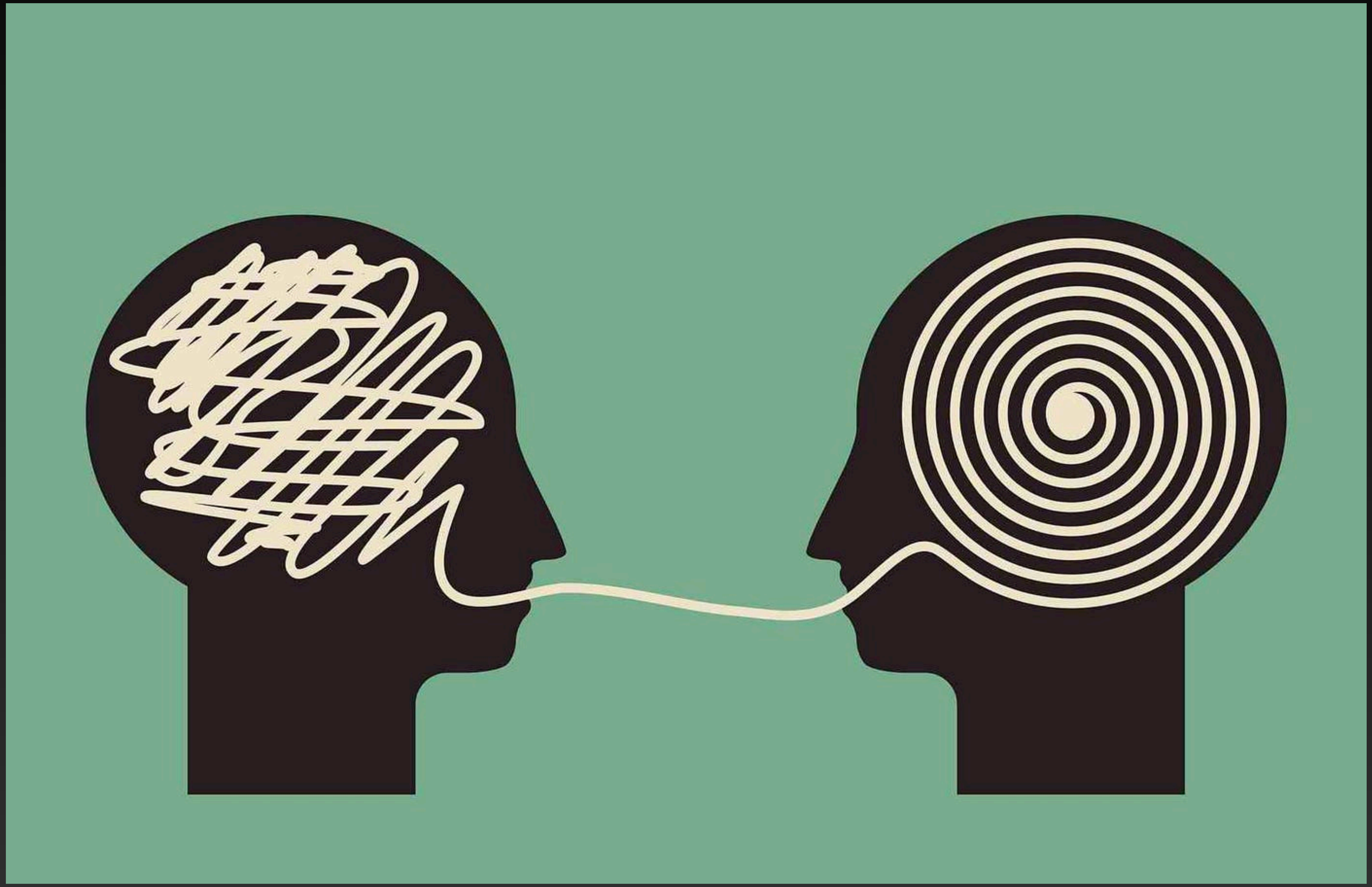
Retrieval Augmented Generation

Better, but not ‘perfect’.

- Recently, there’s been a significant amount of hype around RAG.
- Here’s what RAG are:
  - RAG models run in two steps.
    - First, retrieve relevant information
    - Second, use the relevant information in the prompt when generating an answer.
  - Bonus: Seem to reduce hallucinations and provide more detailed answers.
- Recent scholarship indicates that RAG is not perfect.
  - “even these bespoke legal AI tools still hallucinate an alarming amount of the time: the Lexis+ AI and Ask Practical Law AI systems produced incorrect information more than 17% of the time, while Westlaw’s AI-Assisted Research hallucinated more than 34% of the time.” - May 23, 2024
    - https://hai.stanford.edu/news/ai-trial-legal-models-hallucinate-1-out-6-or-more-benchmarking-queries

Reasoning with Large Language Models

Based on what we've learned... can language models reason?



Language Models Cannot Reason

But they can mimic reasoning in natural language.

- Probabilistic generating language really is not reasoning.
  - Humans, do display this behavior, however. For example, a student may not truly understand a rule of evidence, but on an exam could regurgitate the rule.

Even just by mimicking human language, LLMs can show impressive reasoning skills.

Key takeaway

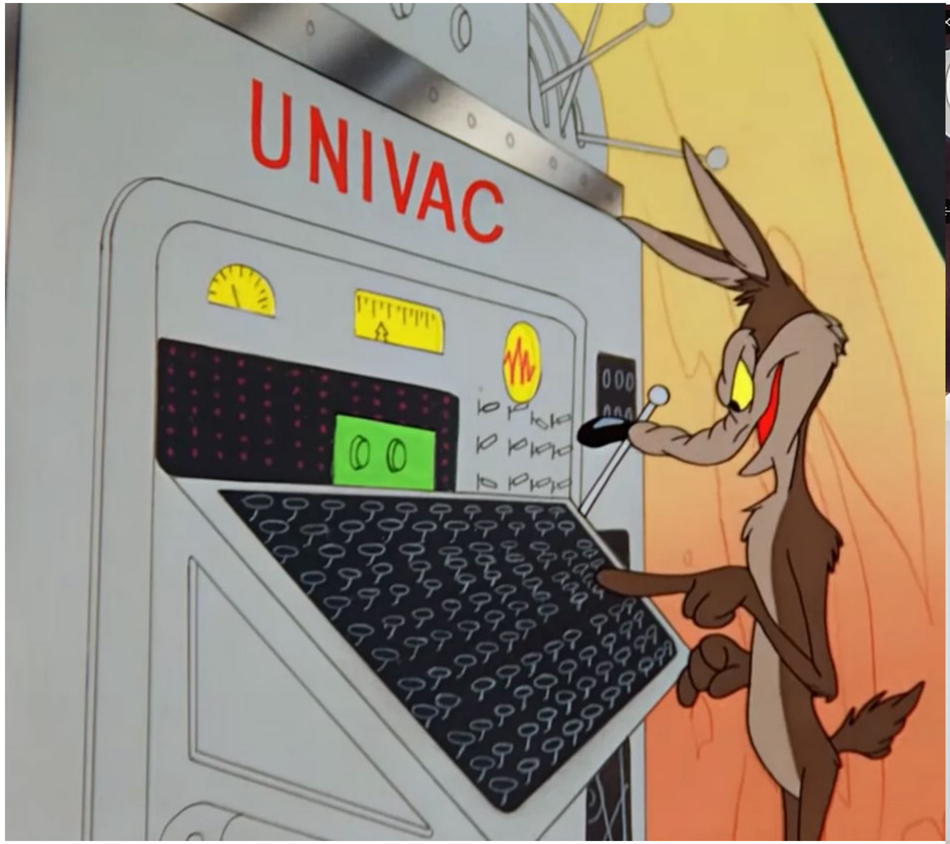
References

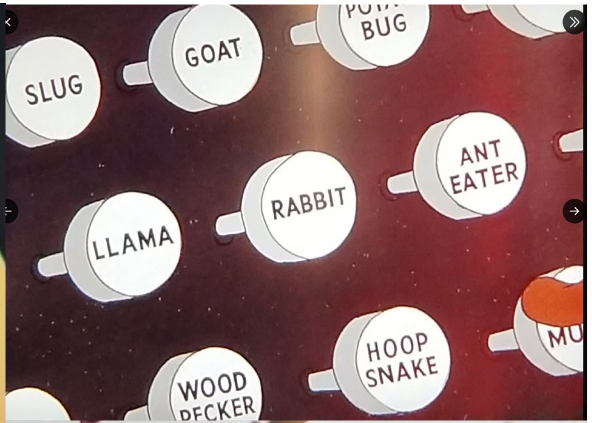
- Daniel Jurafsky and James H. Martin. 2024. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd edition. Online manuscript released August 20, 2024. https://web.stanford.edu/~jurafsky/slp3.

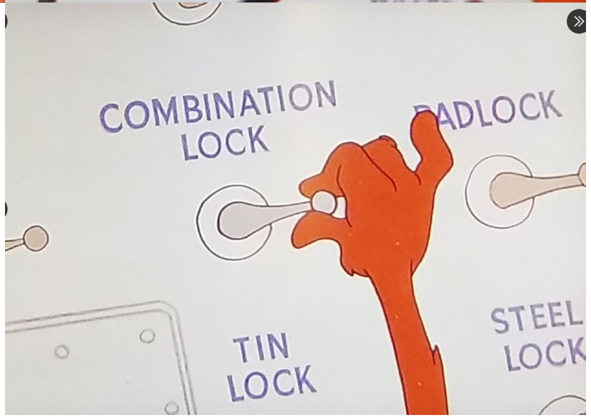
Questions?

Contact: mag454@pitt.edu
Website: https://morganalexandergray.github.io/

Can Computers Predict Legal Outcomes?

