LLM Legal Risk Management, and Use Case Development Strategies to Minimize Risk.

August 2023

PREPARED FOR
AI Village Presentation

PRESENTED BY
Robert J Hudock
Nicholas Maietta
King & Spalding LLP

https://github.com/rhudock/ai-ontology

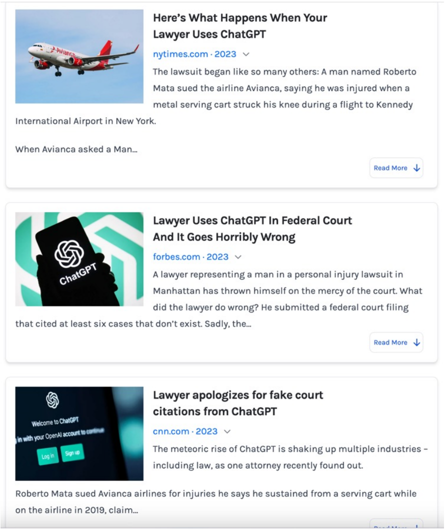
Agenda

Use Cases	Summarize Meetings Dynamic 50 State Surveys Semantic Search
Enabling Technologies	Domain Specific Ontology Semantic Search Large Language Models (LLMs)
Risk Management Frameworks	NIST RMF OWASP LLMs MITRE ATLAS
Legal Obligations; State Laws	50 State Survey Use Case •Virginia •Colorado •Connecticut
Meeting Documentation Generation (Use Case)	Data Risk •De-identification Use Case Model Risk Implementation Risk Regulators
Threat Modeling	

Use Cases

Promise and Peril of Large Language Model (LLM): Use Cases in the Legal Profession

- **Accuracy and Reliability:** LLMs may provide incorrect or incomplete legal information, which can lead to flawed legal advice. (see https://incidentdatabase.ai/cite/541/#r3011)(Avianca personal injury case).
  - Responsibilities regarding nonlawyer assistance
- **Ethical Considerations:** Dependence on LLMs could lead to questions about the due diligence performed by a lawyer, possibly violating professional ethics guidelines. (ABA Model Rules)
- **Over-reliance:** Lawyers may become too dependent on LLMs, hindering their own critical thinking and legal reasoning skills.
- **Client Confidentiality:** Using LLMs could lead to unintentional data breaches or sharing of confidential client information if not properly safeguarded.



Here's What Happens When Your Lawyer Uses ChatGPT

nytimes.com · 2023

The lawsuit began like so many others: A man named Roberto Mata sued the airline Avianca, saying he was injured when a metal serving cart struck his knee during a flight to Kennedy International Airport in New York.

When Avianca asked a Man...

Lawyer Uses ChatGPT In Federal Court And It Goes Horribly Wrong

forbes.com · 2023

A lawyer representing a man in a personal injury lawsuit in Manhattan has thrown himself on the mercy of the court. What did the lawyer do wrong? He submitted a federal court filing that cited at least six cases that don't exist. Sadly, the...

Lawyer apologizes for fake court citations from ChatGPT

cnn.com · 2023

The meteoric rise of ChatGPT is shaking up multiple industries – including law, as one attorney recently found out.

Roberto Mata sued Avianca airlines for injuries he says he sustained from a serving cart while on the airline in 2019, claim...

ABA Model Rules

Duty of Competence: Rule 1.1 of ABA Model Rules.

- Lawyers must provide competent representation to clients.
- Competence entails legal knowledge, skill, thoroughness, and preparation.
- Lawyers must stay informed on current technology.
- 2012: ABA added Comment 8 to Rule 1.1.
- Comment 8 emphasizes the importance of understanding changes in both law and technology.
- Lawyers need to be aware of both benefits and risks of relevant technology.

Duty of Confidentiality: ABA Model Rule 1.6.

- Lawyers owe clients a duty to maintain their information confidentially.
- Lawyers must prevent unauthorized or inadvertent disclosure and access to client information.
- Using AI may involve sharing client information with third-party vendors.
- Lawyers must ensure clients' information is safeguarded.
- Lawyers should discuss confidentiality safeguards with AI providers.
- Use AI only if confident in the security of client's confidential information.

Illustrative Example:
Prompt: Who is Robert Hudock?
GPT-4 Not Supper Accurate

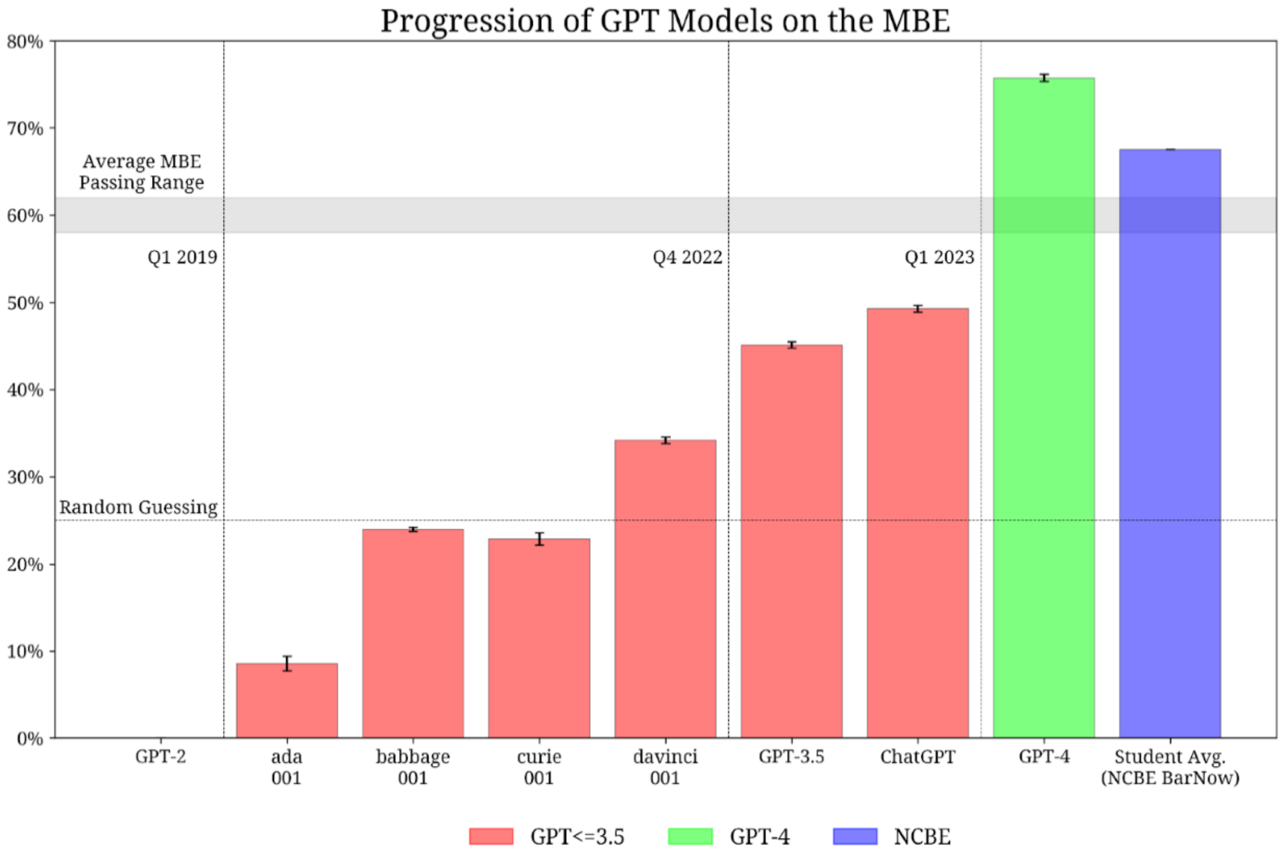
- Robert Hudock is a recognized expert in cybersecurity law. As of my last knowledge cutoff in September 2021, he is a partner at the law firm ~~Hudock Employment Law Group~~. He specializes in the area of cybersecurity, privacy, and data protection, advising clients on how to comply with relevant regulations, managing risk, and responding to cyber incidents.
- Mr. Hudock has significant experience in information security and privacy issues, having worked as a privacy attorney, ~~a systems engineer~~, and an information security professional in the past. He has provided guidance to clients in a variety of sectors, including healthcare, financial services, and technology.
- He has certifications like the ~~Certified Information Privacy Professional/US (CIPP/US) from the International Association of Privacy Professionals~~ and the Certified Information Systems Security Professional (CISSP) from the International Information System Security Certification Consortium.

ABA Model Rule of Professional Conduct 5.3: "Responsibilities Regarding Non-lawyer Assistance"

- **Rule 5.3** extends requirements from **Rule 1.6,** which covers confidentiality, to legal assistance from software and software professionals.
  - Third-party software professionals and AI legal software fall under this rule.
  - Software professionals need awareness of a lawyer's duties to clients.
- Lawyers should educate software professionals on: Legal confidentiality; Evidentiary privilege.
- Digital data privacy and security when using AI software tools.
- Lawyers should understand software development processes (e.g., Agile) and tech concepts related to data privacy, security, and AI to communicate effectively with tech professionals.
  - Software, whether AI-driven or not, should adhere to the model rules, especially regarding client data's privacy and security.
  - While many software professionals prioritize data privacy/security, they may need more training on client confidentiality and privilege.
  - No federal U.S. law mandates software professionals to follow data privacy/security rules.
  - Good legal references for data privacy/security include the EU's GDPR and the U.S.'s CCPA. Other references include U.S. federal laws on health and financial privacy.

