


Article

# Identifying Key Issues in Artificial Intelligence Litigation: A Machine Learning Text Analytic Approach

Wullianallur Raghupathi \*, Aditya Saharia and Tanush Kulkarni 

Gabelli School of Business, Fordham University, 140 W. 62nd Street, New York, NY 10023, USA; saharia@fordham.edu (A.S.); kultanush@gmail.com (T.K.)

\* Correspondence: raghupathi@fordham.edu

## Abstract

The rapid proliferation of artificial intelligence (AI) systems across high-stakes domains—with global AI adoption accelerating since 2023—has created an urgent need to identify which AI challenges and issues are materializing into real-world harms so that policymakers can develop targeted regulations, organizations can implement effective risk management, and accountability mechanisms can address actual rather than speculative problems. Public concern has risen sharply: 52% of Americans now feel more concerned than excited about AI (up from 38% in 2022), and 80% believe government should maintain AI safety rules even if development slows. Yet existing approaches exhibit critical limitations that impede evidence-based governance. Ethics frameworks, while establishing normative principles across 84+ published guidelines, remain aspirational rather than empirical. Survey-based studies capture perceptions from over 48,000 respondents globally but measure expectations rather than documented harms. Incident databases catalog over 1200 AI failures but depend on media coverage, systematically overrepresenting high-profile cases while underrepresenting routine organizational problems. This study addresses this gap by analyzing 347 AI-related U.S. litigation cases using machine learning text analytics, providing empirical evidence of AI problems that have crossed the threshold from abstract concern into documented legal conflict. Employing Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF) topic modeling with coherence validation (NMF achieving 0.276 NPMI vs. LDA's 0.164), the analysis identifies nine distinct AI issue areas with specific case distributions: cybersecurity vulnerabilities and data breaches (116 cases, 33.4%), intellectual property and AI ownership (61 cases, 17.6%), AI misrepresentation and inflated claims (59 cases, 17.0%), criminal justice and algorithmic due process (37 cases, 10.7%), employment automation (33 cases, 9.5%), privacy and surveillance (31 cases, 8.9%), platform accountability (21 cases, 6.1%), algorithmic bias (19 cases, 5.5%), and government AI deployment (6 cases, 1.7%). The findings reveal a systematic mismatch between AI ethics discourse—which emphasizes fairness and transparency—and litigation patterns, where data security (33.4%) and intellectual property (17.6%) dominate while algorithmic bias comprises only 5.5% of cases. Most disputes are addressed through existing legal frameworks (First Amendment, Lanham Act, FOIA, Title VII) rather than AI-specific regulation, underscoring the urgent need for governance mechanisms aligned with empirically documented AI challenges.



Academic Editor: Stefan Fischer

Received: 11 November 2025

Revised: 17 December 2025

Accepted: 22 December 2025

Published: 25 December 2025

**Copyright:** © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)[Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).**Keywords:** artificial intelligence; AI litigation; machine learning; text analytics; topic modeling; LDA; NMF; AI governance; algorithmic accountability; cybersecurity; intellectual property; algorithmic bias; AI ethics; legal text analysis

## 1. Introduction

The rapid deployment of artificial intelligence (AI) systems across high-stakes domains including employment, criminal justice, healthcare, finance, and digital platforms has outpaced the development of mechanisms capable of systematically identifying and governing the problems that arise in practice. While growing literature addresses AI ethics, fairness principles, and legal accountability [1,2], much of this work remains normative or anticipatory. As a result, there is limited empirical evidence regarding which AI issues and challenges are materializing into consequential real-world problems, as opposed to those that remain speculative or aspirational concerns.

This study focuses on AI issues and challenges, not legal doctrine. Litigation is used strictly as an empirical lens rather than as the object of analysis. Litigation records provide a distinctive and high-stakes source of evidence in which AI-related problems have progressed beyond abstract concern into documented conflict. Each case reflects a situation in which the design, deployment, or representation of an AI system produced outcomes sufficiently consequential to trigger formal dispute. Accordingly, this study does not attempt to catalog all potential AI risks. Instead, it is intentionally bounded to AI challenges that have generated documented litigation in U.S. courts, thereby prioritizing empirical precision over conceptual breadth.

Existing approaches to identifying AI issues exhibit important limitations. Ethics frameworks articulate normative aspirations but do not indicate which problems recur during real-world deployment [1]. Survey-based studies capture perceptions and expectations rather than documented harms [3,4]. Incident databases depend heavily on media reporting and voluntary disclosure, which systematically overrepresent high-profile failures while underrepresenting routine but consequential AI problems embedded in organizational processes [5–7]. Consequently, policymakers and researchers lack empirical insight into which AI challenges repeatedly cross a threshold of real-world consequence. Litigation data addresses this gap by capturing AI failures that have already produced sustained, adversarial conflict.

The necessity of this research is amplified by recent developments. Since 2023, the widespread adoption of generative and automated AI systems has accelerated rapidly, while regulatory and institutional responses remain fragmented and evolving. Early disputes concerning AI failures are already shaping accountability practices in courts and organizations. Absent systematic empirical analysis, AI governance debates risk being driven by perception, principle, or media salience rather than documented patterns of harm. This makes litigation-based analysis both timely and necessary.

Three research questions guide this study. First, which categories of AI issues and challenges most frequently generate real-world conflict? Second, what latent thematic structures emerge when machine-learning methods are applied to litigation text, and how do these structures relate to substantive AI challenges? Third, how do empirically identified AI challenges map onto existing accountability and governance mechanisms? Each research question is addressed by a distinct analytical component. Frequency analysis, co-occurrence networks, and manual categorization address the first question; unsupervised topic modeling using Latent Dirichlet Allocation (LDA) [5] and Non-negative Matrix Factorization (NMF) [6] addresses the second; and cross-model interpretation and institutional mapping address the third. The study analyzes 347 AI-related U.S. litigation cases to ensure internal methodological coherence across questions and methods.

This study makes four bounded contributions. First, it develops an empirically grounded taxonomy of AI issues and challenges derived from documented real-world conflict rather than normative principle or hypothetical risk. Second, it demonstrates that unsupervised machine-learning methods applied to litigation text can recover substantively

meaningful AI issue structures while revealing procedural dimensions often absent from ethics-based mappings. Third, it shows that many AI challenges are currently addressed through adaptations of existing accountability mechanisms rather than AI-specific governance tools. Fourth, it highlights a systematic mismatch between dominant AI ethics discourse and the AI challenges most frequently implicated in practice. These contributions are intentionally limited to AI-related litigation in the United States through 2024 and should be interpreted within that empirical scope.

The remainder of this paper proceeds as follows. Section 2 reviews existing approaches to identifying AI issues, comparing methodological strengths and limitations across ethics frameworks, surveys, audits, incident databases, and computational text analytics. Section 3 details the data collection, preprocessing, and machine learning methods employed. Section 4 presents the empirical results, including the nine AI issue areas identified and the hierarchical topic structure revealed by LDA and NMF analysis. Section 5 interprets these findings and discusses their implications for AI governance. Section 6 acknowledges limitations and suggests directions for future research. Section 7 concludes with a synthesis of contributions and practical implications.

## 2. Literature Review

Research on identifying artificial intelligence issues has proliferated across multiple disciplines, employing diverse methodological approaches that yield complementary but distinct insights. This literature review synthesizes scholarship on AI ethics, algorithmic bias, and societal impacts through a comparative lens, examining how different methodologies—survey research, systematic literature reviews, computational text analytics, incident databases, empirical audits, and litigation analysis—surface different dimensions of AI challenges. By comparing methodological strengths, limitations, and findings, this review establishes the scholarly context for using litigation data as a novel window into real-world AI issues [7–9].

The organization of this review reflects a methodological rather than topical structure. Rather than cataloging AI issues domain by domain, we examine how researchers have approached the fundamental question: What are the key contemporary challenges posed by AI systems, and how do we know? This comparative framework reveals that methodological choices shape which AI issues become visible, how they are characterized, and what populations' experiences are captured.

### 2.1. Systematic Literature Reviews of AI Ethics Principles

A substantial body of research has sought to map the landscape of AI ethics through systematic analysis of principles, guidelines, and ethical frameworks published by governments, corporations, and research institutions.