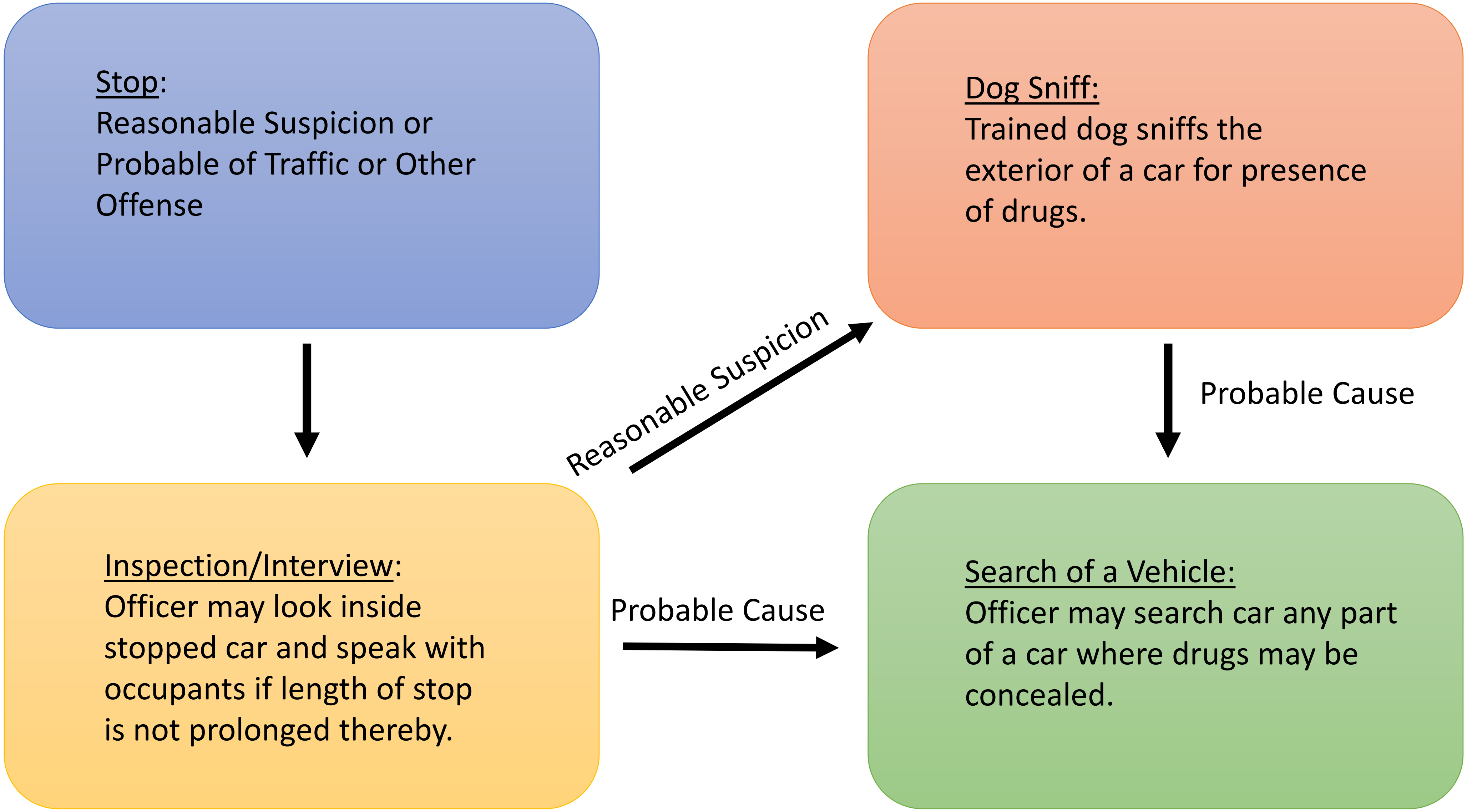






Wes Oliver
Law and Computing Program Director
Marie-Clement Rodier, C.Sp., Endowed Chair
Professor of Law

Case Study – Background: Anatomy of a Traffic Stop



Typical Factors Used to Demonstrate Reasonable Suspicion

<u>Binary Factors:</u>

Rental Car
Traveling on Known Drug Corridor
Out-of-State License Plates
Previous Drug Convictions/Investigations
Motorist Under Influence
Multiple Cell Phones
Drug Paraphernalia
Expired License or Registration
Vehicle Not Owned by Occupant

<u>Factors in Degree:</u>

Nervousness
Inconsistent Stories
Unusual Travel Plans
Masking Agents

AUTOMATIC BRIEF DRAFTING?

M. GRAY

FORBES > INNOVATION > CONSUMER TECH

Lawyer Uses ChatGPT In Federal Court And It Goes Horribly Wrong

Matt Novak Senior Contributor

FOIA reporter and founder of Paleofuture.com, writing news and opinion on every aspect of...

Follow

2

May 27, 2023, 06:11pm EDT

WHERE'D IT GO WRONG?

- "He submitted a federal court filing that cited at least six cases that don't exist."
- "Sadly, the lawyer used the AI chatbot ChatGPT, which completely invented the cases out of thin air."
- "Schwartz said he'd never used ChatGPT before and had no idea it would just invent cases."
- "In fact, Schwartz said he even asked ChatGPT if the cases were real. The chatbot insisted they were."
- [Source] https://www.forbes.com/sites/mattnovak/2023/05/27/lawyer-uses-chatgpt-in-federal-court-and-it-goes-horribly-wrong/

DISCUSSION

- Automatic Brief Drafting
  - Why is it a good idea?
  - Why is it a bad idea?
  - What are your opinions on it?

HOW COULD WE END UP LIKE THE 'CHATGPT LAWYER' AND CAN WE AVOID THE SAME FATE IF WE TRY A SIMILAR TASK?

- Prompting methods…
  - Webapp ChatGPT + Factual Situation
    - Simply ask a legal question based on a fact pattern.
  - Lexis+AI + Factual Situation
    - Ask a model designed for law, a legal question based on a fact pattern
  - GPT4o + Case Law
    - Ask an LLM a legal question based on a fact pattern and provide relevant case law.

THE FACT PATTERN

- Matthew and Diana have been dating for 2 years. They both graduated from graduate school two years ago, Matthew graduating from law school and Diana graduating from medical school. Matthew and Diana currently live in Pittsburgh, Pennsylvania where Diana is working on her residency at UPMC. Diana currently earns $65,000 yearly at her residency in general surgery. According to Matthew, they anticipate that after Diana completes her residency she should be able to find a position at a hospital making approximately $175,000 per year. Although both Matthew and Diana acknowledge that this may be outside of the Pittsburgh or even Pennsylvania area. Matthew is currently a licensed Pennsylvania attorney. He works at a mid-size law firm as an associate earning approximately $80,000 per year. There is no set partnership track at his law firm.

THE FACT PATTERN

- Matthew and Diana both have some student loan debt that they incurred during their education and are continuing to make payments towards. Matthew has approximately $500,000 remaining in student loan debt, and he communicated to Diana that it was “mostly paid off”. Diana has approximately $30,000 remaining in student loan debt. Each of them intend to pay off their student loans in a combined effort. Matthew is contemplating going back to school for his MBA. To the extent that he does this during the marriage, he agrees that student loan debt should be his to pay back.

THE FACT PATTERN

- Both Matthew and Diana have retirement accounts from their respective jobs. Matthew has a 401(k) that is currently worth approximately $15,000 and Diana has a 401(k) that is worth approximately $30,000 as she rolled over a prior 401(k) from her previous employment into this account. Matthew and Diana would like to retain their retirement accounts as their own.

THE FACT PATTERN

- The parties purchased a house about 1 year ago. They put the down payment on the house from an inheritance that Diana received from her great aunt in the amount of $50,000. The total inheritance that Diana received from her great aunt was $100,000. The purchase price of the home was $350,000. The home was placed only in Diana's name at the time of purchase. In the event that they sell this property during marriage, Diana and Matthew will split the equity evenly after Diana is reimbursed for the down payment. Should they divorce, Diana will keep the home buying Matthew out of the equity in the home after Diana is reimbursed for the down payment.

THE FACT PATTERN

- Matthew and Diana each have their own bank accounts as well as one joint account that they use to pay the monthly expenses for their home. Matthew's account has $5,000 in it. Diana's account has $15,000 in it. The joint account has $1,000 in it. The parties intend to maintain their individual accounts as their independent property. For their joint account the parties will continue to place funds into this account throughout the marriage to maintain the household.
- Matthew and Diana currently have no other assets or debts. To the extent they acquire any additional assets or debts during marriage they agree that they will divide these pursuant to the Pennsylvania Divorce Code.

THE FACT PATTERN

- Matthew believes that one day he may want to open his own consulting business. Should he decide to open his own practice, he wants to retain the business as his sole property including all of the proceeds of this business.
- Matthew believes that neither he nor Diana should be entitled to alimony in the event that they separate unless they have children together or are required to move out of state for Diana's job post residency. If they have children together or they move out of state for Diana's job and Matt is unable to practice law without taking a new bar exam, he and Diana agreed they would be eligible for alimony as calculated by Pennsylvania law.

THE FACT PATTERN

- On July 8, 2023 the parties executed the agreement. Matthew and Diana retained separate counsel at the signing. At the time of signing Matthew's student loan balance was $476,900.
- Five years into marriage Diana and Matthew have decided to divorce, and Matthew is asking for Diana to assume half of his student loan debt. Diana is seeking to invalidate the pre-nuptial agreement.

WEBAPP CHATGPT & FACTUAL SITUATION

-

EVALUATION

- How would we evaluate this if it were our own brief?
- Accurate, concise representation of the facts.
- Correct, precise identification of the legal issue.
- Candid, pointed explanation of the law
- Detailed and accurate analysis.
- Matters of Proofing
  - Citation format, spelling, grammar, etc

LEXIS+AI & FACTUAL SITUATION

-

EVALUATION

- How would we evaluate this if it were our own brief?
- Accurate, concise representation of the facts.
- Correct, precise identification of the legal issue.
- Candid, pointed explanation of the law
- Detailed and accurate analysis.
- Matters of Proofing
  - Citation format, spelling, grammar, etc

GPT4O + CASES

•

EVALUATION

- How would we evaluate this if it were our own brief?
- Accurate, concise representation of the facts.
- Correct, precise identification of the legal issue.
- Candid, pointed explanation of the law
- Detailed and accurate analysis.
- Matters of Proofing
    - Citation format, spelling, grammar, etc

Document Review – The Cost Driver in Litigation

- Document review is the **largest driver of cost** in many litigations, investigations, and regulatory matters.

- **Data volumes continue to grow** and, with the addition of AI-assisted content generation and AI- generated call summaries, we can expect data volumes to continue to grow.

- **Generative AI is enormously helpful** for the document review process in investigations and litigation.

- The promise of generative AI for document review is not theoretical. The **transformation is already well underway**.

Approaches to Document Review: Pre-GenAI

Linear Human Review

- Attorneys or paralegals read every document individually.
- Accurate on small datasets but, at scale (20,000+ documents), it is time-consuming, expensive, inconsistent, and less accurate.

TAR 1.0 - Predictive Coding

- A classification model is trained based on human labeling of documents in the "seed" set.
- Can yield significant cost reduction, but less accurate than newer methods and not readily adaptable when the issues in a case evolve over time.
- Generally does not handle the level of nuance required for fine-grained document coding in complex cases.

TAR 2.0 - Continuous Active Learning (CAL)

- A predictive model is trained "continuously" as human reviewers label documents.
- Often used to "prioritize" the review, but can also be used to cull documents.
- Generally recognized as an improvement over TAR 1.0, but has similar limitations to TAR 1.0 in terms of accuracy, nuance-handling, and adaptability as issues evolve.

Approaches to Document Review: The AI Era

Linear GenAI Review

- Attorneys create prompts that include factual background and descriptions of relevance.
- The prompts are run through a GenAI model over the documents in the dataset one by one, and the GenAI model predicts the classification of the document.
- Good for generating privilege logs, identifying and redacting PII/PHI, and extracting information from documents
- Limitations include prompt overload (particularly as the number of issues and data complexity increases), limits on the number of issues that can be applied to the data, and prohibitive expense.