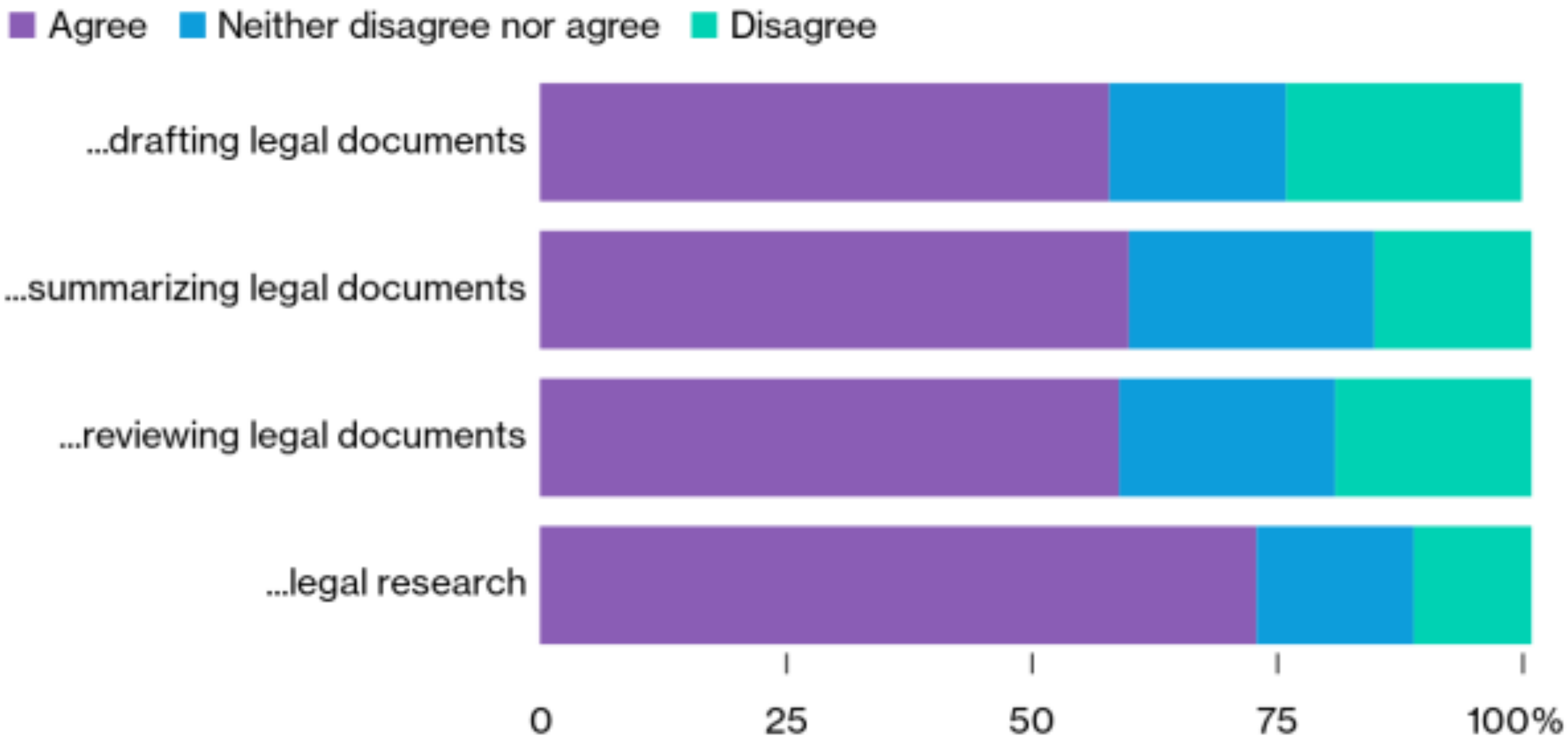
Unauthorized Practice of Law (UPL)

- Use of LLMs could potentially constitute the "practice of law" for UPL purposes
- Definition of "practice of law" vague
  - E.g., *In re Wolf* (FL 2009) ("if the giving of such advice and performance of such services **affect important rights of a person** under the law, and . . . requires that the persons giving such advice **possess legal skill and a knowledge of the law greater than that possessed by the average citizen**").
- Types of risks:
  - **Non-lawyers** preparing documents or providing advise for others (see, *UPL v. Parsons* in Texas re: Quicken Family Lawyer)
  - **Lawyers** assisting non-lawyer in UPL



Attorneys Recognize the Potential for Generative AI in Legal Practice

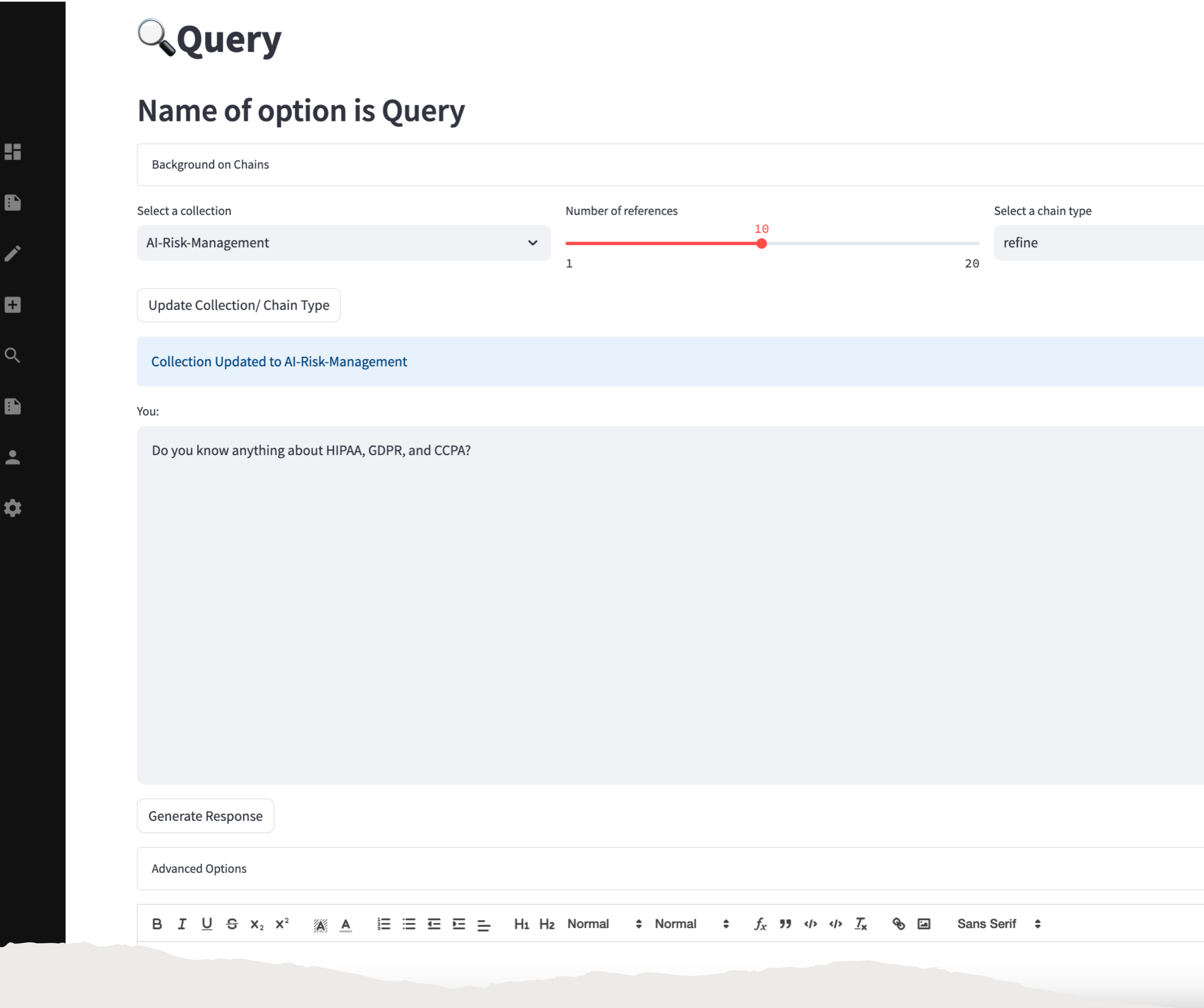
'Generative AI can be used for...'



Source: Bloomberg Law's State of Practice Survey 2023, conducted from March 21, 2023 to April 14, 2023. Respondents were asked to what extent they agree with the statement.

Bloomberg Law

https://news.bloomberglaw.com/bloomberg-law-analysis/analysis-three-considerations-for-attorneys-using-generative-ai (visited August 9th, 2023)



Query

Name of option is Query

Background on Chains

Select a collection: AI-Risk-Management

Number of references: 10

Select a chain type: refine

Update Collection/ Chain Type

Collection Updated to AI-Risk-Management

You:

Do you know anything about HIPAA, GDPR, and CCPA?

Generate Response

Advanced Options

Internal Use Cases

Remove sensitive identifiers from content to allow data to be used with an LLM with appropriate approvals.

Summarize Meetings

Attorney with constance conference calls, trouble keeping up with the decisions made in the meeting, news issues, and/ or documenting meetings to address any legal risks.