The foundational study in this area is Jobin, Ienca, and Vayena's [1] analysis of 84 AI ethics documents published by private companies, research institutions, and public sector organizations worldwide. Using scoping review methodology adapted from Arksey and O'Malley [10], the researchers identified eleven recurring ethical themes and found global convergence around five core principles: transparency, justice and fairness, non-maleficence, responsibility, and privacy. However, their in-depth thematic analysis revealed substantial divergence in how these principles are interpreted, why they are deemed important, what domains they address, and how they should be implemented. This finding—apparent consensus masking fundamental disagreement—has shaped subsequent scholarship on the limitations of principles-based AI ethics [11].

Methodologically, Jobin et al. relied on document analysis of publicly available ethics guidelines, which captures normative aspirations but not implementation realities. Their

geographic analysis revealed that 88% of documents were published after 2016, with private companies (22.6%) and governmental agencies (21.4%) as the primary issuers. This distribution suggests the limitations of relying solely on official guidelines: they represent institutional positions rather than lived experiences of AI impacts.

## 2.2. Survey-Based Approaches to AI Issues

Survey research represents the dominant empirical methodology for identifying AI issues through public and expert opinion data. Prior research on identifying key AI issues has relied on public opinion surveys and expert elicitation.

The Pew Research Center has conducted the most methodologically rigorous U.S.-specific studies, surveying 5410 adults via the American Trends Panel using address-based sampling with a 93% response rate [3]. Key findings reveal that 52% of Americans feel more concerned than excited about AI—up from 38% in 2022—with dominant concerns centered on employment displacement, privacy erosion, and algorithmic decision-making in high-stakes contexts. Notably, a significant expert-public gap exists: 56% of AI experts foresee positive societal impacts versus far fewer among the public.

International surveys have extended this research globally. The Stanford Human-Centered AI Institute's annual AI Index Report synthesizes multiple survey sources to track global public opinion on AI. The 2024 and 2025 reports [4] draw on Ipsos surveys of over 23,000 adults across 32 countries, revealing significant regional variation in perceived AI challenges. Key findings demonstrate that 83% of respondents in China and 80% in Indonesia view AI benefits as outweighing drawbacks, compared to only 39% in the United States and 36% in the Netherlands [4]. Global nervousness about AI rose 13 percentage points from 2022 to 2024, reaching 52% [12]. These cross-national comparisons reveal how AI issues are perceived differently across cultural and economic contexts.

The University of Toronto's Global Public Opinion on Artificial Intelligence survey covered 23,882 respondents across 21 countries, while the Melbourne Business School study surveyed over 48,000 people in 47 countries, finding that while 66% use AI regularly, only 46% trust it [13,14]. The Brookings/GRAIL AI SHARE database represents a significant meta-analytic effort, aggregating approximately 1800 survey questions from 218 studies conducted between 2014 and 2023 [15]. This synthesis revealed consistent patterns: broad public support for AI regulation spanning political affiliations, persistent concerns about employment and privacy, and limited trust in either technology companies or governments to implement oversight effectively. The Gallup/SCSP survey found that 80% believe government should maintain AI safety rules even if development slows, and 97% agree AI should be regulated [16].

Survey approaches offer important advantages: scale, statistical representativeness when professionally designed, and the ability to capture public attitudes systematically [17,18]. However, they measure perceptions rather than actual harm, and response options can constrain which issues surface. Critically, surveys capture what people think about AI rather than documenting how AI systems cause problems in practice.

## 2.3. Empirical Audits and Bias Studies

A distinct research tradition has employed empirical audits to identify AI issues through direct testing of algorithmic systems. This approach surfaces concrete harms that may not appear in public opinion or ethics guidelines.

The seminal work in this area is Buolamwini and Gebre's [19] Gender Shades study, which audited commercial facial recognition systems and found error rates of 0.8% for lighter-skinned males versus 34.7% for darker-skinned females. This study catalyzed both academic research and corporate reforms by demonstrating that AI sys-

tems can systematically underperform for marginalized populations [20]. The audit methodology—testing actual systems against diverse benchmarks—provides evidence of real-world harm that principles-based approaches cannot capture.

ProPublica's investigation of the COMPAS recidivism prediction instrument represents investigative journalism adopting audit methodology [21]. The analysis found that Black defendants were twice as likely to be falsely flagged as high-risk compared to white defendants. Subsequent academic work by Chouldechova [22] and Kleinberg et al. [23] formalized the mathematical impossibility of simultaneously satisfying multiple fairness metrics when base rates differ across groups, revealing that algorithmic bias reflects fundamental tensions rather than mere technical failures.

Healthcare AI has received substantial audit attention. Obermeyer et al.'s [24] study in *Science* found that a widely used healthcare algorithm systematically underestimated illness severity for Black patients because it used healthcare costs as a proxy for need—and Black patients historically receive less care at equivalent illness levels. This finding illustrates how bias can emerge not from discriminatory intent but from proxy variables that encode historical inequities.

The audit tradition provides rigorous evidence of specific AI failures but faces scalability limitations [15,25]. Each audit requires substantial resources to conduct, and findings may not generalize across systems. The approach excels at documenting particular harm but cannot provide comprehensive mapping of the AI issue landscape.

#### 2.4. AI Incident Databases

The emergence of AI incident databases represents a systematic attempt to catalog documented AI failures. The AI Incident Database (AIID), launched in 2020, now contains over 1200 documented incidents of AI systems causing harm [26,27]. Incidents are classified by affected parties, harm types, and AI technologies involved, creating a structured repository of AI failures.

The MIT AI Risks Repository provides a complementary classification system organized around technical failure modes and social impacts [28]. The OECD's AI Incidents and Hazards Monitor offers international coverage with standardized taxonomy aligned with the OECD AI Principles [29]. Collectively, these databases enable researchers to identify patterns across incidents and track how AI harms evolve over time.

Incident databases offer structured documentation but depend on media coverage and voluntary reporting. High-profile failures involving major technology companies are well-documented, while harms affecting less visible populations may go unreported. The databases also tend to capture discrete incidents rather than systematic patterns of harm.

#### 2.5. Computational Text Analytics Applied to AI Discourse

Computational approaches have been employed to analyze large-scale text corpora related to AI, though primarily focused on news media, social media, and research literature rather than legal documents.

The Stanford HAI partnership with Quid analyzed 6.69 million English posts from Reddit, Twitter, and news sources from 2016 to 2024, revealing temporal evolution of AI discourse topics and sentiment patterns [30,31]. The analysis found increased attention to AI safety, job displacement, and regulation following major AI releases, with sentiment becoming more negative over time. Topic modeling revealed distinct discourse communities with different AI concern profiles.

The European Union Fundamental Rights Agency [32] analyzed media coverage of AI bias across member states, finding systematic underreporting of algorithmic discrimination despite documented incidents. This meta-analysis of coverage patterns revealed that media

attention does not proportionally reflect documented harms, suggesting that media-based approaches may systematically undercount certain AI issues.

### 2.6. Domain-Specific AI Issue Studies

Substantial research has examined AI issues within specific domains. Healthcare AI has attracted particular attention: a JAMA study surveying 13,806 patients found concerns about privacy, diagnostic errors, and reduced human contact [33], while a Lancet Digital Health systematic review synthesized attitudes across 48 studies [34]. Employment AI issues have been documented through both investigative journalism—such as Reuters’ reporting on Amazon’s abandoned AI recruiting tool that discriminated against women—and academic research examining algorithmic hiring practices [35,36]. Table 1 summarizes the methodological approaches to identifying AI issues.