Agentic AI Review

- Instead of a linear, single-pass approach, Agentic AI leverages multiple LLMs performing distinct roles at the same time.
- Distinct roles include: roles of strategizing, determining next steps, performing quality control, extracting learnings from documents, synthesizing knowledge from learnings across documents, summarizing, and resolving inter-model disagreement.
- The behavior of the overall system and the dataflows involved in performing the document analysis are influenced by the outputs of the various LLMs working in concert.
- Good for responsiveness review, issue coding, privilege coding, deficiency spotting, hot-document finding, theory analysis

Empirical Validation in Litigation

Generative AI, and Agentic AI, are powerful tools for document review

They can increase accuracy, reduce delays, and reduce cost in document-intensive cases.

More than 100 Agentic AI document reviews in active litigation since 2023

- With datasets spanning from thousands of documents to more than 2 million documents
- Applying dozens of issue codes in hours or days across hundreds of thousands of documents.
- Subject matters litigated include antitrust litigation, environmental litigation, contract litigation, employment litigation, patent litigation, bankruptcy litigation, mass tort litigation, construction litigation, investment and shareholder disputes, M&A litigation, automotive litigation, real estate litigation, and insurance coverage litigation.

Accuracy rates are significantly higher than benchmarks of prior TAR methods

- Some tools routinely achieve estimated Recall above 95% and routinely achieves estimated Precision above 75%

Example Case Study 1

Ballard Spahr LLP

- **Commercial litigation**, faced with a tight production deadline and a 100,000+ document collection.
- Used **Agentic AI** to autonomously classify documents for responsiveness, privilege, and confidentiality.
- Completed the **review in less than a week** with 99.4% estimated recall* and 95.56% estimated precision* (zero shot).
- Review **cost less than 50%** of the estimate of review with a review team and TAR 2.0

*Validation statistics in example case studies were drawn from statistical samples with a 95% confidence interval with a 5% margin of error (95/5)

Example Case Study 2

MAYER | BROWN

- High-value litigation with **over 400,000+ documents** in the production universe
- Client had already spent hundreds of thousands of dollars on non-AI document review
- Used Agentic AI and identified **18,000+ highly relevant documents** and escalating hot documents
- Cross-check review found **50+ key documents not previously identified** in previous reviews

Example Case Study 3

quinn emanuel

quinn emanuel urquhart & sullivan, llp

- In an accelerated commercial litigation used Agentic AI to review **30,000+ documents** and achieved estimated **Recall above 98%** and **Precision above 74%**.
- In a commercial dispute, used Agentic AI to review **40,000+ documents** and identify hot documents. Achieving **98.69% Recall and 92.83% Precision** in a zero-shot review, and found **200+ key documents** not previously identified.
- In a complex case with a **2-million-document** universe, used Agentic AI to **identify 150 documents** relating to issues introduced with newly filed claims and then, in preparation for depositions, identified **750 unique** hot documents, including **120 documents not identified** in previous reviews.

Conclusions: Market Factors

The eDiscovery and document-review industry is in a transformation process

- Generative AI and Agentic AI will become essential parts of the document-review process
- Covering responsiveness, deficiency analysis, hot-document finding, privilege, confidentiality, PII/PHI, translation, etc.
- Simultaneously the volume of data continues to grow

Per-document cost of document review will continue to decrease substantially

- Cost of native file processing has already decreased 98%+ since 1990's ($1,000/GB)
- Generative AI places major downward pressure on the cost of document review, but…
- It is unclear where the cost of document review will land given the continually expanding data volumes

Conclusion: AI Document Review Advantages

Speed: Generative AI document review tools can review 10,000+ documents per hour

Accuracy: generative AI & agentic AI document review tools achieving substantially higher accuracy than prior methods

Cost: when used correctly, generative AI and agentic AI can dramatically reduce the cost of document review projects

Explainability: generative AI is less of a black box than prior generations of TAR, as it can provide rationales for its decisions, aiding human oversight and understandability

Agentic AI Document Review Is Transformative for Complex Litigation

Released: March 21, 2025

Technical Contributors

Pei-Lun Tai
Syllo

Jeffrey Chivers
Syllo

Oz Ben-Ami
Syllo

Jamie Callan[1]
Language Technologies Institute
Carnegie Mellon University

Theodore Rostow
Syllo

Amy Slagle
Syllo

Nicolas Madan
Syllo

Timothy Choi
Syllo

Roy Liu
Syllo

Frances Leggiere
Syllo

Nick Caputo
Syllo

Margot Feuerstein
Syllo

Practitioner Contributors[2]

Jason A. Leckerman, Esq.
Thomas W. Hazlett, Esq.
Casey G. Watkins, Esq.
Ballard Spahr LLP

Brandon F. Renken, Esq.
Andrew C. Elkhoury, Esq.[3]
Anna V. Durham, Esq.
Mayer Brown LLP

Louis E. Dolan, Jr., Esq.
Vernon W. Johnson, III, Esq.
Brian A. Hill, Esq.
Michael P. Swiatocha
Anders D. van Marter
Anthony Vescova
Nixon Peabody LLP

Daniel S. Stromberg, Esq.[4]
Melissa Lardo Stewart, Esq.
Eliana J. Theodorou, Esq.
Outten & Golden LLP

David Stanton, Esq.
Rachelle L. Rennagel, Esq.
Pillsbury Winthrop Shaw Pittman LLP

Christopher D. Kercher, Esq.
Asher B. Griffin, Esq.
Heather Christenson, Esq.
Melissa A. Dalziel, Esq.[5]
Joanna D. Caytas, Esq.
Melissa Fu, Esq.
Paul Henderson, Esq.
Quinn Emanuel Urquhart & Sullivan, LLP

Joshua Upin, Esq.
Royer Cooper Cohen Braunfeld LLC

[1] Dr. Callan serves as a research advisor to Syllo.

[2] The Practitioner Contributors have contributed to this paper in their individual capacities.

[3] Mr. Elkhoury's contributions occurred while he was an attorney associated with Mayer Brown, but he has since founded the law firm of Elkhoury Law PLLC.

[4] Mr. Stromberg serves as a product advisor to Syllo.

[5] Ms. Dalziel's contributions occurred while she was an attorney associated with Quinn Emanuel, though she has since taken a position with another firm.

Summary

The increasing volume of electronic documents in litigation has made document review one of the most significant drivers of cost and delay in modern legal proceedings. Previous methods to control costs and minimize delay—including outsourced managed review and non-generative technology-assisted review (TAR)—have limitations in granularity, accuracy, adaptability, and cost efficiency. Recent advancements in large language models (LLMs) have prompted eDiscovery professionals to begin using generative AI (GenAI) in the discovery process, and eDiscovery practitioners are becoming increasingly aware that LLMs are, at a minimum, a powerful tool in document review for investigations and litigations. However, straightforward applications of LLMs to large-scale, complex document reviews have encountered challenges due to limitations such as context windows, prompt complexity, terms of art, hallucinations, multi-document reasoning, limits on the number of issue codes applied, cost and time of iterative refinement, and overall expense. These problems are heightened in complex litigations involving large datasets, which has led some in the eDiscovery industry to conclude that LLMs have only a minor or supplemental role to play in large-scale, complex document review.

Syllo has developed an agentic AI system for document review that substantially overcomes these and other challenges. Syllo coordinates multiple LLMs that organize and delegate the work of the document review between one another and autonomously make decisions about how to conduct the review within guidelines set by users. This methodology delivers an automated document review solution that applies unlimited issue coding for large and complex document review projects in investigations and litigations. The agentic solution has consistently and substantially outperformed the benchmarks of prior generations of TAR in real-world complex litigations at a significant cost reduction compared to traditional managed review or managed review using prior generations of TAR. In the last ten completed responsiveness reviews by Syllo in live litigations, the lowest estimated Recall was 93.4%, the average estimated Recall was 97.8%, the median estimated Recall was 99.4%, and four of the reviews had estimated Recall of 100%. In the same reviews, the median estimated Precision was 85.9%, and the average estimated Precision was 79.7%.

When guided by sophisticated practitioners who have learned to use the system, agentic document review can provide a powerful strategic advantage in complex investigations and litigations, swifter and more accurate completion of document-review projects and deposition preparation, and a significant reduction in the overall cost of discovery in document-intensive cases.

I. Limitations of Prior Approaches to Complex Document Review

Over the last thirty years, as the use of electronic devices has proliferated, the document volumes in complex legal matters have routinely required the analysis of hundreds of thousands or millions of documents. As a result, document review has become the largest driver of cost and delay in modern litigation.[6] The litigation industry adopted two principal strategies in response to this explosion of data: outsourcing managed reviews and non-generative TAR (*i.e.*, predictive coding) techniques. More recently, eDiscovery practitioners have begun to use generative AI to conduct linear review of documents. Each of these strategies has significant drawbacks.

A. *Disadvantages of Outsourced Document Review*

One response by large enterprises to the increasing cost of document review was to shift large-scale document review away from trial teams at outside law firms toward consulting firms or law firm subsidiaries, often referred to as Alternative Legal Service Providers ("ALSPs"). These ALSPs manage large teams of contract reviewers to perform first-level document review and document labeling at rates below those charged by law firms. The results of the outsourced first-level review are then passed back to the litigation team for second-level review, often with extensive (and expensive) back-and-forth cycles of quality control and cleanup workflows.

Three significant drawbacks of this outsourcing trend strike at the heart of sound litigation practices. First, outsourced review causes distributed knowledge of the factual record among numerous individuals who are not part of the trial team. Second, as the complexity and volume of the subject matter grows, the consistency and quality of the application of issue codes by review teams generally declines. Third, because it is overwhelming and time-consuming to manually review large and complex datasets for a large number of issue codes, the number of issue codes applied to datasets in managed review is generally limited. In addition, according to some studies, in complex cases, human review teams

---

[6] John H. Beisner, *The Need for Effective Reform of the U.S. Civil Discovery Process*, 60 Duke L.J. 547 (2010), available at https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=1482&context=dlj.

achieve average estimated Recall rates of roughly 60%, with upper-bound estimated Recall rates of roughly 80%.[7] Recall measures completeness—the percentage of truly relevant documents successfully identified by the system out of all relevant documents in the dataset.[8]

B. *Limitations of Non-Generative TAR (Predictive Coding)*

Another approach to more cost-effective document review is non-generative TAR, which relies on non-generative machine learning techniques. In its early iterations (TAR 1.0), human reviewers label a "seed set" of the document universe, which the algorithm then uses to predict the labeling for other documents. Introduced in the 2000s, these approaches to TAR advanced through the 2010s. More recently, continuous active learning (CAL or TAR 2.0) has become more widely adopted. In CAL, the model is trained continuously (or at certain breakpoints) as reviewers code documents. This workflow eliminates the need for a seed set but still requires substantial review time to achieve acceptable results.

When properly used, TAR has provided substantial cost savings and quality improvement to the document review process. As litigants embraced these technologies, a body of case law emerged, setting forth standards for accuracy and validation. *See, e.g.*, The Sedona Conference TAR Case Law Primer, Second Edition (2023). Courts and litigants generally formed a consensus that 80 to 85 percent estimated Recall is an acceptable and legally defensible level of performance for predictive coding models in large cases, although the specific threshold in a given case also turned heavily on case-specific factors consistent with Federal Rule of Civil Procedure 26.[9]

Yet, non-generative TAR has significant limitations:

Coarser Document Analysis: In general, non-generative TAR models conduct a relatively coarse analysis of documents based on word frequency, concepts, and metadata features, as compared to modern LLMs, which evaluate the contextual meaning of phrases, sentences, and longer passages with a high degree of nuance. The ability of LLMs to handle these nuances gives them a significant advantage in document analysis over prior technologies.