Semantic Document Collections

System allows an attorney to create collections of documents to specific areas of the law. The system ingests content and indexes this content with a domain specific ontology.

Query Knowledge Base

As a lawyer in a highly technical domain with complex legal requirements I need to be able to apply historical knowledge and ulations to assist a client with a novel problem or issue.

Surveys

Enabling Technologies: LLMs, ontologies, semantic databases.

Large Language Models (LLMs)

- LLMs have vast knowledge bases, which can be fine-tuned with legal texts, case law, statutes, and legal theories.
- This provides them with a foundational understanding of legal topics and the ability to generate contextually relevant responses.

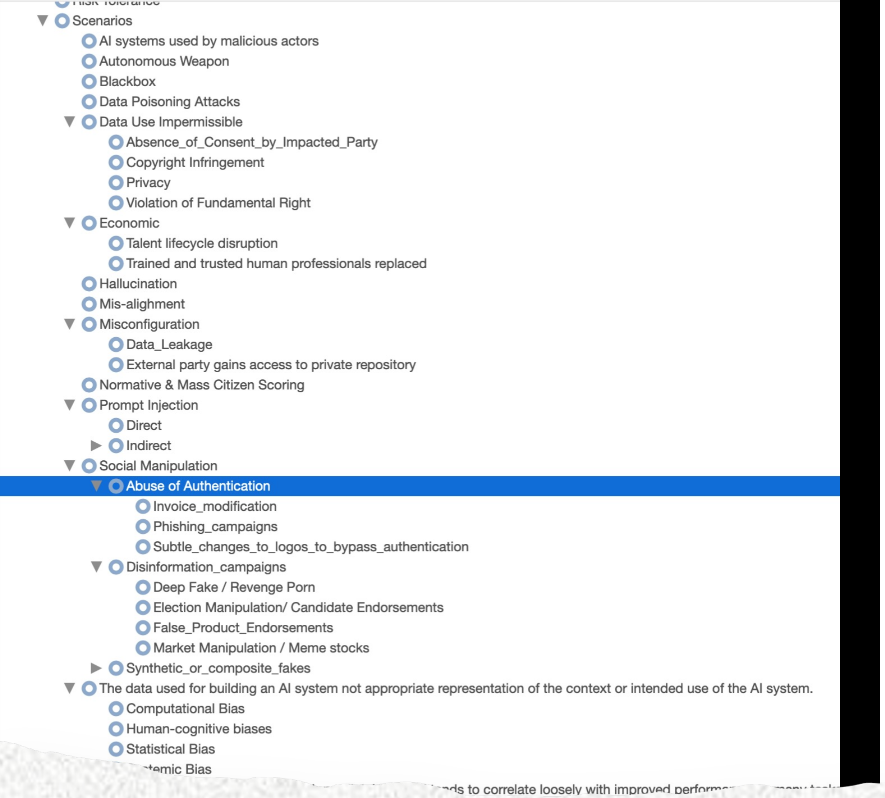
Ontologies

- Structured framework for organizing and defining information.
- Applied to a specific domain like law, defines legal terms, relationships, statutes in a structured manner.
- Ontologies allows LLMs to “understand” and ”navigate” complex legal concepts following a structured definition of the domain.
- *See Sample Privacy Ontology (Slide 21)*

Semantic Database

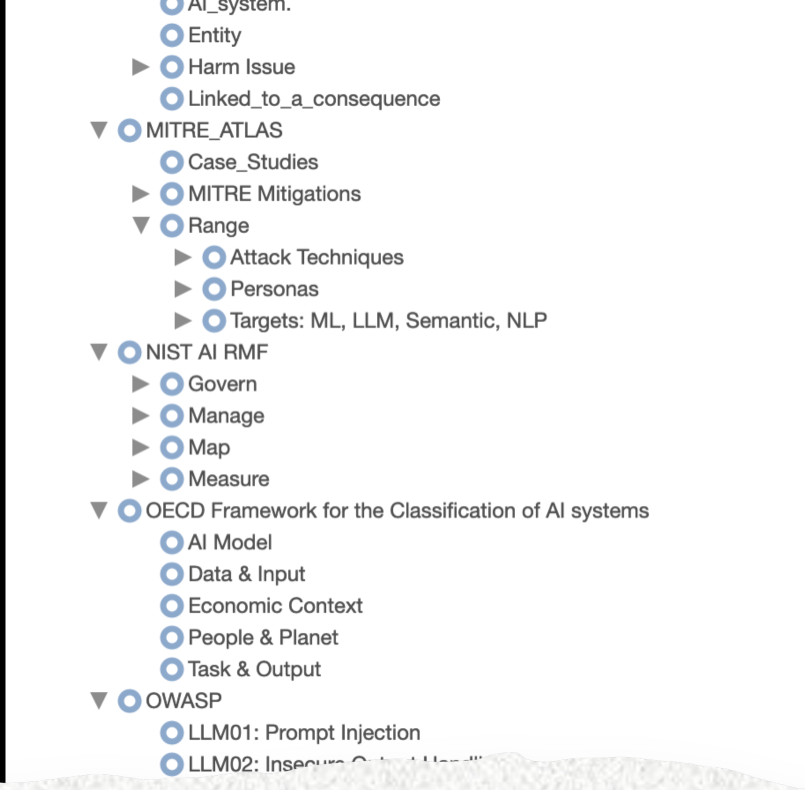
- Offer a more refined source of information.
- Designed to understand the meaning and context behind data.
- When querying a semantic database for legal matters, the LLM can pull out relevant case law, statutes, and precedents that are directly applicable to a specific query, rather than offering generic legal advice.

Risk Management

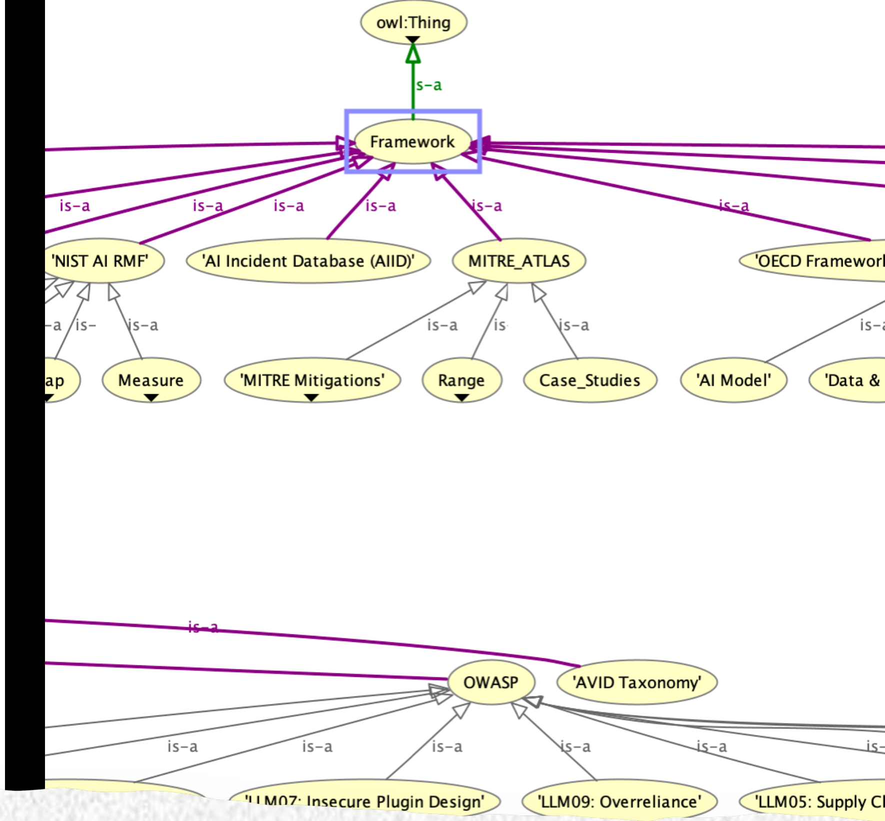


- Scenarios
  - AI systems used by malicious actors
  - Autonomous Weapon
  - Blackbox
  - Data Poisoning Attacks
  - Data Use Impermissible
    - Absence_of_Consent_by_Impacted_Party
    - Copyright Infringement
    - Privacy
    - Violation of Fundamental Right
  - Economic
    - Talent lifecycle disruption
    - Trained and trusted human professionals replaced
  - Hallucination
  - Mis-alighment
  - Misconfiguration
    - Data_Leakage
    - External party gains access to private repository
  - Normative & Mass Citizen Scoring
  - Prompt Injection
    - Direct
    - Indirect
  - Social Manipulation
    - Abuse of Authentication
      - Invoice_modification
      - Phishing_campaigns
      - Subtle_changes_to_logos_to_bypass_authentication
    - Disinformation_campaigns
      - Deep Fake / Revenge Porn
      - Election Manipulation/ Candidate Endorsements
      - False_Product_Endorsements
      - Market Manipulation / Meme stocks
    - Synthetic_or_composite_fakes
  - The data used for building an AI system not appropriate representation of the context or intended use of the AI system.
    - Computational Bias
    - Human-cognitive biases
    - Statistical Bias
    - temic Bias

ds to correlate loosely with improved perform



- AI_system.
- Entity
- Harm Issue
- Linked_to_a_consequence
- MITRE_ATLAS
  - Case_Studies
  - MITRE Mitigations
  - Range
    - Attack Techniques
    - Personas
    - Targets: ML, LLM, Semantic, NLP
- NIST AI RMF
  - Govern
  - Manage
  - Map
  - Measure
- OECD Framework for the Classification of AI systems
  - AI Model
  - Data & Input
  - Economic Context
  - People & Planet
  - Task & Output
- OWASP
  - LLM01: Prompt Injection
  - LLM02: Insec