**Table 1.** Comparison of Methodological Approaches to Identifying AI Issues.

Methodology	Data Source	Strengths	Limitations	AI Issues Surfaced
Systematic Reviews	Ethics guidelines	Captures normative consensus	Aspirational, not empirical	Transparency, fairness, privacy principles
Surveys	Public/expert opinion	Scale, representativeness	Perceptions vs. actual harms	Employment, privacy concerns
Audits	Direct system testing	Concrete evidence of harm	Resource-intensive, limited scope	Bias in specific systems
Incident Databases	Reported failures	Structured documentation	Reporting bias	High-profile AI failures
Text Analytics	Media, social media	Large-scale patterns	Coverage bias	Discourse trends
Litigation Analysis	Court cases	Documented real harms	Access to justice barriers	Actionable AI issues

### 2.7. Research Gaps and Positioning

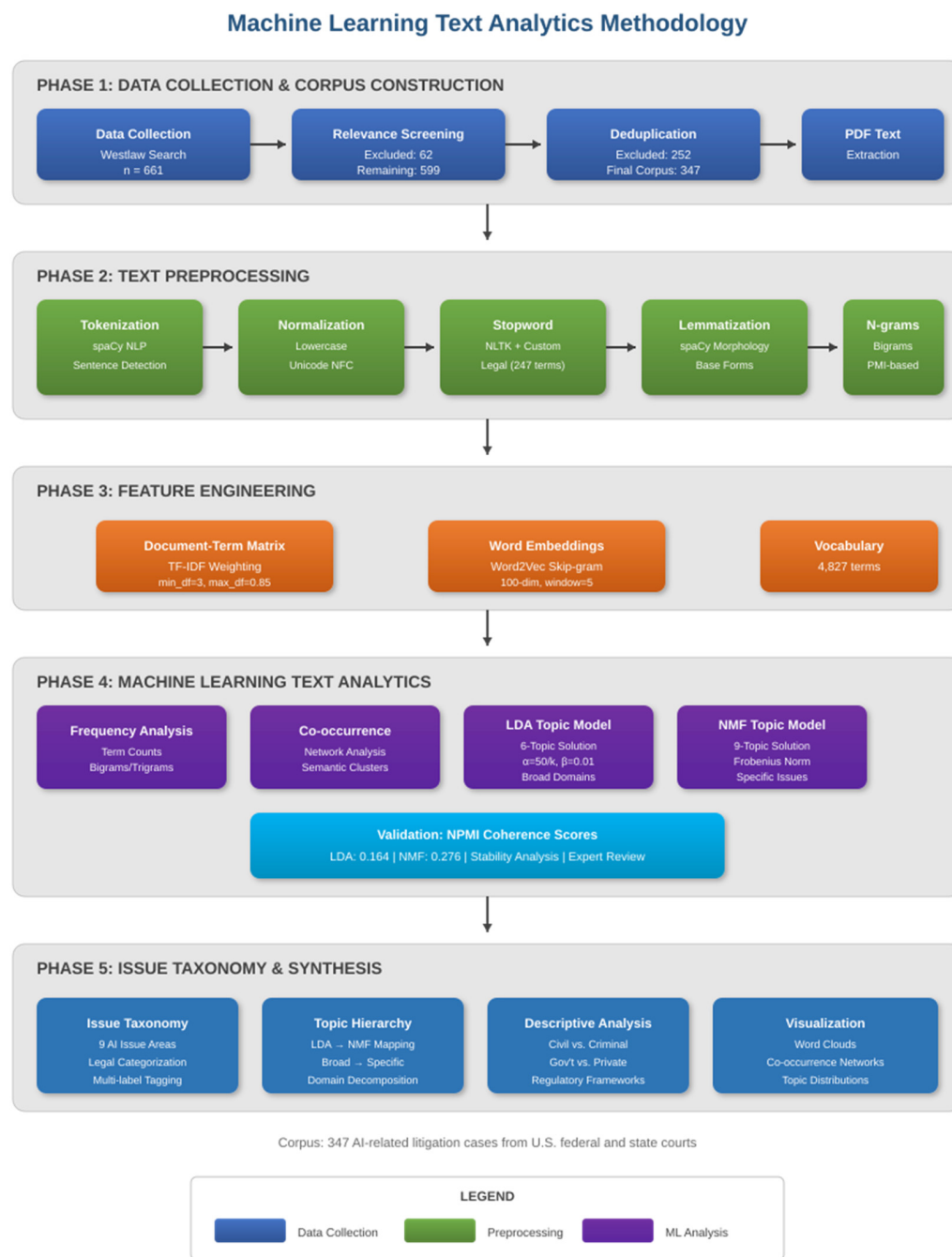
This comparative analysis reveals several gaps that motivate the present study. First, most approaches rely on opinion data (surveys), normative documents (ethics guidelines), or media coverage rather than documented instances of AI causing consequential harm. Second, computational text analytics has been applied extensively to news and social media but rarely to legal documents, despite litigation representing a particularly authoritative source of documented AI problems. Third, existing approaches tend to capture either high-level principles or specific incidents, with limited ability to map the full landscape of AI issues at intermediate granularity.

Litigation data addresses these gaps by providing documented cases where AI issues became serious enough to trigger formal action—neither speculative concerns nor isolated incidents, but recurring patterns of real-world problems. The present study applies machine learning text analytics to this underexplored data source, contributing both substantive findings about contemporary AI issues and methodological demonstration of litigation-based text analytics.

## 3. Data and Methods

This section details the methodological framework employed to identify AI issues from litigation data. We describe the rationale for using litigation as a data source, the corpus construction process, text preprocessing procedures, and the machine learning text analytics methods applied. Each methodological choice is explained with reference to its

appropriateness for legal text analysis and its contribution to surfacing AI issues. Figure 1 outlines the methodology for this study.



**Figure 1.** End-to-end methodology flowchart for machine learning text analytics of AI litigation corpus.

### 3.1. Litigation as a Data Source for AI Issues

This study uses litigation cases as an unconventional but powerful data source for identifying AI issues. Unlike surveys, expert panels, or media analysis, litigation data reflects situations where AI-related problems became consequential enough to trigger formal action. Each case represents a crystallized instance of an AI issue—whether involving algorithmic decision-making, data practices, automated systems, or AI-enabled products and services. The corpus thus provides an empirical foundation for understanding which AI challenges are most salient in practice.

Legal documents possess several characteristics that make them particularly suitable for text analytics. First, they are produced under institutional constraints that promote precision and completeness: attorneys must articulate claims clearly to succeed, and courts require detailed factual recitations. Second, legal language, while specialized, follows conventions that facilitate systematic analysis. Third, the adversarial nature of litigation surfaces multiple perspectives on AI issues, as plaintiffs articulate harms while defendants present counterarguments. Fourth, the formal record-keeping of the judicial system ensures comprehensive documentation unavailable in other contexts [37].

The choice to analyze litigation data complements rather than replaces other methodological approaches. Surveys capture public perceptions; audits document technical failures; incident databases catalog reported harms. Litigation captures a distinct phenomenon: AI issues that have crossed a threshold of consequentiality sufficient to motivate formal legal action. This threshold filters out speculative concerns while retaining documented real-world impacts.

### 3.2. Corpus Construction

The initial dataset comprised 661 case documents retrieved from Westlaw using queries related to artificial intelligence and machine learning. Search terms included ‘artificial intelligence’, ‘machine learning’, ‘algorithm’, ‘automated decision’, and related terminology. The search encompassed federal and state courts across the United States, including district courts, appellate courts, and specialized tribunals. The Westlaw database was utilized ([www.westlaw.com](http://www.westlaw.com)). While several independent searches were conducted and repeated, the most recent was carried out on 2 December 2025.

Following relevance screening, 62 cases were excluded as not substantively related to AI. These exclusions included cases where AI terminology appeared only in passing reference, cases involving technologies not meeting standard AI definitions, and cases where automated systems did not involve learning or adaptive components. Deduplication procedures removed an additional 252 duplicate or near-duplicate records arising from multiple filings in the same case, appeals of the same underlying dispute, and parallel state and federal proceedings. The final analytic corpus consisted of 347 unique AI-related cases, forming the basis for all subsequent analysis.

### 3.3. Text Preprocessing Pipeline

Raw legal documents require substantial preprocessing before machine learning analysis. Legal texts contain formatting artifacts, citation conventions, procedural boilerplate, and domain-specific vocabulary that must be managed appropriately. Our preprocessing pipeline was implemented in Python programming language using established NLP libraries and followed best practices for legal text analysis [38].

**Tokenization.** Documents were first tokenized—segmented into individual words and phrases—using the spaCy natural language processing library [39]. Tokenization decisions affect downstream analysis: we preserved hyphenated terms (e.g., ‘machine-learning’) as single tokens and managed legal citation formats (e.g., ‘42 U.S.C. § 1983’) to prevent fragmentation of meaningful units. Sentence boundary detection employed spaCy’s statistical model trained on legal and general English text.