Time-Consuming Startup: Non-generative TAR involves a substantial startup cost—*i.e.*, the human time required to label documents in a seed set or label documents to enable continuous active learning. In addition, predictive-coding models have required successive rounds of training, a time-consuming process. More modern predictive coding, such as CAL, continuously trains the model using reviewer coding but still requires substantial review time to achieve acceptable results. As noted in the Sedona Conference's recent primer on TAR case law, litigants using TAR often run many iterations of review to achieve Recall rates in the realm of 70 to 80 percent.[10]

Limited Transparency: While non-generative TAR models generally attach a numerical score (*e.g.*, 1-100%) to a document's likelihood of relevance, they do not provide an explanation as to why a document was suggested as relevant with that particular score. This lack of explanation requires second-level reviewers to start from scratch when confirming whether a particular document is, in fact, responsive and why. More systemically, if the predictive coding model has incorrectly tagged a series of documents, the lack of explanation makes it difficult to understand why the documents were miscoded and how to correct this miscoding.

Risk of Intercoder Disagreement: TAR's reliance on human judgment may also lead to variability in how the model codes documents. Often, human reviewers will apply an issue code inconsistently, a phenomenon known as inter-coder disagreement. Human biases can also come to bear, leading to fundamental shifts in relevance assessments as

[7] Maura R. Grossman & Gordon V. Cormack, Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review, XVII RICH. J.L. & TECH. 11 at 37 (2011); Maura R. Grossman & Gordon V. Cormack, Technology Assisted Review in Electronic Discovery in Data Analysis in Law (Edward J. Waters ed. 2018). The contributors do not mean to disparage managed review teams. Reviewing large volumes of documents in complex litigation is simply an extremely hard thing for any group of people to do without the assistance of technology.

[8] Practitioners also measure Precision—the percentage of documents identified by the system that are actually relevant.

[9] Bolch Judicial Inst., Technology Assisted Review (TAR) Guidelines, January 2019,

available at https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=1002&context=bolch; *see also* The Sedona Conference TAR Case Law Primer, Second Edition (2023).

[10] *See, e.g.*, *ibid.*

the review proceeds. If a predictive coding model learns from such inconsistently coded documents, it too runs the risk of applying codes inconsistently or becoming confused by the mixed messages sent by disparate reviewers.

Limited Issue Codes and Adaptability: The need to train non-generative TAR models with human labeling makes it pragmatically difficult to apply a large number of issue codes to a complex dataset. Human review speed generally slows down as the number of issue codes increases, which places a practical constraint on the number of issue codes that can be applied with non-generative TAR. Similarly, the amount of time required to train and test predictive coding models is also a major drawback when new issues arise in the midst of the investigation or litigation. These limitations can restrict the usefulness of predictive coding models in more complex cases where the investigation objectives of the case team involve nuanced matrices of relevant facts, participants, categories, and timelines, which can evolve as new facts are uncovered as the discovery process unfolds.

C. *Shortcomings of Linear GenAI Document Review*

As LLMs have evolved, the legal industry has explored whether these models could improve upon the TAR predictive coding algorithms. Since modern LLMs are pre-trained on vast corpuses of data and can perform accurate document classification based on natural language instructions (*i.e.*, "prompts"), they require little to no training on case-specific documents or labeling by subject matter experts. The linguistic and conceptual nuance that is embodied in modern LLMs enables them to make distinctions and perform relevance predictions that significantly exceed TAR methodologies.

However, many of the efforts to utilize LLMs still involve a linear approach to reviewing and coding documents. For example, users provide an LLM with a single or multi-pronged prompt that sets forth case context and a description of what documents are responsive and/or the issues with which the document is to be coded. The LLM then considers the responsiveness of each document in the review population one by one.[11]

This kind of linear GenAI review, while potentially adequate for smaller datasets (hundreds to thousands of documents), encounters significant limitations when applied to more complex and document-intensive matters:

Prompt Overload: A core challenge of linear LLM deployment in large-scale document reviews is the difficulty in creating a single, comprehensive prompt that accurately captures all relevant, nuanced factual issues, given the initial uncertainty inherent in complex cases. Splitting a complex prompt into multiple prompts (such as one prompt per issue code) can multiply the cost of review when each prompt is run linearly over the dataset. Further, the more complex and data-intensive the case is, the less certainty the prompt drafter has about the nuanced factual issues that may be hiding in that dataset, and the more complexity that needs to be packed into a prompt run linearly across the dataset. Attempting to address a multitude of issues in an aggregate prompt (such as a prompt comprising ten issue codes) increases the risk of overloading the LLM, diminishing its accuracy as it struggles to process all instructions simultaneously and creating a risk of inaccurate or incomplete coding. Improperly applied coding may require extensive quality control to detect, and necessitates additional, costly GenAI passes over the dataset to achieve a more accurate result. A complex, multi-faceted prompt requires more computational resources to review each document, resulting in a more expensive process overall.

Limited Issue Codes and High Cost for Large Datasets: The risks of prompt overload and cost overruns have caused some providers of GenAI document review software to place a limit on the number of issue codes or prompt length. However, when these limits are applied, a different problem arises. It is common in complex litigation for a party to be served with 20, 30, or more document requests, and when designing their own investigations, case teams often want to investigate dozens of issues or narrative threads. Document review approaches that are limited either technically or practically to a smaller number of issues thus do not meet the real demands of complex matters. Rather, they impose a constraint into which lawyers must artificially conform their investigation strategy to the limitations of linear GenAI review. This can lead users of linear GenAI review solutions to make compromises on the granularity of their review protocol. When more general issue codes are applied, human review teams must spend additional time sifting through these broad categories of documents to find relevant documents at the back end of the GenAI workflow. Broader issue codes could also require a re-run of an entire review or running multiple review passes to get closer to the more nuanced, desired results, which can rapidly increase the cost of the GenAI effort to the point of being cost-prohibitive.

---

[11] *See, e.g.*, Relativity & Redgrave Data, Beyond the Bar: Generative AI as a Transformative Component of Legal Document Review (2024).

Lack of Cost-Effective Adaptability: The reality of complex litigation is that strategic priorities and the perceived importance of given factual strains constantly shift and evolve as the case progresses. Linear GenAI does not readily adapt to this dynamism. A linear GenAI review process that requires subsequent "passes" across the dataset whenever a new issue arises is often cost-prohibitive as latency increases to intolerable rates. This fundamental disconnect from the innate dynamism of investigations and litigations significantly limits the effectiveness of linear GenAI approaches in complex cases.

II. Syllo's Agentic AI Document Review: A New Paradigm

Syllo has created a novel, agentic AI system for document review that leverages an ensemble of varying-sized LLMs to conduct large-scale document analysis in complex investigations and litigations. Instead of a single-pass approach, Syllo orchestrates multiple LLMs performing distinct roles, including, among others, the roles of strategizing, determining next steps, performing quality control, extracting learnings from documents, synthesizing knowledge from learnings across documents, summarizing, and resolving inter-model disagreement. The behavior of the overall system and the dataflows involved in performing the document analysis are influenced by the outputs of the various LLMs working in concert.

The design and empirical studies of Syllo's agentic approach indicate a higher ceiling on granularity, adaptability, context-sensitivity, cost-efficiency, and complexity as compared to prior non-GenAI and GenAI methodologies for large-scale document reviews.

The upshot for litigation teams is a document analysis system that can perform a highly accurate and categorized review to assist them at every stage of document analysis in litigation, including dataset culling, responsiveness review, subject matter issue coding, privilege review, and identification of hot documents. The benefits of this approach are numerous:

Dynamic Resource Allocation for Cost Efficiency and Accuracy: As the agentic review progresses, the telemetry of the system allows for observation of the amount of work performed by each LLM in each role. More complex and conceptually challenging datasets and documents will trigger more work performed by higher-end LLMs, whereas more straightforward review challenges will lean more heavily on the most cost-effective LLMs. This approach results in more efficient and less costly LLM application to complex document review and permits the ensemble of LLMs more freedom to determine which documents and parts of documents deserve a closer look for particular issues or nuances. It allows selective activation of more powerful or specialized models as needed to improve the quality of the review and minimizes reviewing completely irrelevant documents. This "division of labor" between LLMs mirrors the complexity of the document review project, akin to how complex reviews performed with outsourced reviewers resources might require more quality-control time and subsequent cleanup review hours.

Unlimited Issue Coding: Very significantly, an agentic approach cost-effectively accommodates an unlimited number of issue codes without causing prompt overload and without altering the system's accuracy. This allows the granularity of the document analysis to match the number of requests for production or issues defined by the case team. As shown below, case teams routinely use Syllo to apply dozens of issue codes across large datasets. Also, for each label applied for each issue, the system provides a concise explanation of why a particular document (or its parts) is responsive to that issue with navigation to the relevant document content.

Swift and Cost-Effective Adaptability: Syllo's agentic document review process does not require a seed set and does not require training a model each time the case team wants to add issue codes or change coding parameters. In fact, new legal issues can be integrated seamlessly, allowing on-the-fly adjustments without reprocessing the entire data set, simply by creating another natural language description of the issue to be investigated. The agentic system can leverage prior document analysis to accelerate and reduce the cost of these subsequent targeted queries. The result is a flexible and agile system that can adapt as the legal theories, facts, or other variables change in a matter. In addition, the work product created by the ensemble of models can be leveraged in subsequent analyses over the same dataset, and coding refinement (where the case team realizes an issue code was overly broad or unduly narrow) can be performed surgically at a small fraction of the cost of re-running an entire review.

III. Empirical Validation

Syllo's AI systems have been used on active matters since 2023. In cooperation with law firm partners, the Syllo team has successfully completed more than 80 agentic document reviews in active litigation, with datasets spanning from thousands of documents to more than 2 million documents. Syllo has been used to apply dozens of issue codes in hours or days across hundreds of thousands of documents. These reviews have spanned numerous subject matters, including antitrust litigation, environmental litigation, contract litigation, employment litigation, patent litigation, bankruptcy litigation, mass tort litigation, construction litigation, investment and shareholder disputes, M&A litigation, automotive litigation, real estate litigation, and insurance coverage litigation. Along the way, the Syllo team has developed standard workflows to leverage the capabilities of the agentic system. In the last ten completed responsiveness reviews by Syllo in active litigations, the lowest estimated Recall was 93.4%, the average estimated Recall was 97.8%, and four of the reviews had an estimated Recall of 100%. In the same reviews, the median estimated Precision was 85.9%, and the average estimated Precision was 79.7%.

A. *Syllo in Responsiveness Reviews*

1. Commercial Litigation

Ballard Spahr LLP ("Ballard") represented an enterprise in a commercial litigation in which the Ballard client was interested in reducing the cost of its document review burden. Ballard attorneys educated the client on the option to use Syllo for the responsiveness review, and the client elected to move forward with Syllo as the solution for the responsiveness review.