Risk Management Frameworks

Resources to Inform Risk Management Process

AI Incident Database

- Index real-world harms or near harms caused by AI system deployment.
- Learn from experiences to prevent or mitigate negative outcomes.
- Available at https://incidentdatabase.ai/
- AVID Database of Example Vulnerabilities

Taxonomies

- AVID: Taxonomy. Defines potential risks at different stages of a developing an AI system.
    - Security, Ethics, Performance.
    - Lifecycle
- The CSET AI Harm Taxonomy characterizes AI incidents and classifies harms of relevance to the public policy community.[2]
- MITRE Adversarial Threat Landscape for Artificial-Intelligence Systems (https://atlas.mitre.org/)

Ontologies

- Legal Matter Standard Specification (LMSS) SALI. https://github.com/sali-legal/LMSS; see also SALI LMSS Search Tool, A RESTful API for searching and tagging text with the SALI Legal Matter Standard Specification (LMSS) ontology.
- Securing Artificial Intelligence (SAI); AI Threat Ontology.[1]
- Stanford Legal Taxonomy (https://taxonomy.legal/)

1 | https://www.etsi.org/deliver/etsi_gr/SAI/001_099/001/01.01.01_60/gr_SAI001v010101p.pdf

2 | Mia Hoffmann, Mina Narayanan, Ankushi Mitra, Yu-Jie Liao, and Heather Frase, “CSET AI Harm Taxonomy for AIID and Annotation Guide,” (available at https://github.com/georgetown-cset/CSET-AIID-harm-taxonomy.git)

Legal Obligations & State Laws

Sources of Obligations

Laws and Regulations:

General (e.g., FTC Section 5)

Sectoral-focused (e.g., HIPAA, GLBA, FCRA; Employment)

Privacy-focused

- US: State Laws (next slide), including CCPA's upcoming rulemaking on ADM
- EU/UK: Guidance on GDPR applicability to AI from DPAs such as France's CNIL and UK's ICO; CNIL Action Plan on AI; EDPB task force on ChatGPT

Technology-focused

- US: California's Chat Bot Disclosure Law
- EU: AI Act; China: Interim Measures for the Management of Generative Artificial Intelligence Services

Regulatory actions, direction, and case

US

- FTC: Inquiry into OpenAI; model deletion cases (*Edmodo*); qualifying harms (*Kochava*);
- FDA: Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan
- Regulator guidance/priorities: CFPB/DOJ/EEOC/FTC Joint Statement

EU: Italy's Garante on ChatGPT; Clearview AI fines

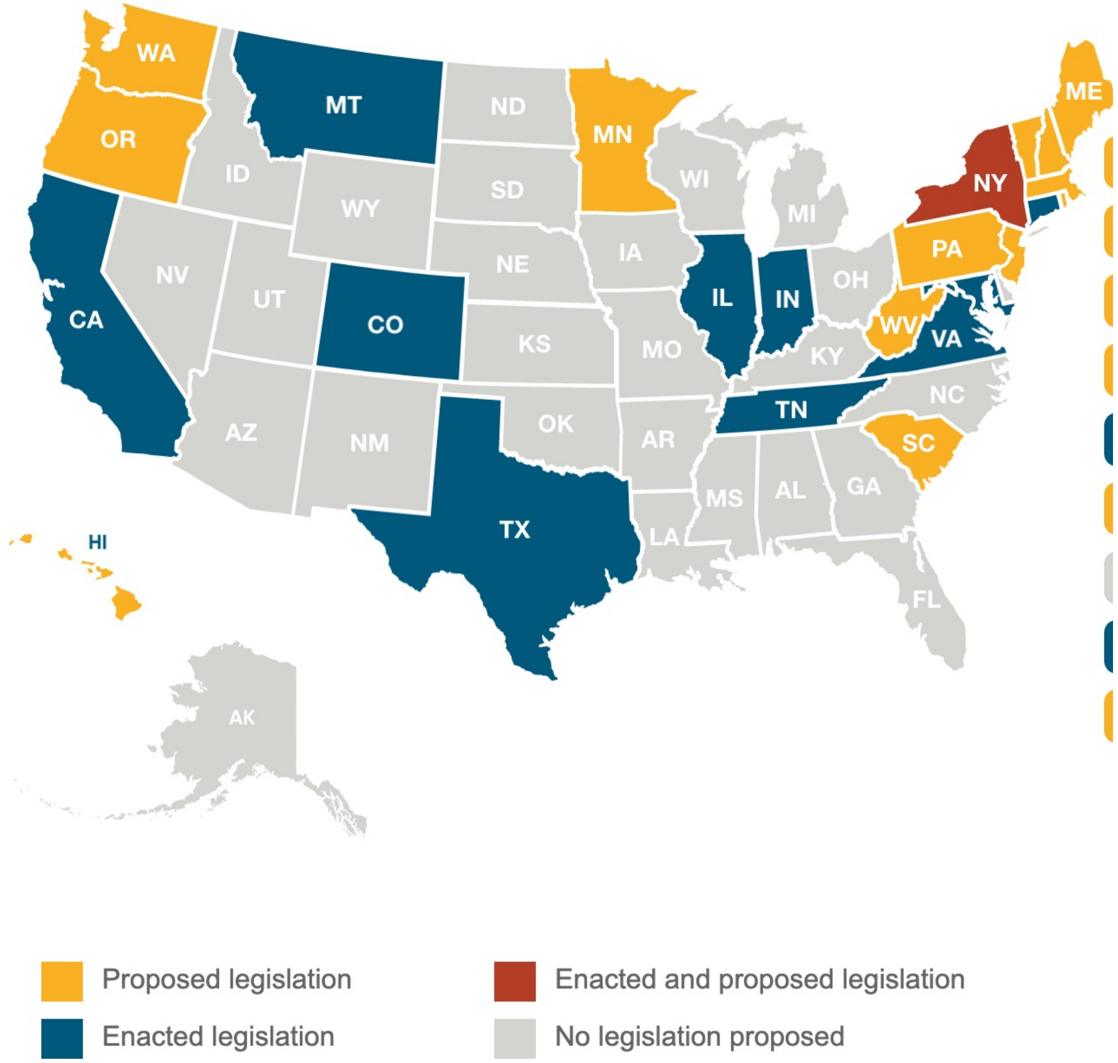
Self-regulations, such as voluntary commitments in the White House' "Ensuring Safe, Secure, and Trustworthy AI" (available at https://www.whitehouse.gov/wp-content/uploads/2023/07/Ensuring-Safe-Secure-and-Trustworthy-AI.pdf)

Contracts, such as AUPs (e.g., ChatGPT usage restrictions) and license restrictions (e.g., Llama 2 Community License)

Best practices / code of conduct, such as Google's Responsible AI practices

STATE-BY-STATE AI LEGISLATION SNAPSHOT

- Opt-Out of Automated Decision Making
- Inference Opt-Out
- Data Privacy Impact Assessment (DPIA)
- Express consent from individuals before using AI to analyze their personal data
- Reasonable Security Controls
- Chatbot Disclosure Rules (CA)
- Restrictions for Use in Hiring (IL, NYC)



Opt-Out Rights: Automated profiling of personal data

Virginia Consumer Data Protection Act[1]

- **Definition of Profiling**
  - Automated processing of personal data
  - To predict personal aspects such as:
    - Economic situation
    - Health
    - Personal preferences
    - Behavior, reliability, movement, or location
- **Consumer Rights: Right to opt out of profiling for:**
  - Decisions producing legal or significant effects
  - Example: Deciding on providing or denying services
- **Controller Responsibilities:**
  - Provide a clear privacy notice to consumers
  - Notice must include purpose for processing data
  - If used for profiling, it must be explicitly stated

Colorado Privacy Act (CPA)[2] **and Associated Regulations**[3]

- Allows consumers to opt-out of personal data processing for specific profiling decisions that produce "legal or similarly significant effects".
- Defined such "legal or [] significant effects" as those impacting provision or denial of finance, housing, education, health care, criminal justice, etc.
- Controllers must conduct a **Data Protection Impact Assessment (DPIA)** for processing posing a high risk to consumers.
- Risky processing includes profiling that can cause:
  - Unfair/deceptive treatment or unlawful impact on consumers.
  - Financial/physical harm.
  - Offensive intrusions into privacy.
  - Any other substantial injury to consumers.
- Automated decision-makers must:
  - Ensure their operations don't pose the aforementioned risks.
  - Include them in their DPIA.
- "Profiling" defined as **any form of automated processing of personal data** to evaluate, analyze, or predict personal aspects concerning an identified or identifiable individual's **economic situation, health, personal preferences, interests, reliability, behavior, location, or movements.**

1 | Available at: https://law.lis.virginia.gov/vacodefull/title59.1/chapter53/
2 | Available at: https://leg.colorado.gov/sites/default/files/documents/2021A/bills/2021a_190_rer.pdf
3 | Available at: https://coag.gov/app/uploads/2023/03/FINAL-CLEAN-2023.03.15-Official-CPA-Rules.pdf