**Case Normalization.** All text was converted to lowercase to ensure that ‘Algorithm’, ‘ALGORITHM’, and ‘algorithm’ were treated as the same term. While case sensitivity can carry meaning in some contexts (e.g., proper nouns), the benefits of normalization for frequency analysis and topic modeling outweigh the minimal information loss in this application.

**Stopword Removal.** Stopwords—high-frequency words with limited semantic content—were removed using a two-tier approach. First, we applied the standard English stopword list from the Natural Language Toolkit (NLTK), which includes articles, prepositions, and common verbs (e.g., ‘the,’ ‘of,’ ‘is,’ ‘have’). Second, we developed a custom legal stopword list to remove procedural terminology that appears frequently across all legal documents but carries minimal discriminative value for identifying AI issues. This custom list included terms such as ‘plaintiff,’ ‘defendant,’ ‘court,’ ‘motion,’ ‘order,’ ‘pursuant,’ ‘hereby,’ and similar procedural vocabulary. The combined stopword list contained 247 terms. Stopword removal reduces noise and allows substantively meaningful terms to emerge in frequency and topic analyses.

**Lemmatization.** Words were reduced to their base or dictionary forms through lemmatization using spaCy’s morphological analyzer. This process converts inflected forms to a common representation: ‘algorithms,’ ‘algorithmic,’ and ‘algorithmically’ map to ‘algorithm’; ‘discriminated,’ ‘discriminating,’ and ‘discrimination’ map to ‘discriminate’. Lemmatization consolidates related terms, increasing statistical power for detecting patterns while preserving more semantic information than the alternative stemming approach, which can produce non-word stems. For legal text, lemmatization appropriately manages the formal register and Latinate vocabulary common in judicial writing.

**N-gram Extraction.** Beyond individual words (unigrams), we extracted bigrams—two-word sequences—to capture multi-word concepts that lose meaning when separated. Legal and technical terminology frequently involves compound phrases: ‘facial recognition,’ ‘trade secret,’ ‘due process,’ ‘machine learning’. Bigram extraction employed collocation detection based on pointwise mutual information (PMI) scores, retaining word pairs that co-occur more frequently than expected by chance. The final vocabulary included both unigrams and high-PMI bigrams, enabling the analysis to capture both single-word and phrasal concepts.

### 3.4. Document-Term Matrix Construction

The preprocessed text was transformed into a document-term matrix (DTM) a numerical representation where rows correspond to documents (cases) and columns correspond to terms (words and bigrams). Each cell contains a value representing the term’s presence or importance in that document. The DTM provides the foundation for all subsequent quantitative analyses.

**Term Frequency (TF).** The baseline representation counts raw term occurrences within each document. A document mentioning ‘patent’ five times receives a value of 5 in the corresponding cell. Term frequency captures the prominence of concepts within individual documents.

**TF-IDF Weighting.** For analysis requiring discrimination between documents, we applied term frequency-inverse document frequency (TF-IDF) weighting. TF-IDF downweights terms that appear in many documents (reducing the influence of common legal vocabulary) while upweighting terms distinctive to documents or document subsets. The formula  $TF\text{-}IDF(t,d) = TF(t,d) \times \log(N/DF(t))$  multiplies term frequency by the logarithm of the total document count divided by the document frequency. This weighting scheme highlights terms that characterize specific AI issue areas rather than appearing generically across the corpus.

**Frequency Thresholds.** We applied minimum and maximum document frequency thresholds to filter the vocabulary. Terms appearing in fewer than 3 documents ( $\text{min\_df} = 3$ ) were excluded as too rare to support generalizable inference; terms appearing in more than 85% of documents ( $\text{max\_df} = 0.85$ ) were excluded as too common to discriminate between AI issues. These thresholds balance signal and noise, retaining terms frequent enough

for statistical reliability but discriminative enough for meaningful analysis. The resulting vocabulary contained 4827 unique terms.

### 3.5. Distributed Word Representations

While the document-term matrix captures term frequencies, it treats each term as independent, ignoring semantic relationships. We supplemented frequency-based analysis with distributed word representations (word embeddings) that encode semantic similarity in dense vector spaces.

**Word2Vec Embeddings.** We trained Word2Vec models on the corpus using the skip-gram architecture with negative sampling [40]. Word2Vec learns vector representations by predicting context words from target words, resulting in vectors where semantically similar terms (e.g., ‘privacy’ and ‘surveillance’) have high cosine similarity. We trained 100-dimensional vectors with a context window of 5 words and minimum word frequency of 5. These embeddings informed the co-occurrence network analysis by identifying semantically related term clusters beyond simple co-occurrence.

**Domain Adaptation.** General-purpose word embeddings trained on news or web corpora may not capture legal-specific meanings. For example, ‘party’ in legal text refers to litigants, not social gatherings. Training embeddings directly on the litigation corpus produces representations adapted to legal AI discourse, where ‘algorithm’ clusters with ‘bias’, ‘prediction’, and ‘decision’ rather than general computing terms.

### 3.6. Term Frequency and Co-Occurrence Analysis

Frequency analysis provides the foundation for understanding the vocabulary of AI litigation. We computed term frequencies across the entire corpus to identify the most prominent concepts, revealing the linguistic landscape through which AI issues are articulated in legal contexts.

**Rationale.** High-frequency terms indicate concepts central to AI litigation discourse. While simple, frequency analysis reveals which AI concepts generate the most legal attention—information not available from ethics guidelines or survey research. Terms like ‘data’, ‘privacy’, ‘patent’, and ‘discrimination’ emerge not because experts deemed them important but because litigants and courts invoked them repeatedly.

Co-occurrence analysis extends frequency analysis to examine which terms appear together within documents. When ‘facial’ and ‘recognition’ frequently co-occur, they refer to the same underlying concept. More interestingly, when ‘algorithm’ co-occurs with both ‘discrimination’ and ‘patent’ in different documents, it reveals that algorithmic issues span both civil rights and intellectual property domains.

**Network Visualization.** We constructed a term co-occurrence network where nodes represent terms and edges connect terms appearing in the same documents. Edge weights reflect co-occurrence frequency, and network visualization algorithms position densely connected terms closer together. The resulting network reveals clusters of related AI concepts and bridges between issue areas, providing a map of how AI challenges interconnect in legal discourse.

### 3.7. Topic Modeling: Rationale and Method Selection

Topic modeling algorithms discover latent thematic structures in document collections without requiring predefined categories. Rather than classifying documents into researcher-specified bins, topic models allow themes to emerge from the data—an approach well-suited to exploratory analysis of an evolving domain like AI litigation.

**Why Topic Modeling for Legal AI Text?** Legal documents are lengthy, complex texts that resist simple keyword classification. A particular case may address multiple AI issues—intellectual property, privacy, and procedural matters—in different sections. Topic

models accommodate this complexity by representing documents as mixtures of topics rather than forcing single-category assignment. This probabilistic representation [41] better captures the multi-faceted nature of AI litigation than deterministic classification.

We applied two complementary topic modeling algorithms: Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF). Each algorithm rests on different mathematical foundations and produces diverse types of topics, and their complementary application strengthens confidence in findings that emerge consistently across methods.

### 3.8. Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation is a generative probabilistic model that assumes documents are produced by a two-stage process: first selecting a mixture of topics, then selecting words from those topics [5,42]. LDA has become the standard approach for topic modeling in computational social science and digital humanities.

**Model Specification.** LDA requires specification of the number of topics ( $k$ ) as a hyperparameter. We evaluated models with  $k$  ranging from 4 to 15 topics, assessing coherence scores and interpretability. The six-topic solution provided optimal balance: topics were semantically coherent (high-probability words formed interpretable themes), sufficiently granular to distinguish AI issue areas, and stable across different random initializations.

**Hyperparameters.** We set symmetric Dirichlet priors with  $\alpha = 50/k$  for document-topic distributions and  $\beta = 0.01$  for topic-word distributions. These values encourage moderately sparse topic mixtures (documents drawing from multiple but not all topics) while allowing topics to be characterized by distinctive vocabulary. Model fitting used variational Bayes inference with 500 iterations.