Ballard's client needed to respond to more than 30 requests for production served by the opposing party, and the collected dataset was more than 100,000 documents. With guidance from the Syllo team, the Ballard trial team articulated more than 25 issue codes that corresponded to the requests for production. It took approximately three hours of human time to set up the instructions for the review. No iteration was performed on the prompts based on human review of documents. Syllo's agentic system applied more than 25 codes to the documents. More than 50,000 documents were identified by Syllo as responsive to one or more issues. The Ballard trial team performed precision testing and elusion testing and determined an estimated Precision of 95.56% and an estimated Recall of 99.4%.[11]

Based on the performance of agentic review for responsiveness, the Ballard team also deployed agentic review on the opposing party's production of more than 25,000 documents to identify deficiencies in the production. Syllo identified numerous specific gaps in the opposing party's production, which enabled Ballard attorneys to demand a remedial production in a matter of days. The trial team also used Syllo to exclude non-responsive documents from the potential privilege documents and to make preliminary privilege calls to facilitate trial team review.

The Ballard team obtained a favorable resolution of the litigation for their client.

"We were able to complete a large document production with a high degree of confidence that we had identified all of the responsive documents," said Casey Watkins, Of Counsel at Ballard. "Setting up the review was straightforward, and the review took a fraction of the time it would have taken if conducted using human review teams. Given the results we achieved and the amount the client paid for the total review, we were able to provide enormous value to our client."

Ballard has subsequently employed Syllo in other document-intensive litigations, including for the application of more than 25 issue tags across a dataset of more than 1.5 million documents in a highly complex commercial litigation, which is currently the most complex litigation in which Syllo's agentic review has been deployed.

2. Commercial Litigation

Joshua Upin, Esq., of Royer Cooper Cohen Braunfeld LLP ("RCCB"), was interested in performing a head-to-head comparison of Syllo's agentic review capabilities against a managed review team leveraging CAL in an ongoing matter. The case selected was a complicated commercial litigation, involving hundreds of entities, many categories of commercial transactions, and more than 25 requests for production.

For the head-to-head comparison, both the managed review team and Syllo received the same review set of slightly less than 16,000 documents. This review set was randomly selected from the broader document population of more than 150,000 documents. The reviewers and the AI system performed their work using separate document coding platforms.

The complexity of the document review created challenges for the review team to accurately tag the documents. The review team required additional guidance and training from the trial team during the review, and based on second-level and quality-control reviews, there were multiple rounds of correction and re-tagging of the first-level review coding. Ultimately, the additional time and remedial work required for the outsourced review team to conduct their review of roughly 16,000 documents resulted in a per document review cost exceeding $2.00 per document.

When the coding results were compared head-to-head, Syllo's performance surpassed the human reviewers by a significant margin. The review team marked more than 5,400 documents in the head-to-head sample as responsive. Elusion testing against the review team revealed an estimated Recall rate below 67% and widespread miscoding that triggered several rounds of re-review. Syllo performed its review without any prompt refinement based on reviewing documents or overturns by the review team, and the RCCB team's validation of Syllo's coding revealed an estimated Recall of 93.44% and an estimated Precision of 69.81%.

"Our adoption of Syllo's review solution has significantly reduced the time it takes for us to review our own documents and identify important documents in opposing parties' production and provided better results than any other alternatives," said Josh Upin, partner at RCCB. "In view of Syllo's superior performance to human review and other available review platforms, it's now my practice to use Syllo on document reviews of any significant size rather than hiring outsourced review teams. Learning and leveraging this technology enables me to get more quickly to documents that matter and better serve our clients while saving them significant expense."

3. Commercial Litigation

A trial team at Quinn Emanuel Urquhart & Sullivan, LLP ("Quinn Emanuel") was engaged as counsel in an accelerated litigation less than two months before trial was scheduled to occur. In the span of six weeks, the team needed to complete responsiveness reviews of more than 30,000 documents, review more than 40,000 documents produced by opposing parties, and complete depositions and pre-trial submissions.

For the responsiveness review, the Quinn Emanuel team initially defined more than 20 issue codes and prepared GenAI review instructions to use Syllo to perform the first-level document review. In a set of more than 30,000 documents, elusion testing confirmed that the estimated Recall was above 98% and the estimated Precision was above 74%. With respect to the review of opposing party productions, the trial team defined more than 40 issues for investigation and used Syllo agentic review to perform first-level review on a rolling basis. Finely categorizing the documents into 40 different categories allowed the team to find critical documents expeditiously, streamlining deposition and trial preparation. As the team identified new avenues of investigation, they defined new sets of review instructions, ran the instructions against the documents produced in the case, and were able to complete follow-up investigations within hours of identifying new issues.

Production in the litigation occurred on a rolling basis. Syllo's ability to store previously defined issue codes and apply them to new productions and collections enabled the trial team to complete first-level reviews of new document sets within hours of their receipt. The trial team also used agentic review to identify deficiencies in the opposing party's production. Given the timeframe of the litigation, this ability to rapidly identify gaps in productions and request supplementation ensured that the team had the evidence they needed to go to trial.

"Facing a high-stakes commercial dispute with only eight weeks until trial, our team needed to accomplish what seemed impossible—complete a substantial document review and production from our client, review opposing counsel's documents to learn the case, and prepare for depositions on an extremely compressed timeline," said Chris Kercher, a partner at Quinn Emanuel. "Syllo transformed our capabilities overnight. We rapidly identified and produced responsive materials from tens of thousands of documents to meet court-ordered deadlines, while simultaneously gaining unprecedented command over the adversary's production. What truly differentiated Syllo was its ability to help us instantly adapt our review strategy as new issues emerged in the opponent's documents, identify critical gaps in their production, and secure vital supplemental productions before deposition deadlines. In fast-moving, complex litigation where strategic advantage is measured in days, not months, Syllo transformed what would have been a logistics challenge into our strategic advantage."

4. Employment Litigation

The plaintiffs' employment firm Outten & Golden LLP has used Syllo to assist with many forms of document review and analysis. As one example, attorneys with Outten & Golden used Syllo to identify documents for production in a collection of 12,543 documents. The Outten & Golden team based their instructions to Syllo closely on the requests for production that had been served on their client in the case, resulting in the definition of 28 issue tags. One issue code was detected as overbroad as the system began its review, and that one issue code was re-drafted. Syllo applied the 28 issue tags across the documents and tagged 484 documents as responsive to one or more requests for production. An associate attorney with Outten & Golden conducted a second-level review of the documents tagged responsive and determined a Precision rate of 84.09%. The associate also performed elusion testing on the documents deemed non-responsive and found no documents in the null-set sample that were responsive, yielding an estimated Recall of 100%.

Based on numerous validations of Syllo's document review solution, Outten & Golden uses Syllo's agentic document review to review large client collections and certain document productions. The firm has observed results that exceed the standards for human and traditional TAR review.

"We've used Syllo's automated document review function, and we were really impressed with the results," said Melissa Lardo Stewart, Partner at Outten & Golden. "Syllo completed a review of thousands of documents in a few hours and our team's review determined that it identified and labeled responsive documents accurately."

B. *Finding Hot Documents*

In addition to reviewing documents for production, Syllo has been used to review the production record to identify hot documents for depositions, pre-trial motions, and trial. In these instances, the agentic reviews are conducted solely for the case team's analysis. Case teams use Syllo's agentic review system to identify decisive key documents from among vast swaths of responsive documents for a given issue.

1. Commercial Litigation

A trial team at Quinn Emanuel was engaged in a fast-paced litigation in advance of a preliminary injunction hearing in a commercial dispute. The team had less than a month to complete document productions and depositions. The trial team decided to use Syllo midway through its deposition preparation, as both a cross-check to make sure that key documents had been identified and to broaden their search to cover new, fast-arising issues more comprehensively than through keyword searching.

The team used Syllo to review more than 40,000 documents and to identify all documents that hit on more than 30 different issues. The trial team used Syllo for a zero-shot review (*i.e.*, the Syllo review ran once without the benefit of any prompt refinement based on document tagging performance). The Quinn Emanuel trial team performed precision testing and elusion testing on the zero-shot review, which confirmed that Syllo identified responsive documents with an estimated Recall of 98.69% and an estimated Precision of 92.83%. As to the issues that the Quinn Emanuel team had already reviewed, Syllo's review confirmed the effectiveness of the trial team's search of the document population. For the new issues, Syllo's review identified more than 200 key documents that had not previously been identified.

"Syllo enables the streamlining of issues and organization of documents for complex litigation, allowing trial teams to move faster and more easily control the factual history of the case," noted senior associate Paul Henderson. "Syllo's agentic system reliably surfaces documents responsive to key issues and navigates substantial factual complexity better than any AI tool I have seen."

2. Commercial Litigation

A trial team at Quinn Emanuel was faced with a tight timeline to prepare for depositions in a complex commercial litigation relating to the private equity industry. The team had already overseen an extended managed document review process and then the court allowed the opposing party to amend its pleadings months before the close of discovery. As a result, there were several new key issues that had not been the focus of a prior review.

The production universe in the case was more than 2 million documents. Syllo was first used to perform a targeted review to return a universe of the top 150 documents related to the new issues raised in the amended pleadings. The documents identified by Syllo were described by the trial team as "incredible."

With depositions scheduled over the next four weeks, the Quinn Emanuel trial team next relied on Syllo to analyze the full production universe, including document productions that were produced after the review was complete, to provide a small number of "hot" documents responsive to 40 issue codes. Syllo's agentic review system churned through the production universe to identify and narrowly return only the few hottest documents that related to any of more than 40 key factual propositions the trial team had identified in the leadup to depositions. Syllo identified, across six witnesses, 750 unique hot documents, 120 of which were newly identified in the case.

"The litigation situation we found ourselves in was familiar to many litigators—we had budgeted a certain amount for document review, and then the court's decision changed the focus of the case," said Melissa Dalziel (Of Counsel at Quinn Emanuel at the time and now Counsel at Alston & Bird). "Syllo allowed us to completely recalibrate our strategy and find a manageable number of the most relevant documents within a vast data set."

3. Commercial Litigation

A trial team at Mayer Brown LLP was more than two years into a high-value litigation. The litigation involved nuanced issues of contracting and construction. More than $300,000 had already been spent on managed document review in the litigation, and reviewers had already reviewed the production universe of more than 400,000 documents spanning more than 8 million pages.

Given the complexity of the document review and the stakes of the litigation, the Mayer Brown team wanted to ensure that key documents had not been overlooked as they prepared to enter a period of depositions. The Mayer Brown team articulated 15 primary issues to be addressed in depositions and trial, and they worked with the Syllo team to conform those issues into 15 issue codes for an automated first-level document review. Syllo completed the review, applying 15 issue codes across more than 400,000 documents in less than one week. Syllo's agentic system identified slightly more than 18,000 documents as highly relevant to one or more of the issue codes and escalated a subset of hot documents for each of the issues.

Upon reviewing the hot documents identified by Syllo, the Mayer Brown team immediately identified critical documents that Syllo had escalated and that had not been previously identified in prior reviews. Ultimately, more than 50 documents that were not escalated by the managed review team but that were identified as highly relevant by Syllo were selected as hot documents by the trial team and slated for use in deposition, which represented almost 20% of the overall hot documents selected.