Connecticut

- **Covered Individuals: State residents.**
- Excludes those in commercial, employment contexts, or company roles.
- Controllers doing business in the state or targeting state residents.
  - Applies to those managing personal data of 100,000+ consumers OR 25,000 consumers + 25% revenue from data sales.
- **'Do Not Sell' Requirement:**
  - Defines data sale as exchanging data for money/other value.
  - Controllers must disclose processing & opt-out means.
- **Data Breach Notification:**
  - Processors to assist controllers with data security and breach notifications.
  - Security Measures:
- Controllers must have strong data security practices (administrative, technical, physical).
- **Exemptions:**
  - Exceptions include public health, credit reporting, educational rights, employment data, etc.
- Controllers must conduct a **Data Protection Impact Assessment (DPIA)** for processing posing a high risk to consumers.
- Enforcement & Penalties: Solely by the Attorney General
  From July 1, 2023 - Dec 31, 2024: Controllers get a notice & 60 days to address the violation.
- Violations regarded as unfair trade practices.
- **Consumer Privacy Rights:**
  - Confirm, access, correct, delete personal data.
  - Get a portable copy of personal data.
  - Opt-out of targeted advertising, data sales.
- **Definitions:**
  - Consumer: State resident minus specific roles & contexts.
  - Personal Data: Linked or linkable info to an individual (excludes public/de-identified data).
  - Sensitive Data: Includes racial, health, sexual, biometric, child data, geolocation.
  - De-identification: Making data unlinked to an individual/device.
  - Controller: Entity deciding the purpose/means of data processing.
  - Processor: Entity processing data on the controller's behalf.
  - Third Party: Entity other than consumer, controller, or processor.
  - Profiling: Automated processing . . . to evaluate, analyze or predict personal aspects related to an . . . individual's economic situation, health, personal preferences, interests, reliability, behavior, location . . .

50 State Survey

Objective:

- For an attorney specializing in privacy & security laws to stay updated on state-specific laws.

Key Needs:

- Maintain a comprehensive summary of legal standards for each state.
- Develop a common language for describing each law to facilitate comparison.
- Automatic detection & flagging of changes in laws.

Detailed Requirements:

- Comprehensive summaries of each state's privacy and security laws.
- A systematized description language for ease of comparison.
- Real-time monitoring of legal databases for law amendments or changes.
- Immediate alerts to the attorney on identified changes.
- Enhancing attorney's capability for proactive client advice based on the latest state laws.

Sample Prompt w/ Privacy Ontology

- Your task is to conduct an in-depth examination of a privacy law, focusing on the following areas:
- - Effective Date
- - Covered Individuals
- - Scope: Identify the covered individuals and applicable entities.
- - Inconsistencies or contradictions in terminology
- - 'Do Not Sell' requirement
- - Data Breach Notification requirement
- - Required security safeguards or controls - Exceptions or Exemptions - Enforcement mechanisms, Penalties, and Fines
- - Identify the following privacy rights:
-     - Right to Access
-     - Right to Deletion/Erasure
-     - Prohibition on Discrimination
-     - Right to Opt-out
-     - Opt-in (age)
- . - Right to Data Portability
-     - Right to Rectification
-     - Right to Restriction of Processing
-     Provide definitions for terms like:
-         - Person/Consumer/Data Subject,
-         - Personal information/PII,
-         - Sensitive Data/Private Information,
-         - Business,
-         - Deidentification,
-         - Service Provider/Provider
-         - Controller,
-         - Operator,
-         - Data Collector/Collector.
- Below is a draft analysis:

----------------

{st.session_state.statute_analysis}

----------------

Additional context for refining the analysis:

----------------

{prompt_request}

----------------

Please update the existing analysis based on the new context. Present your findings as a concise report in markdown format.

Updated Analysis of the Colorado Privacy Act (CPA)

Effective Date
The Colorado Privacy Act (CPA) will become effective on July 1, 2023.

Covered Individuals
The CPA covers "Consumers," defined as Colorado residents acting only in an individual or household context. It does not apply to individuals acting in a commercial or employment context.

Scope
The CPA applies to "Controllers," defined as persons that, alone or jointly with others, determine the purposes for and means of processing personal data. It also covers "Data Brokers," defined as Controllers that knowingly collect and sell to third parties the personal data of a Consumer with whom the Controller does not have a direct relationship. Moreover, the Act covers "Processors," defined as a person that Processes Personal Data on behalf of a Controller.

Inconsistencies or contradictions in terminology
The CPA provides definitions for various terms. However, some of these definitions might be subject to interpretation, leading to potential inconsistencies or contradictions.

'Do Not Sell' requirement
The CPA grants consumers the right to opt out of the processing of personal data for targeted advertising or the sale of personal data. The CPA also introduces the concept of a Universal Opt-Out Mechanism, which allows consumers to opt-out from all Controllers they interact with without having to make individualized requests with each Controller.

Data Breach Notification requirement
The CPA does not explicitly mention a data breach notification requirement.

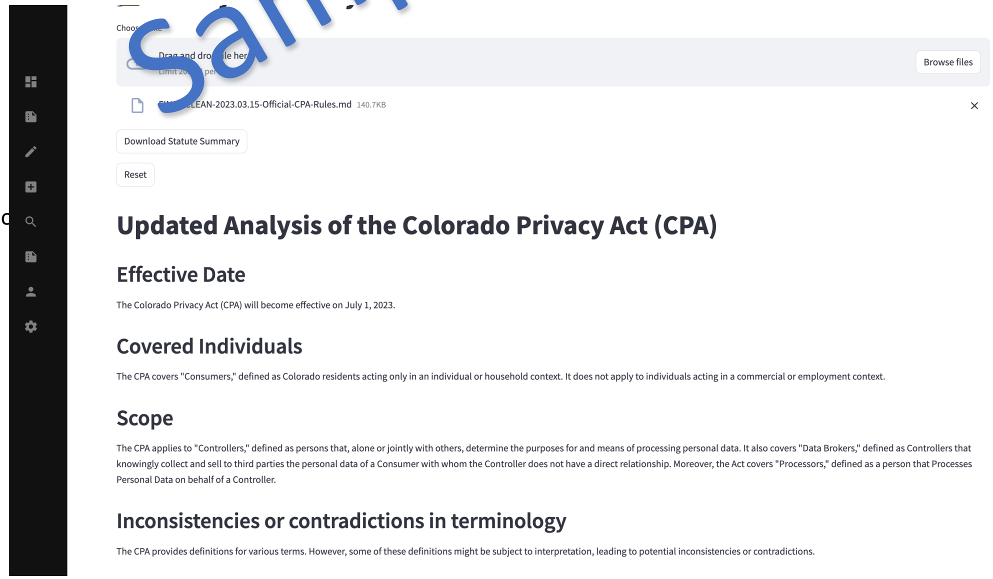
Required security safeguards or controls
The CPA requires Controllers to implement reasonable security measures, consistent with 4 CCR 904-3, Rule 6.09, to protect Personal Data. The safeguards should be designed to protect against unauthorized or unlawful access to or use of Personal Data and the equipment used for the Processing and against accidental loss, destruction, or damage.

Exceptions or Exemptions
The Act is subject to the applicability requirements and exemptions provided in C.R.S. § 6-1-1304.

Enforcement mechanisms, Penalties, and Fines
The CPA enforcement powers are vested in the Colorado Attorney General or District Attorney, as applicable, under C.R.S. §§ 6-1-1310 and 6-1-1311. However, these powers can't infringe upon rights protected by the United States Constitution or Colorado Constitution, including the right to freedom of speech or freedom of the press.

Privacy Rights
The CPA grants several privacy rights, including the right to opt out of the processing of personal data for targeted advertising or the sale of personal data, right to access, right to correction, and right to deletion. The CPA also introduces the right to data portability.

Sample Output



Updated Analysis of the Colorado Privacy Act (CPA)

Effective Date

The Colorado Privacy Act (CPA) will become effective on July 1, 2023.

Covered Individuals

The CPA covers "Consumers," defined as Colorado residents acting only in an individual or household context. It does not apply to individuals acting in a commercial or employment context.

Scope

The CPA applies to "Controllers," defined as persons that, alone or jointly with others, determine the purposes for and means of processing personal data. It also covers "Data Brokers," defined as Controllers that knowingly collect and sell to third parties the personal data of a Consumer with whom the Controller does not have a direct relationship. Moreover, the Act covers "Processors," defined as a person that Processes Personal Data on behalf of a Controller.

Inconsistencies or contradictions in terminology

The CPA provides definitions for various terms. However, some of these definitions might be subject to interpretation, leading to potential inconsistencies or contradictions.

Use Case Deep Dive

Meeting Documentation

Numerous conference calls throughout the day, challenging to keep track of all the decisions, stay updated with new issues, and document instances where the client chooses to absorb a legal risk as a business decision.

Need a system that can effectively and accurately record the key decisions made during these meetings, capture any new issues that emerge, and document instances where the client decides to take on a legal risk.

This will ensure the user does not miss any vital information during a busy day.

Solution

This system could be an AI-powered note-taking application integrated into our conference call platform.

It should process conversations in 30 second increments, extract significant points, and highlight findings and risks. A transcript or recording of the meeting is not retained, and no content is ever written to disk.