**Interpretation.** Each topic is characterized by a probability distribution over the vocabulary. We interpreted topics by examining the 20 highest-probability terms and reviewing representative documents with high topic proportions. Topic labels (e.g., 'Privacy and Surveillance', 'Employment and Algorithmic Management') were assigned based on semantic coherence of top terms and content of exemplar documents.

LDA captures broad thematic domains by pooling statistical signals across documents. Its probabilistic foundation handles document-level noise gracefully, producing topics that reflect corpus-wide patterns rather than idiosyncratic features of individual documents. For AI litigation, LDA reveals the major categories of legal concern at an aggregate level.

### 3.9. Non-Negative Matrix Factorization (NMF)

Non-negative Matrix Factorization decomposes the document-term matrix into two lower-rank matrices: a document-topic matrix and a topic-term matrix [6]. Unlike LDA's probabilistic interpretation, NMF operates through linear algebra, finding an additive parts-based decomposition of the original matrix.

**Complementary Properties.** NMF tends to produce more focused, keyword-driven topics than LDA. Where LDA topics capture broad thematic domains, NMF topics often identify specific constructs or terminology clusters. This property makes NMF valuable for identifying granular AI issues nested within broader categories. For legal text specifically, NMF's additive decomposition aligns with how legal documents combine distinct procedural and substantive components.

**Model Specification.** We fit NMF using coordinate descent optimization with Frobenius norm objective function. The nine-topic solution provided optimal coherence while decomposing LDA's broad categories into finer-grained issues. For example, where LDA identified a general 'Intellectual Property' topic, NMF distinguished between trade secrets, patent inventorship, and copyright issues.

**Hierarchical Relationship.** By applying both LDA ( $k = 6$ ) and NMF ( $k = 9$ ), we discovered a hierarchical topic structure where broad LDA domains decompose into specific NMF issues. This hierarchy reveals how general AI concerns (e.g., ‘data and privacy’) subdivide into actionable categories (e.g., ‘biometric data collection’, ‘algorithmic surveillance’, ‘data breach liability’). The hierarchical structure informs governance by identifying both overarching frameworks needed and specific mechanisms for targeted issues.

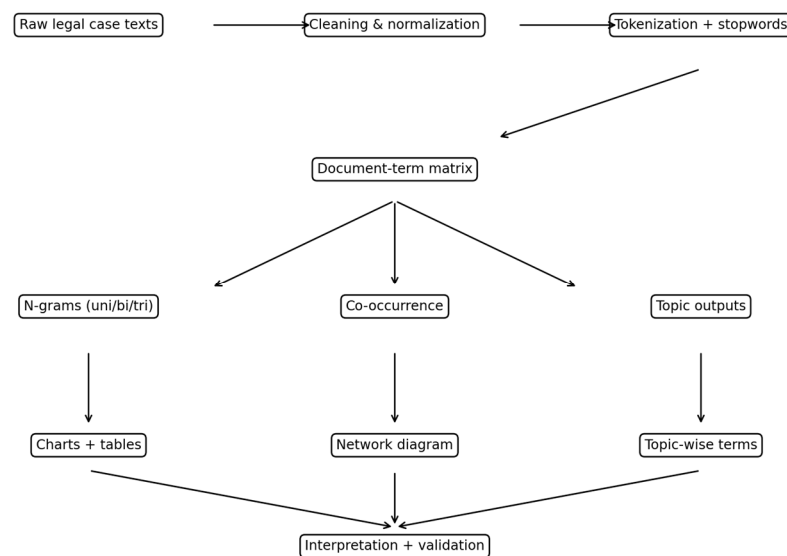
### 3.10. Validation and Model Selection

Topic models require validation to ensure that discovered topics represent meaningful themes rather than statistical artifacts. We employed multiple validation approaches to assess topic quality and guide model selection.

**Coherence Scores.** Topic coherence measures whether high-probability words within a topic tend to co-occur in the corpus. We computed normalized pointwise mutual information (NPMI) coherence scores, which correlate with human judgments of topic interpretability [43]. Higher coherence indicates that topic words form semantically related clusters rather than arbitrary groupings. NMF topics achieved higher average NPMI coherence (0.276) than LDA topics (0.164), consistent with NMF’s tendency toward focused, coherent topics.

**Stability Analysis.** Topic models can produce different results from different random initializations. We assessed stability by fitting models multiple times with different random seeds and measuring topic similarity across runs [44]. Both LDA and NMF produced stable topic structures, with major topics appearing consistently across initializations.

**Internal Robustness Check.** An internal robustness check compared cases with explicit AI terminology (e.g., ‘artificial intelligence’, ‘machine learning’) against those with implicit AI references (e.g., ‘algorithm’, ‘automated decision’). Topic structures remained consistent across subsets, confirming that findings are not driven by superficial keyword presence but reflect deeper thematic patterns. Figure 2 outlines the text analytics pipeline.



**Figure 2.** Text analytics pipeline for surfacing AI issues from litigation data.

## 4. Results

### 4.1. Overview of AI Issues Surfaced

The analysis surfaced nine distinct AI issue areas from the litigation corpus. These represent the key contemporary challenges in AI that have generated sufficient real-world

impact to result in formal cases. Table 2 presents the distribution of cases across issue areas, revealing where AI problems are most concentrated.

**Table 2.** Distribution of AI-Related Cases by Topic, Case Type, and Defendant Type.

Topic	Cases Tagged	Civil	Criminal	Government	Private/Other
Cybersecurity & data breaches in AI systems	116	108	8	18	98
Intellectual property & ownership	61	57	4	28	33
AI misrepresentation, marketing & fraud	59	55	4	33	26
Criminal justice & due process	37	0	37	21	16
Employment law & workplace AI	33	31	2	14	19
Privacy, surveillance & data protection	31	22	9	20	11
Platform liability & consumer protection	21	20	1	12	9
Bias, discrimination & civil rights	19	12	7	9	10
Government use of AI & administrative law	6	4	2	6	0

This table reports the number of cases tagged to each AI issue area, with a heuristic breakdown by civil versus criminal cases and by government versus private defendants. Cybersecurity vulnerabilities represent the largest category, indicating that AI systems’ data dependencies create significant security exposure. Intellectual property challenges including questions of AI-generated content ownership and algorithmic trade secrets constitute the second largest category. Notably, AI misrepresentation cases reveal a distinct challenge: companies overstating AI capabilities or misleading stakeholders about algorithmic functionality. Table 2 reveals important contextual patterns. Criminal AI issues arise exclusively in criminal contexts, highlighting concerns about algorithmic tools in high-stakes settings where liberty is at stake. Privacy and government AI deployment issues disproportionately involve government actors, signaling public sector AI adoption as a distinct challenge area.

Table 3 shows which regulatory frameworks are being applied to AI issues. The prominence of the First Amendment reflects concerns about AI-generated content and platform speech. FOIA invocations signal demands for transparency in government AI systems. Title VII and Equal Protection cases indicate that algorithmic discrimination is being challenged through existing civil rights frameworks. Notably, no AI-specific legislation appears—AI issues are being addressed through adaptation of existing regulatory tools. Table 4 below shows the most frequent terms in the corpus.