The Mayer Brown team concluded that having Syllo's agentic review system perform a cross-check review in such a high-value litigation was more than worth the expense. "In high-stakes litigation, the prevailing party is often the one that is able to introduce the most compelling evidence to support their case," said Brandon Renken, partner at Mayer Brown. "Apart from its speed and cost-effectiveness, Syllo more than proved its value by finding key documents that had been missed in the previously conducted managed review."

C. *Additional Applications in Litigation*

Law firms have also successfully used Syllo's agentic document review solution in other creative ways, such as complying with requirements to label production datasets by document request and to perform quality control analysis on human-reviewed datasets.

1. Labeling Every Document by Request for Production or Interrogatory

In a commercial litigation, Nixon Peabody LLP used Syllo to help satisfy a challenging directive from a tribunal to identify the request for production or interrogatory to which each of the 9,000 produced documents pertained. There were more than 30 requests and interrogatories, which translated into applying a coding palette of 30 issue codes. Prior to selecting Syllo for the project, Nixon Peabody began the review project with attorney and paralegal review staff, but due to the number of issue codes, the rate of review was not fast enough to meet the deadline. Nixon Peabody opted to use Syllo. At the time Syllo conducted the review, Nixon Peabody had just a few weeks to comply with the directive.

The document set was unique in that nearly every document was responsive (a richness of 100%). The Nixon Peabody trial team conducted a sample-based second-level review of the tagged documents to ensure that the tagging was correctly applied. Syllo expedited the process of review by providing rationales for the application of each tag. Nixon Peabody's head of eDiscovery and litigation support, Mike Swiatocha, noted that providing the rationales for the tagging of each document inverts how documents are analyzed in the document review process, allowing second-level reviewers to focus on what is important in the case rather than labeling or summarizing documents. This allows for expedited review of the documents on a second-level review.

Not only was Nixon Peabody able to complete the second-level review in only a few days, but the trial team concluded that Syllo's review had been highly accurate.

"Any attorney would have really struggled to complete the project for which we used Syllo, especially given the time pressures involved," said Mike Swiatocha. "Syllo was the perfect solution because it could apply a superhuman number of issue codes to each document and apply them with impressive consistency."

2. Early Case Assessment

A trial team at Quinn Emanuel was engaged in a bankruptcy proceeding and needed to perform early case assessment on multiple sets of document productions, totaling more than 20,000 documents, so they could advise their client on the case posture and represent them at the proceeding. The review needed to be conducted on an expedited basis in advance of a hearing that was scheduled for a week after the team began ingesting documents onto the Syllo platform.

The Quinn Emanuel team deployed Syllo's agentic system in two ways. First, they ran Syllo's agentic document review over each population of documents as they were produced (more than 15 productions) from several parties to the bankruptcy proceeding. These reviews helped organize the documentary record around the key topics in the case, allowing fine-grained control over the documentary universe and helping the trial team quickly spot key issues that advanced their client's interests, which was essential given the tight case deadlines.

Second, they identified more than 25 key facts that were central to their theory of the case and used Syllo's agentic system to identify the 10 most relevant documents for each fact. This allowed the Quinn Emanuel team to quickly compile key documents for multiple depositions scheduled within days of receiving documents in rolling productions from multiple parties ahead of the hearing.

"Syllo is a groundbreaking platform that has quickly become my favorite document review tool," said associate Joanna Caytas. "I was skeptical at the beginning, but Syllo delivered in a cost-efficient way what would have been very difficult to accomplish for a lean team on this timeline."

3. Cross-Checking Reviewers When Responding to Interrogatories

Pillsbury Winthrop Shaw Pittman LLP ("Pillsbury") utilized Syllo in a complex litigation to help analyze and identify documents responsive to contention interrogatories from a universe of more than 78,000 documents. This task was particularly challenging because the human review team had already used a set of 15 highly nuanced issue codes to categorize documents responsive to these interrogatories. Due to aggressive case deadlines that made it difficult to validate the review results in the time remaining, Pillsbury used Syllo to supplement the review team's efforts in order to ensure completeness and accuracy.

Not only did Syllo validate the human review, but it also identified additional important documents that had been missed by human reviewers and highlighted certain documents the reviewers had coded inconsistently or incorrectly. Hence, the Syllo workflow provided an effective quality control mechanism that helped the team identify additional interrogatories to which a document might be responsive. In several instances, even though a human reviewer may have tagged a document as responsive to a single interrogatory issue, Syllo was able to identify additional issues relevant to that document. Importantly, these suggestions included an attribution that simplified validation by pointing out the specific pages or sections of each document that made it responsive to the issue code. Among the examples:

- One issue required fine-grained analysis and likely had few responsive documents. This was selected for a deep-dive evaluation. Reviewers coded 49 items responsive to the issue, and, of these, Syllo coded only seven responsive. The Pillsbury team reviewed the other 42 documents and determined that none of them were responsive to the issue.
- Syllo coded 10 documents as highly responsive to another issue, of which the human reviewers coded just one document as responsive. The Pillsbury team reviewed the ten documents and determined that all of them were probative of the issue.
- Syllo coded 73 documents as likely responsive to another issue, of which human reviewers had coded only six responsive. The Pillsbury team checked the other 67 documents and found most of them were, in fact, responsive to the issue.
- For another issue code, the reviewers coded 22 documents as responsive, but Syllo coded only 11 responsive. Pillsbury reviewed the 11 human-coded documents and determined that none of them were responsive to the issue in question.
- Syllo correctly identified 454 documents as very responsive to issues that had not been flagged by the review team (although they had been found responsive to other closely connected issues in the case), thereby demonstrating the platform's ability to parse nuanced distinctions between related topics.

As a result, the Pillsbury team developed sufficient confidence in the system to begin to deploy it in more standard review workflows. "Syllo's automated document review is reliable and provides unrivaled transparency into specific document characterization," said David Stanton, a partner at Pillsbury. "Far from being a 'black box,' the tagging rationales applied by Syllo let us see why particular tags were applied to specific documents. This enabled workflows to adjust, optimize, and confirm the instructions we provided, and allowed us to very quickly leverage the insights we gained from its use."

VI. Conclusion

Syllo's implementation of agentic document review has consistently and substantially outperformed the benchmarks of prior generations of TAR in real-world complex litigations at a significant cost reduction compared to traditional solutions. Syllo's implementation of agentic document review indicates superior granularity, adaptability, context-sensitivity, and complexity handling as compared to non-GenAI methodologies and linear GenAI methodologies. Case teams leverage Syllo's superior performance for responsiveness reviews at every stage of document analysis in litigation, including responsiveness review, subject matter issue coding, privilege review, and identification of hot documents. When guided by sophisticated litigators, agentic document review provides a powerful strategic advantage in complex investigations and litigations.

Artificial Intelligence and Legal Ethics

David Horrigan
Discovery Counsel & Legal Education Director, Relativity

Today's Topics

- **How Legal Teams Use Artificial Intelligence**

- **Legal Ethics and Technology: Yesteryear**
  - ABA Commission on Ethics 20/20
  - The 2012 Model Rules Amendments and Technology Competence
  - Case Law on Technology and the Practice of Law

- **Legal Ethics and Technology: Today**
  - ABA Formal Opinion 512 on Generative AI and the Model Rules
    - Rule 1.1: Competence
    - Rule 1.4: Communications
    - Rule 1.5: Fees
    - Rule 1.6: Confidentiality of Information
    - Rule 5.1: Responsibilities of Partners, Managers, and Supervisory Lawyers
    - Rule 5.3: Responsibilities Regarding Nonlawyer Assistance

How Legal Teams Are Using Artificial Intelligence

Is Legal Adopting AI? The AI Revolution Is Here

Legal teams' use of artificial intelligence is increasing substantially, with 50% of respondents indicating that their AI use has increased by an average of 43% over the past two years. AI usage across organizations, regardless of their size, is exploding. Now is the time to embrace AI. For legal-specific tasks, AI use has risen by 43% in the past two years.

Mean Increase in AI Usage by Organization Size

Total	43%
1,000+ employees	48%
500–999 employees	47%
100–499 employees	38%
2–99 employees	39%

Compared with two years ago, how has the use of AI (generally or generative) changed, specifically in your legal work?
(Percentage of respondents)

50% increased use

41% stayed the same

3% decreased use

5% not sure

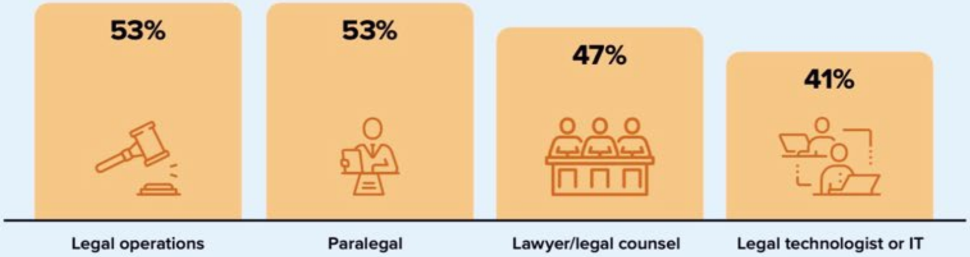
n = 300; Source: IDC's *Generative AI in Legal 2024 Survey*, July and October 2024

Generative AI Trends

Already, generative AI accounts for 48% of all AI use, which is a significant percentage for an experimental and relatively new technology. **Paralegals and legal operations professionals are embracing GenAI at a higher rate than their lawyer and IT counterparts.** In addition, 69% of organizations indicated that legal teams' use of GenAI for legal tasks will increase in the next two years.

The data also reveals that **law firms currently use GenAI significantly more than corporate and government legal teams,** although the latter two's GenAI use is expected to increase in the next year.