The AI tool should allow for a post-meeting review, where I can manually validate and modify the captured points to ensure their accuracy and completeness.

Ultimately, this system will allow me to stay updated with all decisions and issues and maintain a clear record of legal risks the client has chosen to absorb, enhancing my legal advice and overall service to the client.

The attorney can accurately track all decisions made during conference calls.
The system can detect, and document new issues raised during the calls. The system accurately identifies and records instances of assumed legal risk. The process improves the attorney's ability to manage and advise on legal risks. Should be impossible to recover a recording or audio file from

Data Risk

Data Privacy:

Risk Level: High

Assessment: Since this is an attorney's work setting, sensitive and confidential information is exchanged during these conference calls. Even though transcripts or recordings aren't retained, the real-time processing poses a significant privacy risk.

Mitigations: Tokenize Sensitive Identifiers (Encrypt Sensitive Information (AML.M0012)); Contract with appropriate covenants; Use Ensemble Methods (AML.M0006)

Data Quality:

Risk Level: Medium

Assessment: The quality of the data depends on the clarity of the conversations and the accuracy of the AI in extracting points. Background noises, multiple speakers, and legal jargon might affect the quality.

Mitigations: User training to validate meeting summary.

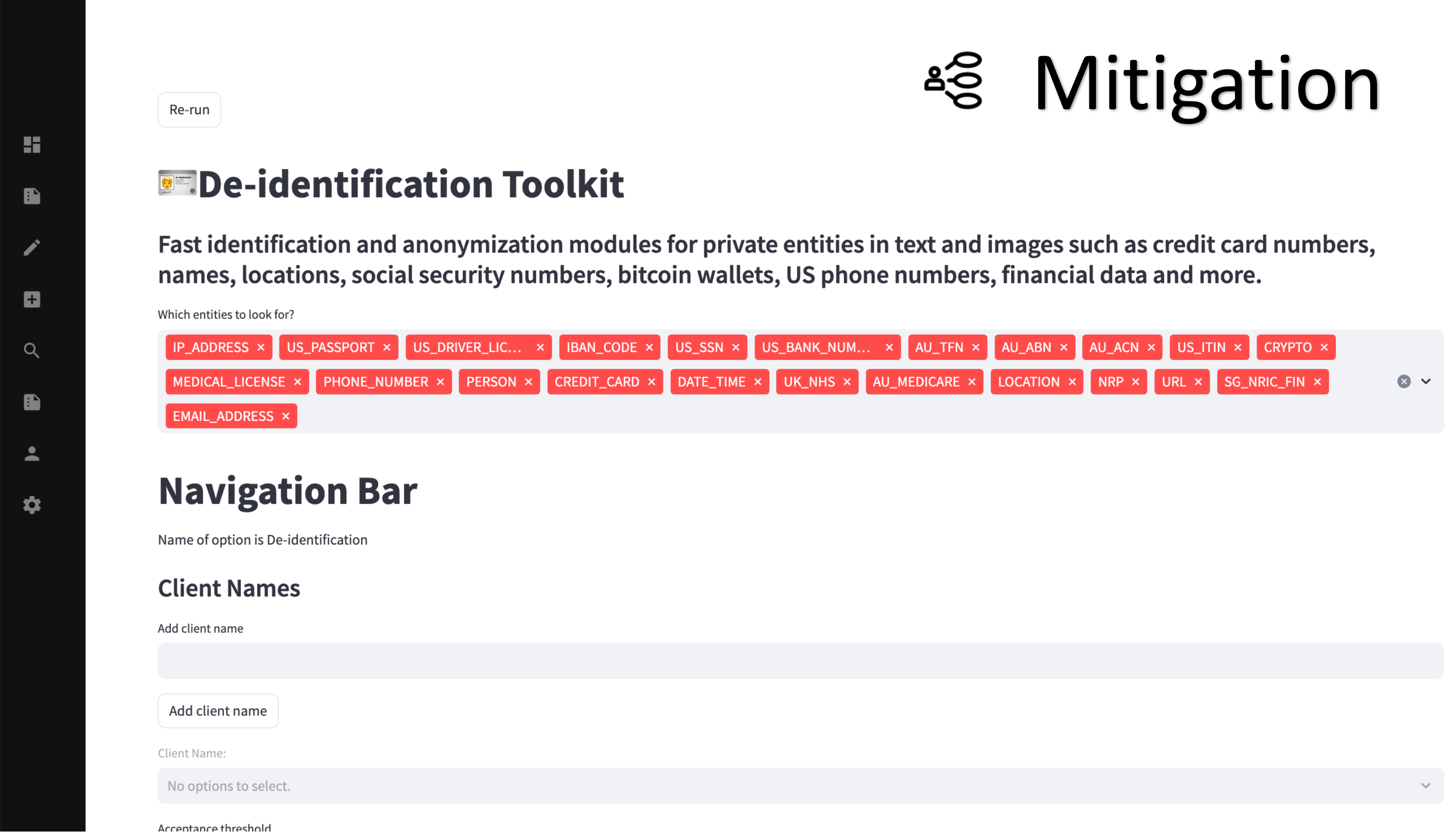
Data Security:

Risk Level: High

Assessment: Without proper security measures, there's a risk of data breaches during the real-time processing of calls, even if no data is written to disk.

Mitigations: Use Ensemble Methods (AML.M0006); Tokenize Sensitive Identifiers (Encrypt Sensitive Information (AML.M0012)); Contract with OpenAI

Mitigation



De-identification Toolkit

Fast identification and anonymization modules for private entities in text and images such as credit card numbers, names, locations, social security numbers, bitcoin wallets, US phone numbers, financial data and more.

Which entities to look for?

IP_ADDRESS × US_PASSPORT × US_DRIVER_LIC... × IBAN_CODE × US_SSN × US_BANK_NUM... × AU_TFN × AU_ABN × AU_ACN × US_ITIN × CRYPTO × MEDICAL_LICENSE × PHONE_NUMBER × PERSON × CREDIT_CARD × DATE_TIME × UK_NHS × AU_MEDICARE × LOCATION × NRP × URL × SG_NRIC_FIN × EMAIL_ADDRESS ×

Navigation Bar

Name of option is De-identification

Client Names

Add client name

Client Name:

No options to select.

Acceptance threshold

Model Risks

Bias and Fairness:

Risk Level: Low

Assessment: While the model may not inherently contain biases related to human attributes, there could be biases based on the language or terminology used, which might lead to incomplete or inaccurate extraction.

Robustness:

Risk Level: Medium

Assessment: Legal jargon, overlapping conversations, or unclear audio can challenge the model's robustness.

Transparency and Explainability:

Risk Level: Medium

Assessment: The attorney might need to understand why certain points were highlighted or missed to trust the AI system completely.

Unclear what ancillary sources were used to train the model.

Implementation Risks

Operational:

Risk Level: Medium

Assessment: Integrating the AI tool into the existing conference platform and ensuring it works in real-time without lags or disruptions.

Economic:

Risk Level: Low

Assessment: If the tool doesn't function as expected, there may be economic implications due to missed information or inaccurate legal risk documentation.

Societal:

Risk Level: Low

Assessment: Misunderstandings due to inaccurate AI outputs might lead to societal implications, especially if the legal risks involve public matters.

Regulatory and Compliance Risks

Legal Compliance:

Risk Level: High

Assessment: Compliance with attorney-client confidentiality laws and other relevant regulations is paramount.

Ethical Standards:

Risk Level: High

Assessment: Legal practices have strict ethical standards, especially concerning confidentiality. Any breaches or leaks can lead to severe ethical violations.

Stakeholder Risks

User Acceptance:

Risk Level: Medium

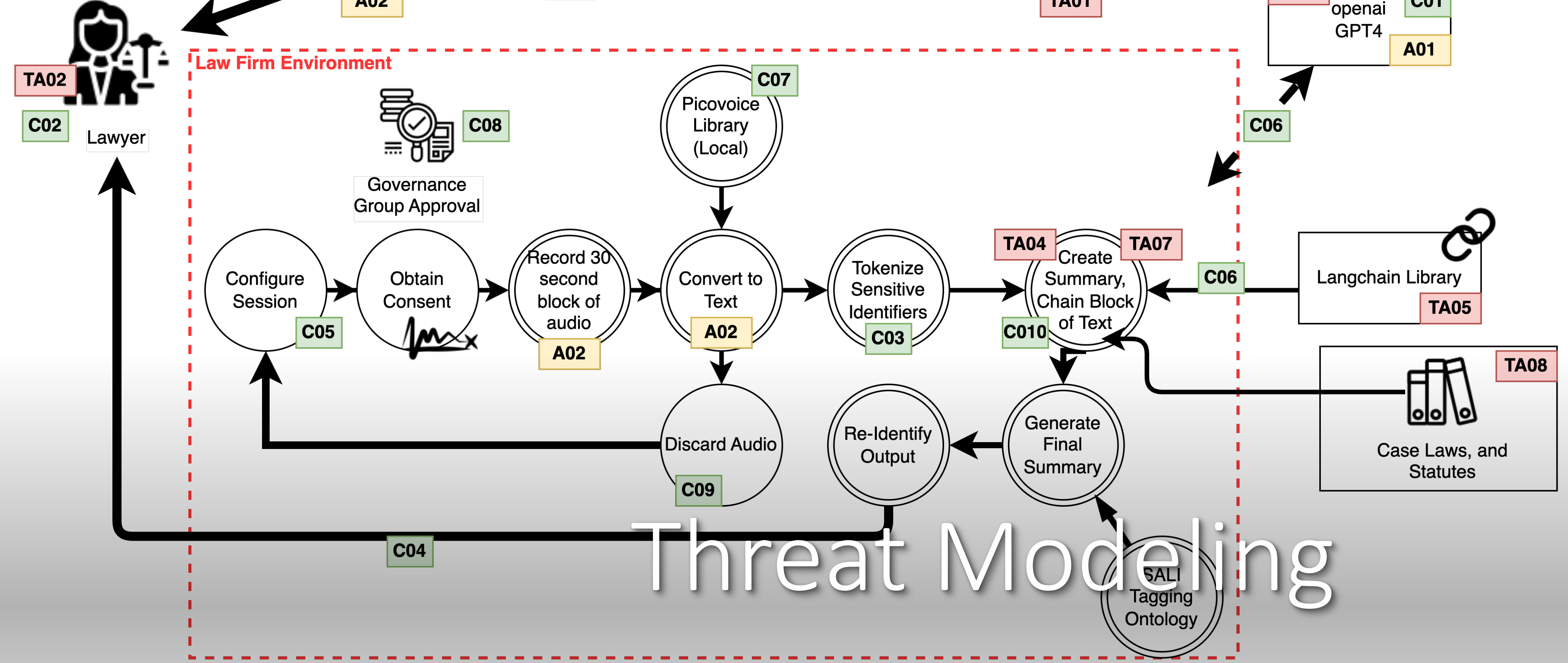
Assessment: The attorney might be skeptical about relying on an AI tool for capturing essential legal discussions initially.