**Table 3.** Regulatory Frameworks Most Frequently Invoked in AI Cases.

Regulatory Framework	Cases Invoking
First Amendment (speech, expression)	11
Lanham Act (trademarks, false advertising)	9
Freedom of Information Act (transparency)	9
Administrative Procedure Act (agency action)	8
Privacy Act (government data)	8
Title VII (employment discrimination)	8
Copyright Act (authorship, ownership)	7
Equal Protection (discrimination)	7

**Table 4.** Most Frequent Terms in AI Issue Corpus.

Term	Frequency
plaintiff	48
Court	31
Class	27
sanctions	26
Patent	22
Trade	21
Ai	18

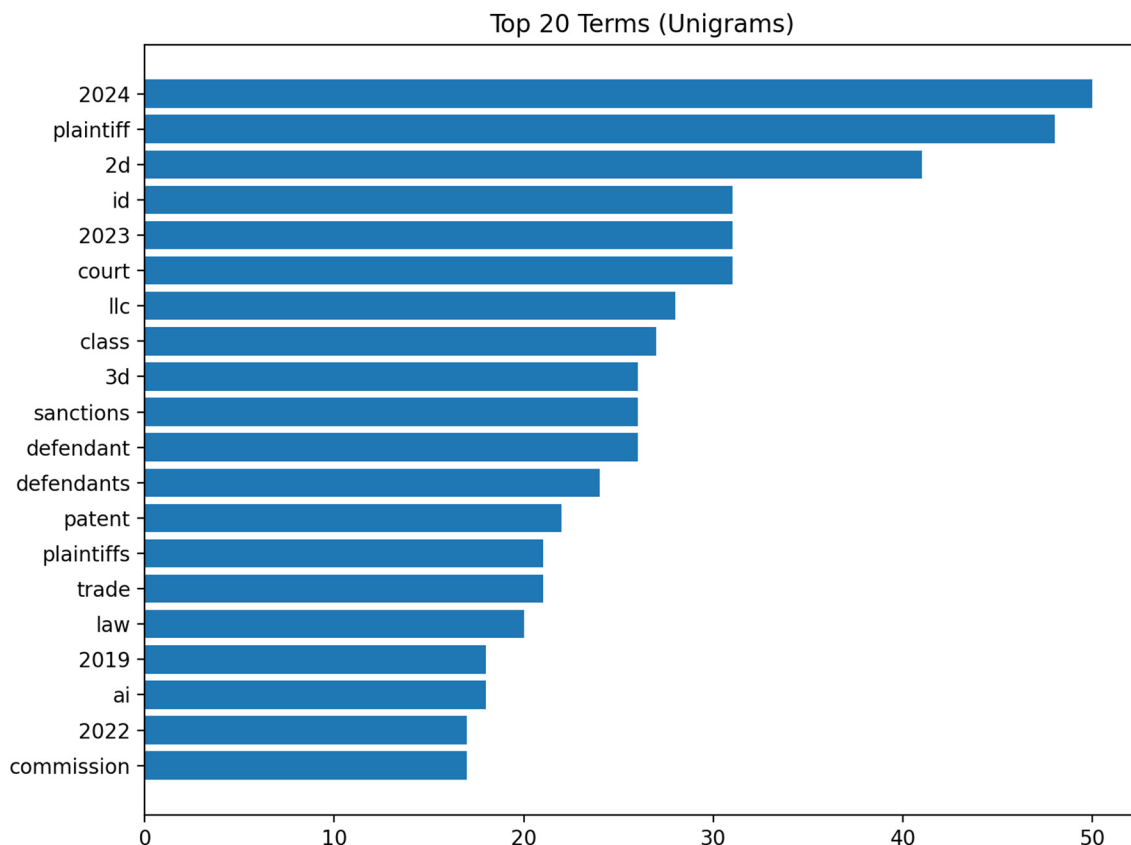
Table 5 presents the key topic, citation count, representative court cases in which artificial intelligence or algorithmic systems are directly implicated in the underlying dispute and a brief description. The cases are included to ground each topic in concrete AI-related issues surfaced by the text analysis, such as automated decision-making, algorithmic opacity, training data practices, and AI-enabled surveillance. The purpose is illustrative rather than doctrinal, highlighting how distinct categories of AI issues manifest in litigation rather than providing comprehensive legal analysis.

**Table 5.** Representative AI-Substantive Cases by Topic.

Topic	Citations of Cases	Representative AI-Substantive Case(s)	Brief AI Issue Description
Bias, Discrimination & Civil Rights	6	State v. Loomis (2016); Mobley v. Workday, Inc. (2023)	Algorithmic risk scoring and automated hiring systems alleged to produce biased or discriminatory outcomes.
Criminal Justice & Due Process	13	State v. Loomis (2016); Houston Fed’n of Teachers v. Houston ISD (2017)	Use of opaque algorithmic tools in high-stakes government decisions raising transparency and fairness concerns.
Privacy, Surveillance & Data Protection	27	Carpenter v. United States (2018); ACLU v. Clearview AI, Inc.	AI-enabled mass data collection, facial recognition, and surveillance practices implicating privacy rights.
Employment Law & Workplace AI	11	Mobley v. Workday, Inc. (2023); EEOC v. iTutorGroup, Inc. (2023)	Automated screening and evaluation systems used in employment decision-making.
Intellectual Property & AI Ownership	24	Thaler v. Vidal (2022); Andersen v. Stability AI Ltd. (2023)	Ownership, authorship, and training data questions arising from generative AI systems.
Platform Liability & Consumer Protection	12	Gonzalez v. Google LLC (2023); Lemmon v. Snap, Inc. (2021)	Algorithmic recommendation and design features alleged to contribute to consumer harm.
AI Misrepresentation, Marketing & Fraud	36	FTC v. Rite Aid Corp. (2023); FTC v. Kurbo, Inc.	Alleged deceptive claims and improper deployment of AI systems in consumer-facing contexts.
Government Use of AI & Administrative Law	13	Houston Fed’n of Teachers v. Houston ISD; State v. Loomis	Public-sector adoption of algorithmic systems affecting rights, benefits, or regulatory outcomes.
Cybersecurity & AI Systems	10	In re Blackbaud, Inc. Data Breach Litigation; FTC v. CafePress Inc.	Security failures in data-driven and algorithm-supported systems handling sensitive information.

#### 4.2. Lexical Patterns in AI Issue Discourse

Text frequency analysis reveals the vocabulary through which AI issues are articulated. High-frequency terms reflect the institutional contexts where AI problems surface, while bigrams and trigrams capture specific AI-related constructs and concerns as indicated in Figure 3.



**Figure 3.** Term Frequency Distribution.

This bar chart displays the relative frequency of key terms extracted from the AI litigation corpus after preprocessing. The horizontal axis represents individual terms while the vertical axis indicates term frequency counts. Terms appearing with higher frequency reflect dominant concepts in the litigation landscape. The distribution reveals that procedural terms (plaintiff, court, class) appear alongside substantive AI-related concepts (patent, trade, AI), indicating that the corpus captures both legal process dimensions and technology-specific issues.

#### 4.3. Co-Occurrence Networks: How AI Issues Cluster

The co-occurrence analysis in Figure 4 reveals which AI-related concepts appear together, exposing the interconnected nature of AI challenges. The network visualization shows that AI issues do not arise in isolation—they cluster into bundles of related concerns.

This network visualization illustrates the co-occurrence relationships among AI-related concepts within the litigation corpus. Nodes represent key terms, with node size proportional to term frequency. Edges connect terms that frequently appear together in the same documents, with edge thickness indicating co-occurrence strength. Clusters of densely connected nodes reveal conceptual groupings—for example, intellectual property terms cluster together, as do privacy and surveillance concepts. The network structure demonstrates that AI issues do not arise in isolation but form interconnected bundles of related legal and technical challenges.





#### 4.5. Hierarchical Topic Structure

Complementing the issue-based taxonomy, unsupervised topic modeling reveals latent thematic structure in the corpus in Table 6. The LDA six-topic solution captures broad AI issue domains, while the NMF nine-topic solution reveals finer-grained procedural and doctrinal patterns.

**Table 6.** Broad AI Issue Domains (LDA 6-Topic Solution).

Topic	AI Issue Domain	Share	Key Indicators
1	Data Access, Platforms, and Scraping	14.9%	defendants, securities, exchange
2	Creative Works, Authorship, and IP	11.7%	trade, intellectual, copyright
3	Employment and Algorithmic Management	14.6%	amazon, defendant, Microsoft
4	Consumer Harm and Deceptive AI	20.4%	LLC, civil, signed, court
5	Privacy and Surveillance	19.9%	plaintiff, sanctions, patent
6	Government and Regulatory AI	18.6%	law, trial, state, injunction

The six LDA topics reveal the broad thematic domains structuring AI litigation in Figure 6.



**Figure 6.** Composite Visualization of Broad AI Issue Domains (LDA).

Topic 1 (Data Access, Platforms, and Scraping) captures disputes over automated data collection, web scraping technologies, and platform terms of service—reflecting tensions between AI systems' data requirements and proprietary interests.