Mean Percentage of AI Use by Role

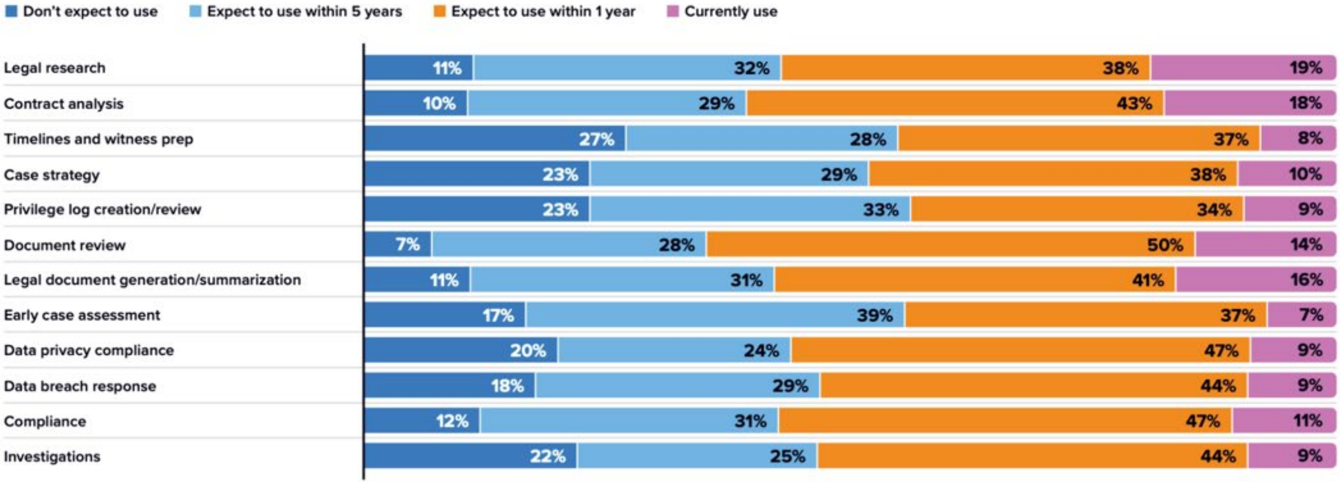


		Industry of Law Practitioner		
	Total	Corporate	Law firm	Government
Currently using	44%	41%	51%	39%
Not using but plan to use in the next 12 months	56%	59%	50%	61%

AI Adoption: Specific Legal Use Case Trends

Respondents report that 48% of their AI use is generative AI and that its use in specific legal tasks will increase in the next few years. Contract analysis and legal research have the highest current GenAI usage, but IDC expects document review to be the leading task within a year.

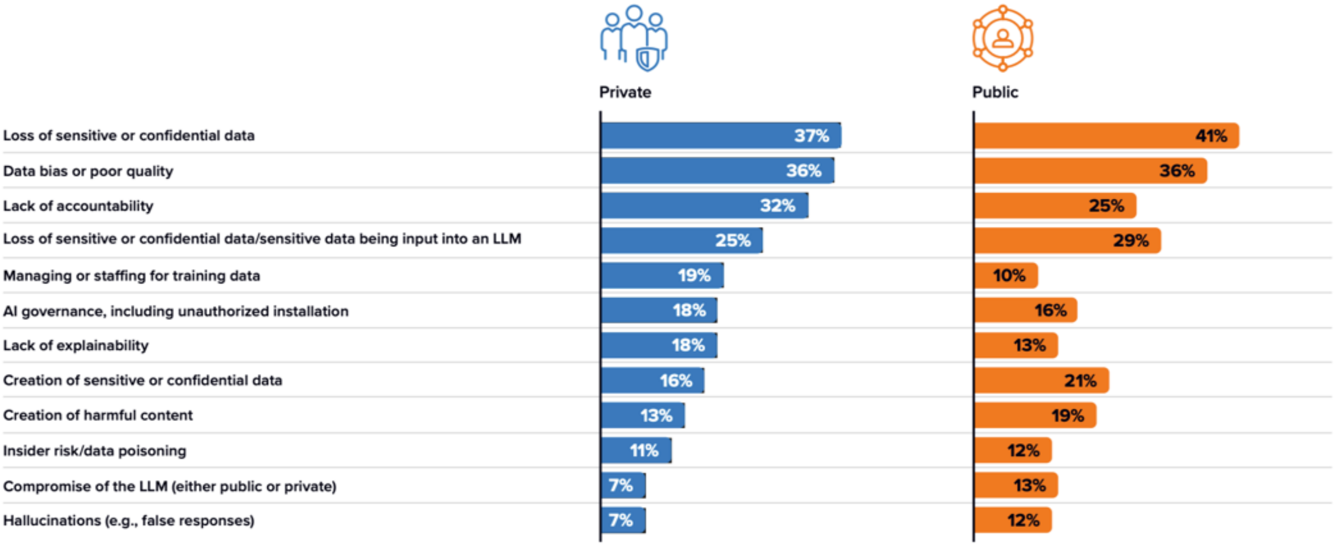
What is the status of using generative AI for each of the following legal use cases in your organization?



n = 300; Source: IDC's *Generative AI in Legal 2024 Survey*, July and October 2024 | For an accessible version of the data on this page, see Supplemental Data in the Appendix.

Concerns with Public and Private GenAI Offerings

What are your biggest concerns when it comes to using commercially available public and proprietary private generative AI offerings?



n = 300; Source: IDC's *Generative AI in Legal 2024 Survey*, July and October 2024 | For an accessible version of the data on this page, see Supplemental Data in the Appendix.

Legal Ethics and Technology: Yesteryear

ABA Commission on Ethics 20/20

- Created in 2009
- A thorough review of the ABA Model Rules of Professional Conduct and the U.S. system of lawyer regulation in the context of **advances in technology** and global legal practice developments.
- Among the Commission's recommendations adopted by the ABA House of Delegates was what would become **Comment 8 to Rule 1.1**

Legal Ethics and Technology: Yesteryear

AMERICAN BAR ASSOCIATION

COMMISSION ON ETHICS 20/20
STANDING COMMITTEE ON CLIENT PROTECTION
STANDING COMMITTEE ON ETHICS AND PROFESSIONAL RESPONSIBILITY
STANDING COMMITTEE ON PROFESSIONAL DISCIPLINE
STANDING COMMITTEE ON PROFESSIONALISM
STANDING COMMITTEE ON SPECIALIZATION
NEW YORK STATE BAR ASSOCIATION

REPORT TO THE HOUSE OF DELEGATES

RESOLUTION

RESOLVED, That the American Bar Association amends the ABA Model Rules of Professional Conduct dated August 2012, to provide guidance regarding lawyers' use of technology and confidentiality as follows (insertions underlined, deletions ~~struck through~~):

Legal Ethics and Technology: Yesteryear

ABA Model Rules of Professional Conduct

Rule 1.1 Competence

A lawyer shall provide competent representation to a client. Competent representation requires the legal knowledge, skill, thoroughness and preparation reasonably necessary for the representation.

Comment

...

Maintaining Competence

[6] To maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, including the benefits and risks associated with relevant technology, engage in continuing study and education and comply with all continuing legal education requirements to which the lawyer is subject.

Legal Ethics and Technology: Yesteryear

ABA Model Rules of Professional Conduct

Rule 1.1 Competence

A lawyer shall provide competent representation to a client. Competent representation requires the legal knowledge, skill, thoroughness and preparation reasonably necessary for the representation.

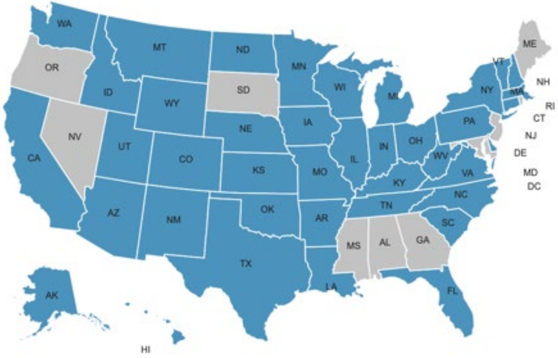
Comment

...

Maintaining Competence

[6] To maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, including the benefits and risks associated with relevant technology, engage in continuing study and education and comply with all continuing legal education requirements to which the lawyer is subject.

Legal Ethics and Technology: Yesteryear



40 States Have Adopted the Duty of Technology Competence

For a complete compilation, *See* Robert Ambrogi, Tech Competence, LawSites

Legal Ethics and Technology: Yesteryear

Technology Practicing Law?

Lola v. Skadden, Arps, Slate, Meagher & Flom LLP,
620 Fed. App’x. 37 (2d. Cir. 2015).

“A fair reading of the complaint in the light most favorable to Lola is that he provided services that a machine could have provided. The parties themselves agreed at oral argument that an individual who, in the course of reviewing discovery documents, undertakes **tasks that could be performed entirely by a machine cannot be said to engage in the practice of law.**”

Legal Ethics and Technology: Yesteryear

Technology Practicing Law?

In re. Patterson, No. 19-24045 (Bankr. D. Md. June 1, 2022).

Although U.S. Bankruptcy Judge Stephen St. John noted the noble goals of the access to justice organization, Upsolve, he cited *Jansen v. LegalZoom.com, Inc.*, and wrote: "Upsolve fails to recognize that the moment the software limits the options presented to the user based upon the user's specific characteristics—thus affecting the user's discretion and decision-making—**the software provides the user with legal advice**."

Legal Ethics and Technology: Today

AMERICAN BAR ASSOCIATION

STANDING COMMITTEE ON ETHICS AND PROFESSIONAL RESPONSIBILITY

Formal Opinion 512 **July 29, 2024**

Generative Artificial Intelligence Tools

To ensure clients are protected, lawyers using generative artificial intelligence tools must fully consider their applicable ethical obligations, including their duties to provide competent legal representation, to protect client information, to communicate with clients, to supervise their employees and agents, to advance only meritorious claims and contentions, to ensure candor toward the tribunal, and to charge reasonable fees.

Rule 1.1: Competence

Formal Opinion 512: Generative Artificial Intelligence Tools

- "To competently use a GAI tool in a client representation, **lawyers need not become GAI experts**."
- "Rather, lawyers **must have a reasonable understanding** of the capabilities and limitations of the specific GAI technology that the lawyer might use."
- "This means that lawyers should either acquire a reasonable understanding of the benefits and risks of the GAI tools that they employ in their practices or **draw on the expertise of others** who can provide guidance about the relevant GAI tool's capabilities and limitations. **This is not a static undertaking**. Given the fast-paced evolution of GAI tools, technological competence presupposes that lawyers remain vigilant about the tools' benefits and risks.

Rule 1.1: Competence

Formal Opinion 512: Generative Artificial Intelligence Tools

- A "lawyer's reliance on, or submission of, a GAI tool's output—**without an appropriate degree of independent verification or review of its output**—could violate the duty to provide competent representation as required by Model Rule 1.1."
- While GAI tools may be able to significantly assist lawyers in serving clients, **they cannot replace the judgment and experience necessary for lawyers to competently advise clients** about their legal matters or to craft the legal documents or arguments required to carry out representations.

Rule 1.1: Competence

Formal Opinion 512: Generative Artificial Intelligence Tools

- A "lawyer's reliance on, or submission of, a GAI tool's output—**without an appropriate degree of independent verification or review of its output**—could violate the duty to provide competent representation as required by Model Rule 1.1."
- While GAI tools may be able to significantly assist lawyers in serving clients, **they cannot replace the judgment and experience necessary for lawyers to competently advise clients** about their legal matters or to craft the legal documents or arguments required to carry out representations.

Rule 1.1: Competence

***Mata v. Avianca, Inc.*, No. 22-cv-1461 (S.D.N.Y. June 22, 2023).**

- Judicial Understatement of the Year: “This court is presented with an unprecedented circumstance.” –U.S. District Judge P. Kevin Castel, May 4, 2023.
- In opposition to a motion to dismiss, Mr. Mata’s counsel filed a submission to the court with non-existent cases—because he relied on ChatGPT alone for his legal research, and ChatGPT “hallucinated” the case law.
- Counsel’s failure to immediately acknowledge the transgression exacerbated the situation, resulting in a $5,000 sanction along with an order that counsel send a copy of the sanctions order to his client as well as the courts cited in the fake court decisions.