Misuse:

Risk Level: Low

Assessment: There's a minimal risk of misuse since the tool is primarily for note-taking and doesn't retain data.

Proper training for the attorney on using the tool and continuous monitoring of its accuracy will also be crucial.



Threat Modeling

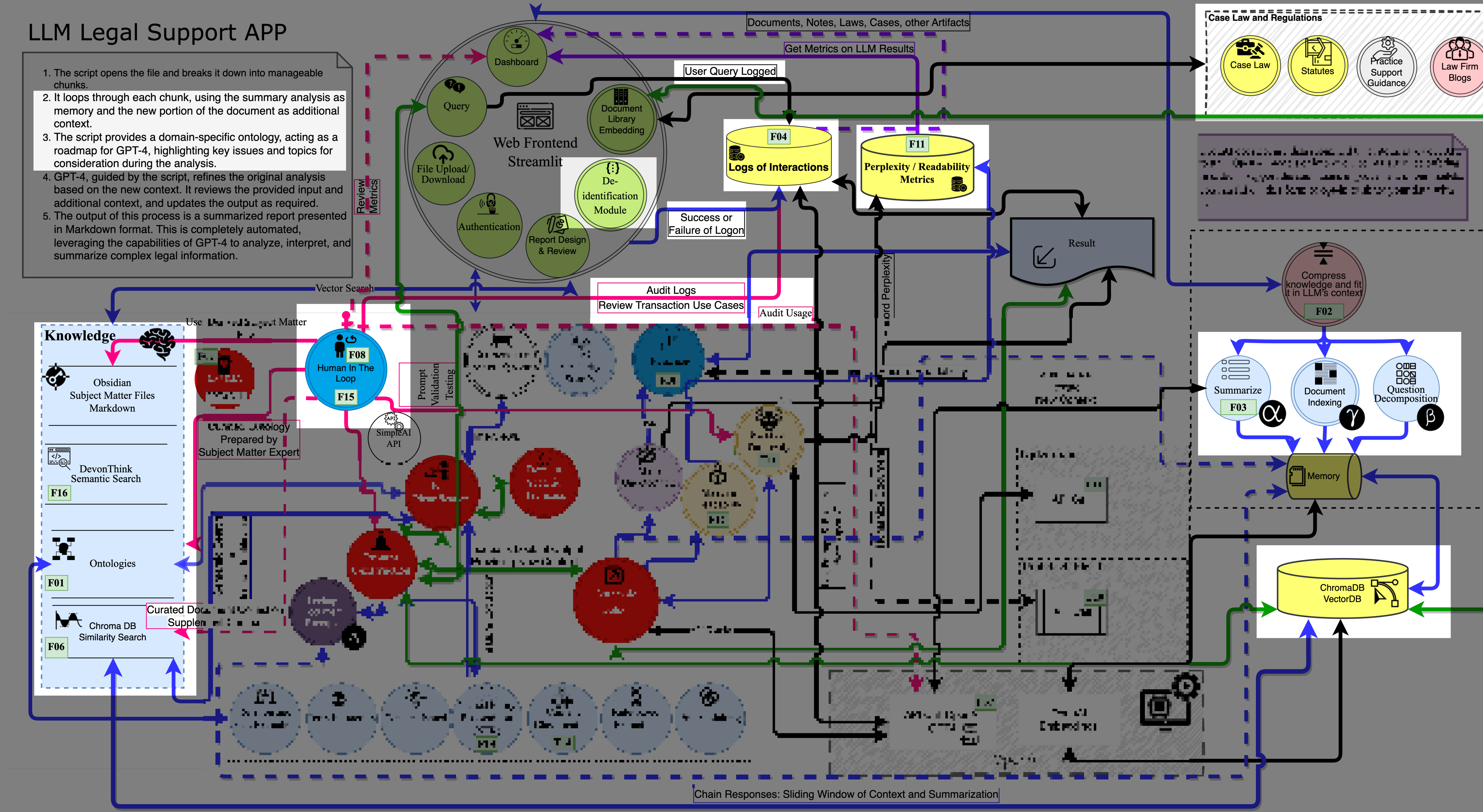
Assets	
ID	**Description**
A01	OpenAI API Key
A02	Client Confidential Information

Threat	
ID	**Description**
TA01	Hacker Steals API Key
TA02	Attorney Relies on Summary w/out Reviewing Resulting in Malpractice (OWASP LLM09)
TA03	Recording w/out Consent (single party consent states)
TA04	Prompt Modified: Malicious / Unintentional (OWASP LLM01)
TA04	Sensitive Information Disclosure (OWASP LLM06)
TA05	Supply Chain Compromise (OWASP LLM05)
TA06	Non-Compliance w/ Retainer Agreement Prohibits Use of LLMs
TA07	Hallucination
TA08	Copyright/ Terms of Use Violations

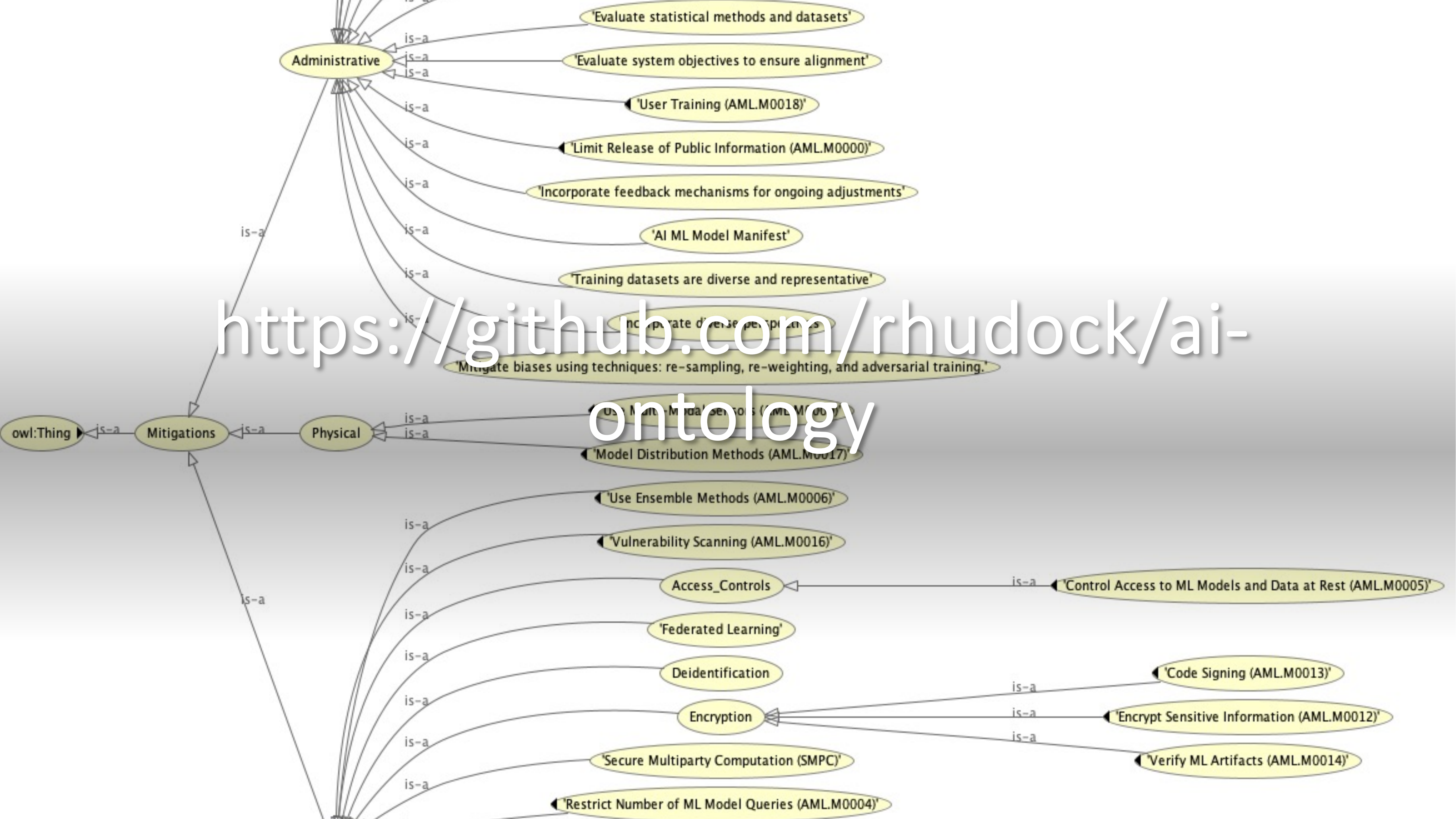
Security Controls	
ID	**Description**
C01	Contract with OpenAI Protecting Privacy
C02	User Training To Check Results
C03	Tokenize Sensitive Identifiers
C04	Human in The Loop
C05	Consent of Client to Use Tool
C06	SSL/ TLS Encryption
C07	Use Local Speech to Text Library
C08	Firm Governance Policy, Requires Risk Assessment of LLM Projects
C09	Data Minimization
C10	Ontology Describing Domain of Subject Matter / Domain of Conversation
C11	Audit Logs