Topic 2 (Creative Works, Authorship, and IP) encompasses intellectual property disputes involving AI-generated content, questions of authorship attribution, and copyright infringement claims that have intensified with generative AI proliferation.

Topic 3 (Employment and Algorithmic Management) addresses workplace AI applications including algorithmic hiring tools, automated performance evaluation, and worker surveillance systems—with major technology companies frequently appearing as parties.

Topic 4 (Consumer Harm and Deceptive AI) represents the largest domain (20.4%), capturing cases involving misleading AI product claims, consumer protection violations, and fraudulent representations of AI capabilities.

Topic 5 (Privacy and Surveillance) encompasses AI-enabled data collection, biometric surveillance, and inferential privacy harms—reflecting growing concerns about algorithmic monitoring capabilities.

Topic 6 (Government and Regulatory AI) captures public sector AI deployment, including challenges to agency algorithmic decision-making, demands for transparency under FOIA, and constitutional challenges to government AI systems. Together, these six domains provide a comprehensive map of the AI issue landscape as manifested in litigation.

Next, Table 7 summarizes the 9 topics uncovered in the NMF analysis.

**Table 7.** Specific AI Issue Areas (NMF 9-Topic Solution).

Topic	Specific AI Issue	Cases	Key Indicators
1	Procedural and Filing Issues	22	services, signed, division
2	Frivolous AI Claims and Sanctions	35	sanctions, appeal, fees, costs
3	Individual Complaints and Grievances	27	plaintiff, se, defendant
4	Class-wide AI Harms	28	class, certification, damages
5	AI in Adjudicative Settings	56	state, trial, jury, evidence
6	Trade Secrets and Algorithmic IP	21	trade, secrets, injunction
7	AI Inventorship and Patents	33	patent, natural, thaler, dabus
8	Government Transparency and FOIA	76	commission, agency, foia
9	AI Contracts and Business Disputes	49	LLC, agreement, license

The nine NMF topics in Figure 7 provide finer-grained resolution of AI litigation issues, revealing both substantive and procedural dimensions. Topics 1–5 capture procedural patterns: Topic 1 (Procedural and Filing Issues) reflects routine case administration; Topic 2 (Frivolous AI Claims and Sanctions) identifies cases where AI-related allegations were deemed unsubstantiated, resulting in court sanctions; Topic 3 (Individual Complaints and Grievances) captures pro se litigants raising AI concerns; Topic 4 (Class-wide AI Harms) encompasses class action litigation alleging systematic AI-related injuries; and Topic 5 (AI in Adjudicative Settings) addresses AI tools used in legal proceedings themselves, including evidence authentication and expert testimony disputes.



**Figure 7.** Composite Visualization of Specific AI Issue Areas (NMF).

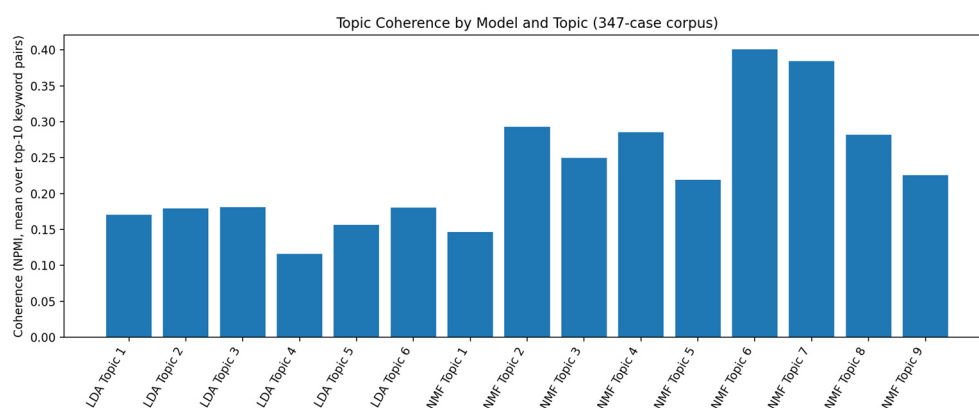
Topics 6–9 reflect substantive AI issue areas with greater specificity than the LDA domains. Topic 6 (Trade Secrets and Algorithmic IP) focuses narrowly on proprietary algorithm protection and misappropriation claims. Topic 7 (AI Inventorship and Patents) captures the prominent line of cases challenging whether AI systems can be named as patent inventors, including the widely cited DABUS/Thaler litigation. Topic 8 (Gov-

ernment Transparency and FOIA) represents the largest NMF topic (76 cases), encompassing Freedom of Information Act requests seeking disclosure of government AI systems and algorithmic decision-making criteria. Topic 9 (AI Contracts and Business Disputes) addresses commercial disputes over AI product licensing, service agreements, and vendor relationships.

The NMF decomposition reveals that AI litigation involves not only substantive technology issues but also distinctive procedural patterns—a dimension absent from ethics-focused AI issue taxonomies. The prominence of government transparency cases (Topic 8) and sanctions for frivolous claims (Topic 2) underscores how institutional and procedural factors shape which AI issues enter the litigation system and how they are resolved.

#### 4.6. Model Validation and Coherence

NMF topics exhibit higher average NPMI coherence (0.276) than LDA topics (0.164) in Figure 8, indicating that the finer-grained NMF topics capture tighter semantic clusters.



**Figure 8.** Topic Coherence Comparison: LDA vs. NMF.

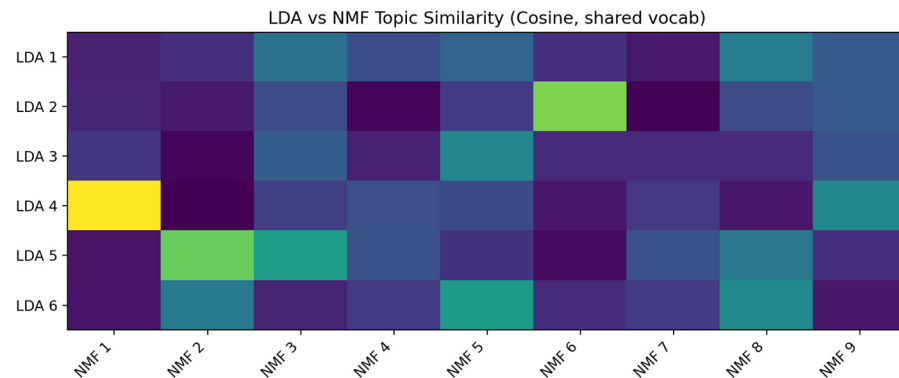
This comparative visualization displays the Normalized Pointwise Mutual Information (NPMI) coherence scores for topics generated by Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF). Higher coherence scores indicate that the top words within each topic are more semantically related, suggesting more interpretable and meaningful topic structures. The chart demonstrates that NMF achieves higher average coherence (0.276) compared to LDA (0.164), indicating that NMF topics capture tighter semantic clusters in this corpus. This validation metric supports the use of both methods in triangulation while suggesting that NMF may provide more precise issue decomposition.

#### 4.7. Mapping Broad Domains to Specific Issues

Comparing LDA and NMF topics reveals how broad AI issue domains decompose into specific challenges.

This heatmap matrix in Figure 9 displays the alignment between broad LDA topics (rows) and more granular NMF topics (columns). Cell intensity indicates the degree of overlap between topic pairs, measured by shared top terms or document assignments. Darker cells represent stronger alignment, revealing how LDA's broader thematic domains decompose into NMF's more specific issue areas. The matrix demonstrates that certain LDA topics map cleanly to individual NMF topics (indicating conceptual coherence), while others distribute across multiple NMF topics (indicating that the broad domain encompasses distinct sub-issues). Figure 9 displays a heatmap comparing topic similarity between LDA and NMF models using cosine similarity on a shared vocabulary. The color scale follows a sequential colormap, progressing from dark purple through blue and teal to green and yellow. Dark purple and indigo cells represent low similarity values, roughly

in the 0.0 to 0.2 range, indicating that those topic pairs share very little vocabulary. As the colors shift toward blue and then teal, the similarity increases to moderate levels, approximately 0.4 to 0.6. The brightest cells—those appearing in green and yellow—signal high similarity, approaching 0.8 to 1.0, meaning those topic pairs are strongly aligned in their word distributions.