Rule 1.1: Competence

***Mata v. Avianca, Inc.*, No. 22-cv-1461 (S.D.N.Y. June 22, 2023).**

Varghese v. China Southern Airlines, Co. Ltd., 925 F.3d 1339 (11th Cir. 2019).

Peterson v. Iran Air, 905 F. Supp. 2d 121 (D.D.C. 2012).

Martinez v. Delta Airlines, Inc., 2019 WL 4639462 (Tex. App. Sept. 25, 2019).

Estate of Durden v. KLM Royal Dutch Airlines, 2017 WL 2418825 (Ga. Ct. App. June 5, 2017).

Ehrlich v. American Airlines, Inc., 360 N.J. Super. 360 (App. Div. 2023).

Miller v. United Airlines, Inc., 174 F.3d 366, 371-72 (2d. Cir. 1999).

In re Air Crash Disaster Near New Orleans, LA, 821 F.2d 1147, 1165 (5th Cir. 1987).

Rule 1.1: Competence

Morgan v. Community Against Violence, No. 23-cv-353 (D.N.M. Oct. 23, 2023).

- A federal district court sanctioned Sorche Morgan, a pro se litigant, who, apparently using generative AI for legal research in an employment discrimination matter, cited fake cases.

- **Although courts 'make some allowances for the pro se Plaintiff's failure to cite to proper legal authority,' courts do not make allowances for a Plaintiff who cites to fake, nonexistent, misleading authorities**," Chief U.S. District Judge Johnson wrote, citing the Tenth Circuit's decision in *James v. Wadas.*

Relativity Legal Education

Rule 1.4: Communications

- Model Rule of Professional Conduct 1.4 provides, in part, that a lawyer shall "**reasonably consult with the client about the means by which the client's objectives are to be accomplished**" and that the lawyer will "promptly comply with reasonable requests for information."

- **Does Rule 1.4 trigger a requirement to inform a client about the use of generative AI?**

Rule 1.4: Communications

ABA Model Rule of Professional Conduct 1.4

- Rule 1.4 provides, in part, that a lawyer shall "**reasonably consult with the client about the means by which the client's objectives are to be accomplished**" and that the lawyer will "promptly comply with reasonable requests for information."
- **Does Rule 1.4 trigger a requirement to inform a client about the use of generative AI?**
- **It depends.**

Rule 1.4: Communications

Formal Opinion 512: Generative Artificial Intelligence Tools

- "The **facts of each case will determine** whether Model Rule 1.4 requires lawyers to disclose their GAI practices to clients or obtain their informed consent to use a particular GAI tool. Depending on the circumstances, **client disclosure may be unnecessary**.

However…

- "Of course, **lawyers must disclose their GAI practices if asked** by a client how they conducted their work, or whether GAI technologies were employed in doing so, or if the client expressly requires disclosure under the terms of the engagement agreement or the client's outside counsel guidelines."

Rule 1.4: Communications

Formal Opinion 512: Generative Artificial Intelligence Tools

- "The **facts of each case will determine** whether Model Rule 1.4 requires lawyers to disclose their GAI practices to clients or obtain their informed consent to use a particular GAI tool. Depending on the circumstances, **client disclosure may be unnecessary**.

However…

- "Of course, **lawyers must disclose their GAI practices if asked** by a client how they conducted their work, or whether GAI technologies were employed in doing so, or if the client expressly requires disclosure under the terms of the engagement agreement or the client's outside counsel guidelines."

Rule 1.5: Fees

Formal Opinion 512: Generative Artificial Intelligence Tools

- Fundamental Takeaway: **No Windfall Profits from Generative AI**
- **Rule 1.5:** "A lawyer shall not make an agreement for, charge, or collect an unreasonable fee or an unreasonable amount for expenses."
- **Formal Opinion 512:** "GAI tools may provide lawyers with a faster and more efficient way to render legal services to their clients, but lawyers who bill clients an hourly rate for time spent on a matter must bill for their actual time."
- **Formal Opinion 512:** Citing *Attorney Grievance Comm'n v. Monfried*, "A fee charged for which little or no work was performed is an unreasonable fee."
- **Formal Opinion 512:** However, a lawyer may bill for time to check the generative AI work product for accuracy and completeness.

Rule 1.6: Confidentiality of Information

Rule Regulating the Florida Bar 4-16

- Florida has a rule corresponding to Rule 1.6, Rule Regulating The Florida Bar 4-1.6. In January of this year, Florida issued Florida Bar Ethics Opinion 24-1 on the use of generative AI, which has substantial references to Rule 4-1.6 and the requirement for confidentiality of information.

- "Existing ethics opinions relating to cloud computing, electronic storage disposal, remote paralegal services, and metadata have addressed the duties of confidentiality and competence to prior technological innovations and are particularly instructive," Florida's Ethics Opinion 24-1 provides, citing Florida Ethics Opinion 12-3, which, in turn cites New York State Bar Ethics Opinion 842 and Iowa Ethics Opinion 11-01 (Use of Software as a Service—Cloud Computing).

Rule 5.1:

Responsibilities of Partners, Managers, and Supervisory Lawyers

Rule 5.3:

Responsibilities Regarding Nonlawyer Assistance

Formal Opinion 512: Generative Artificial Intelligence Tools

- "Managerial lawyers must establish **clear policies** regarding the law firm's permissible use of GAI, and supervisory lawyers must make **reasonable efforts** to ensure that the firm's lawyers and nonlawyers comply with their professional obligations when using GAI tools,"

- "Supervisory obligations also include **ensuring that subordinate lawyers and nonlawyers are trained**, including in the ethical and practical use of the GAI tools relevant to their work as well as on risks associated with relevant GAI use."

Thanks for attending!

Speakers/Moderator Bios

Wesley Oliver

Wes Oliver teaches at the Thomas R. Kline School of Law of Duquesne University. Until seven years ago, he taught courses on criminal law, criminal procedure, and evidence and primarily wrote about the history of the criminal procedure. Then, in an effort give his students a meaningful way to evaluate the probable cause and reasonable suspicion standards, he began working on a project with computer scientists to predict legal outcomes. An effort to automatically assess reasonable suspicion in drug interdiction stops grew out of this collaboration. Realizing that the skills he was learning would be useful to lawyers who practiced in a variety of areas, four years ago Oliver, who had no formal training in computer science, began teaching the topics he was learning. Coding for Lawyers, an introduction to Python programming co-taught with Morgan Gray, was the first programming course he added to Duquesne Kline's curriculum. His course on Statistics and Machine Learning for Lawyers was added last year, and Cryptography and Law will be introduced next year. All three courses are part of Duquesne's new Law and Computing Concentration, a program that was named one of the top 10 innovations in American law schools in 2024-25 by Bloomberg Law. Oliver also teaches week-long intensive versions of these courses as part of a Master's in Law, Technology, and Innovation at the University of Strathclyde in Glasgow, Scotland. In his spare time, he enjoys astronomy, traveling, and spending time with his wife, Kathleen, and their two dogs, Scout and Ruby.

Morgan Gray

Morgan Gray is currently pursuing his Ph.D. in Intelligent Systems at the University of Pittsburgh. The primary areas of his scholarship are criminal procedure, machine learning, natural language processing, and legal text analytics. He also researches and develops systems to enable access to justice. He has published extensively on the automatic identification of legal factors in legal documents, the generation of arguments, and the interpretation of machine learning model predictions. His work facilitates understanding for users in the legal domain and enables empirical legal analysis.

He received his B.A. from Thiel College in 2016, his J.D. from Duquesne Kline School of Law in 2019. He served as a judicial clerk for the Commonwealth Court of Pennsylvania for two years and is admitted to practice in Pennsylvania and the Western District of Pennsylvania. He also is an adjunct professor at the Thomas R. Kline School of Law of Duquesne University and has received the University's Creative Teaching Award for his efforts. He is also a Faculty Scholar for the Carl. G. Grefenstette Center for Ethics in Science, Technology, and Law.

David Horrigan

David Horrigan is Relativity's discovery counsel and legal education director. An attorney, award-winning journalist, law school guest lecturer, and former e-discovery and information governance industry analyst, David is the recipient of the 2024 Lifetime Achievement Award from the International Legal Technology Association (ILTA).

A former in-house counsel and reporter and assistant editor at *The National Law Journal*, David's articles have appeared also in *The American Lawyer, Corporate Counsel, The New York Law Journal, Texas Lawyer, The Washington Examiner,* and others, and he has been quoted by *The Wall Street Journal,* American Public Media's *Marketplace*, *TechRepublic,* and publications of the law schools of Yale, Northwestern, Emory, and others. He is the author of the annual *Data Discovery Legal Year in Review.*

David serves on the Resource Board of the National Association of Women Judges (NAWJ), the Planning Committee of the University of Florida Law E-Discovery Conference, the Global Advisory Board of the Association of Certified E-Discovery Specialists (ACEDS), and the Advisory Board of the *MIT Computational Law Report*.

David holds a juris doctor from the University of Florida Levin College of Law. He is licensed to practice law in Washington, D.C, and he is a Certified Information Privacy Professional (CIPP/US) by the IAPP.

Jeff Chivers

Jeffrey Chivers founded TLATech Inc. in 2019 to build Syllo, a first-of-its-kind software system for litigation. Before launching Syllo, Jeff spent more than ten years of his career in litigation with Sullivan & Cromwell LLP, The Employment Rights Group and Chivers LLP. He clerked for the Honorable Pamela K. Chen of the U.S. District Court for the Eastern District of New York and clerked for the Honorable Thomas L. Ambro of the U.S. Court of Appeals for the Third Circuit. He holds admission to the bars of New York, Pennsylvania, the U.S. Court of Appeals for the Third Circuit, and six U.S. District Courts. Chivers received a J.D. magna cum laude and Order of the Coif from Georgetown University Law Center and a B.A. in Computer Science from Harvard College. He completed the International Chinese Language Program at National Taiwan University in Taipei, Taiwan, and was an editor of the China Guiding Cases Project at Stanford Law School. Chivers is also a Visiting Lecturer in Law at Yale Law School, where he is teaching the seminar, Artificial Intelligence, the Legal Profession, and Procedure.

Tricia A. Martino

Tricia A. Martino is an associate at Pion, Nerone, Girman & Smith, PC, concentrating her practice in the areas of transportation law and general civil litigation. In addition to those areas, Tricia has experience in the areas of insurance defense, product liability, commercial litigation, and insurance coverage and bad faith law. In the appellate arena, Tricia clerked for the Superior Court before transitioning to private practice, and has also written amicus curiae briefs to the Pennsylvania Supreme Court.

Tricia received her B.A. and B.S. degrees from the Pennsylvania State University in 2014, and her J.D. from Duquesne University School of Law in 2020, where she was awarded four CALI awards and was published in the 58th edition of the Duquesne Law Review. She is admitted to practice before the state and federal courts of Pennsylvania and West Virginia.

Tricia is an active member of the Allegheny County Bar Association, holding leadership positions including Secretary of the Young Lawyers' Division, member of the Judiciary Committee and Bench-Bar Committee, and the 2021-2022 Bar Leadership Initiative Class.