LLM Legal Support APP

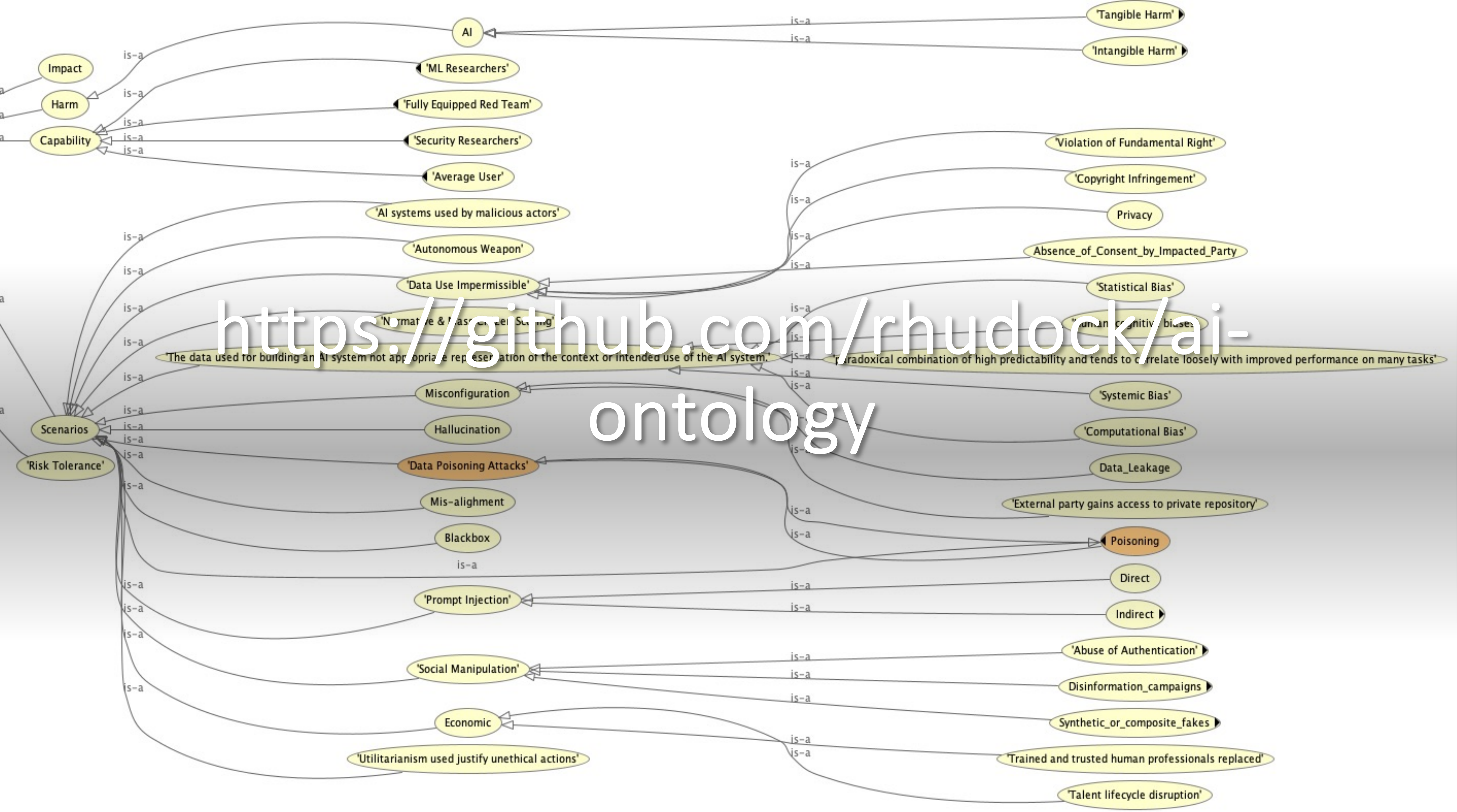
1. The script opens the file and breaks it down into manageable chunks.
2. It loops through each chunk, using the summary analysis as memory and the new portion of the document as additional context.
3. The script provides a domain-specific ontology, acting as a roadmap for GPT-4, highlighting key issues and topics for consideration during the analysis.
4. GPT-4, guided by the script, refines the original analysis based on the new context. It reviews the provided input and additional context, and updates the output as required.
5. The output of this process is a summarized report presented in Markdown format. This is completely automated, leveraging the capabilities of GPT-4 to analyze, interpret, and summarize complex legal information.


Documents, Notes, Laws, Cases, other Artifacts
Get Metrics on LLM Results
User Query Logged
Dashboard
Query
Document Library Embedding
Web Frontend Streamlit
File Upload/ Download
De-identification Module
Authentication
Report Design & Review
Review Metrics
F04
Logs of Interactions
F11
Perplexity / Readability Metrics
Success or Failure of Logon
Result
Case Law and Regulations
Case Law
Statutes
Practice Support Guidance
Law Firm Blogs
Compress knowledge and fit it in LLM's context
F02
Summarize
F03
α
Document Indexing
γ
Question Decomposition
β
Memory
ChromaDB VectorDB
Vector Search
Audit Logs
Review Transaction Use Cases
Audit Usage
Use
Matter
Knowledge
Obsidian Subject Matter Files Markdown
DevonThink Semantic Search
F16
Ontologies
F01
Chroma DB Similarity Search
F06
Curated Doc
Supplem
F08
Human In The Loop
F15
Prompt Validation Testing
SimpleAI API
Ontology Prepared by Subject Matter Expert
ord Perplexity
Chain Responses: Sliding Window of Context and Summarization

https://github.com/rhudock/ai-ontology


'Evaluate statistical methods and datasets'
Administrative
'Evaluate system objectives to ensure alignment'
'User Training (AML.M0018)'
'Limit Release of Public Information (AML.M0000)'
'Incorporate feedback mechanisms for ongoing adjustments'
'AI ML Model Manifest'
'Training datasets are diverse and representative'
'Incorporate diverse perspectives'
'Mitigate biases using techniques: re-sampling, re-weighting, and adversarial training.'
owl:Thing
Mitigations
Physical
'Use Multi-Modal Sensors (AML.M00[illegible])'
'Model Distribution Methods (AML.M0017)'
'Use Ensemble Methods (AML.M0006)'
'Vulnerability Scanning (AML.M0016)'
Access_Controls
'Control Access to ML Models and Data at Rest (AML.M0005)'
'Federated Learning'
Deidentification
Encryption
'Code Signing (AML.M0013)'
'Encrypt Sensitive Information (AML.M0012)'
'Verify ML Artifacts (AML.M0014)'
'Secure Multiparty Computation (SMPC)'
'Restrict Number of ML Model Queries (AML.M0004)'
is-a

https://github.com/rhudock/ai-ontology


AI
'Tangible Harm'
'Intangible Harm'
Impact
Harm
Capability
'ML Researchers'
'Fully Equipped Red Team'
'Security Researchers'
'Average User'
'AI systems used by malicious actors'
'Autonomous Weapon'
'Data Use Impermissible'
'Violation of Fundamental Right'
'Copyright Infringement'
Privacy
Absence_of_Consent_by_Impacted_Party
'Statistical Bias'
'The data used for building an AI system not appropriate representation of the context or intended use of the AI system.'
'paradoxical combination of high predictability and tends to correlate loosely with improved performance on many tasks'
Misconfiguration
'Systemic Bias'
Scenarios
Hallucination
'Computational Bias'
'Risk Tolerance'
'Data Poisoning Attacks'
Data_Leakage
Mis-alighment
'External party gains access to private repository'
Blackbox
Poisoning
Direct
'Prompt Injection'
Indirect
'Abuse of Authentication'
'Social Manipulation'
Disinformation_campaigns
Synthetic_or_composite_fakes
Economic
'Trained and trusted human professionals replaced'
'Utilitarianism used justify unethical actions'
'Talent lifecycle disruption'
is-a

https://github.com/rhudock/ai-ontology

Any Questions