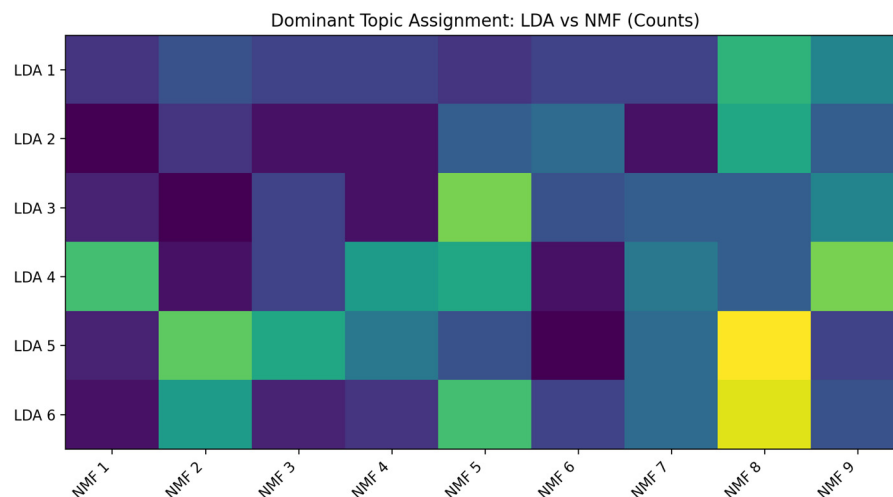


**Figure 9.** Topic Alignment Matrix.

Looking at the matrix, the most striking alignments appear as yellow cells. LDA topic 4 aligns strongly with NMF topic 1, and LDA topic 2 shows high similarity with NMF topic 6. Additionally, LDA topic 5 demonstrates a notable correspondence with NMF topic 2, visible as a bright green cell. Beyond these prominent pairings, there are scattered teal cells indicating moderate alignment, such as LDA topic 3 with NMF topics 5 and 7, and LDA topic 6 with NMF topics 5 and 8.

The predominance of dark purple and blue across most of the matrix tells an important story: the majority of LDA-NMF topic pairs do not align strongly. This is a common and expected finding, as the two algorithms employ fundamentally different mathematical approaches to discovering latent structure. LDA uses a probabilistic generative model while NMF relies on matrix factorization, so they naturally surface somewhat different thematic patterns from the same corpus. The few bright cells scattered throughout the matrix suggest that while certain themes are robust enough to emerge from both methods, each algorithm also captures distinct aspects of the underlying text that the other does not.

This visualization shows how the 347 cases in the corpus distribute across combinations of LDA and NMF topic assignments. Each cell represents a topic pair (one LDA topic and one NMF topic), with values indicating the number of cases assigned to both topics. The distribution reveals which broad AI issue domains (LDA) contain which specific sub-issues (NMF), and which topic combinations are most prevalent in the litigation landscape. Concentrated values along certain rows or columns indicate that particular broad domains strongly associate with specific procedural or substantive patterns. Figure 10 shows the case distribution across dominant topic assignments from both LDA and NMF, again using a sequential colormap from dark purple to yellow. Dark purple cells indicate that very few cases were jointly assigned to that particular LDA-NMF topic pair, while yellow and bright green cells highlight combinations where many cases clustered together. The brightest cell appears at the intersection of LDA 5 and NMF 8, suggesting this pairing captured the largest number of cases. Other notable concentrations appear at LDA 3 with NMF 5, LDA 4 with NMF 5, and LDA 5 with NMF 2. The overall pattern reveals that cases are not evenly distributed across all possible topic combinations but instead concentrate in specific pairings, indicating which thematic intersections are most prevalent in the corpus.



**Figure 10.** Case Distribution Across Topic Pairs.

4.8. Comparing Manual and Algorithmic Issue Identification

The nine AI issue areas presented in Figure 5 were derived through manual categorization based on substantive legal domains, while the LDA and NMF topics emerged through unsupervised algorithmic discovery. Comparing these approaches reveals both convergence and complementary insights, strengthening confidence in the identified AI issues while exposing dimensions each method captures uniquely.

Table 8 reveals strong convergence between manual and algorithmic approaches for several issue areas. Intellectual property issues align clearly with LDA Topic 2 (Creative Works/IP) and decompose in NMF into distinct subtopics for trade secrets (Topic 6) and patent inventorship (Topic 7), with the latter capturing the prominent DABUS/Thaler AI inventorship litigation. Employment automation maps consistently to LDA Topic 3, while NMF reveals that employment cases divide between individual grievances managed through adjudicative processes (Topic 5) and class-wide algorithmic harms (Topic 4). Government AI deployment shows strong alignment with both LDA Topic 6 and NMF Topic 8, with the latter specifically capturing FOIA-based transparency demands.

**Table 8.** Mapping Manual Issue Categories to Algorithmic Topics.

Manual Issue Category	LDA Topic Alignment	NMF Topic Alignment
IP & AI Ownership	Topic 2: Creative Works, IP	Topics 6, 7: Trade Secrets, Patents
Algorithmic Bias	Topic 3: Employment (partial)	Topic 5: Adjudicative Settings
Employment Automation	Topic 3: Employment	Topics 4, 5: Class Actions, Adjudicative
Privacy & Surveillance	Topic 5: Privacy	Topic 8: Gov’t Transparency
Criminal Justice	Topic 6: Gov’t/Regulatory	Topic 5: Adjudicative Settings
Platform Accountability	Topics 1, 4: Platforms, Consumer	Topic 9: AI Contracts
AI Misrepresentation	Topic 4: Consumer Harm/Deceptive	Topics 2, 9: Sanctions, Contracts
Government AI	Topic 6: Gov’t/Regulatory	Topic 8: Gov’t Transparency/FOIA
Cybersecurity	Distributed (no single topic)	Distributed (procedural focus)

The comparison also reveals what each approach captures uniquely. The manual taxonomy excels at identifying substantive AI challenges—the ‘what’ of AI issues—organized around recognizable policy domains. In contrast, NMF topics capture procedural and structural dimensions—the ‘how’ of AI litigation—including sanctions for frivolous AI

claims (Topic 2), class action certification patterns (Topic 4), and contract/licensing disputes (Topic 9). These procedural dimensions, invisible in substantive categorization, reveal important patterns: AI misrepresentation cases frequently involve sanctions for unsubstantiated claims, while cybersecurity cases—though substantively cohesive—distribute across multiple procedural postures.

Notably, cybersecurity, the largest manual category, does not emerge as a distinct algorithmic topic in either LDA or NMF. This divergence reflects how cybersecurity cases, while substantively unified around data breach and security failure issues, exhibit diverse procedural characteristics: some proceed as class actions, others as individual complaints, and still others involve regulatory enforcement. The algorithmic methods thus reveal that ‘cybersecurity’ as an AI issue encompasses heterogeneous litigation patterns unified more by subject matter than by procedural or doctrinal structure.

The complementary nature of these approaches strengthens the study’s findings. Manual categorization provides interpretable, policy-relevant issue areas that align with how practitioners, policymakers, and researchers discuss AI challenges. Algorithmic discovery validates this structure while adding granularity—particularly the substantive/procedural distinction—that pure manual analysis might miss. The convergence between methods on core issue areas (IP, employment, privacy, government AI) increases confidence that these represent genuine, robust categories of contemporary AI challenges rather than artifacts of any single methodological approach. Further, to enhance the rigor and transparency of the study, Appendix A describes the manual verification procedures employed in this study. It summarizes the systematic process used to validate topic model outputs and case categorizations.

## 5. Discussion

This study demonstrates that litigation data, analyzed through machine learning text analytics, provides a systematic method for identifying key contemporary AI issues. The nine issue areas surfaced—cybersecurity vulnerabilities, intellectual property, AI misrepresentation, criminal justice applications, employment automation, privacy and surveillance, platform accountability, algorithmic bias, and government AI deployment—represent challenges that have moved beyond speculation into documented real-world problems.

Several findings merit emphasis. First, cybersecurity emerges as the dominant AI challenge, comprising one-third of the corpus. This suggests that AI systems’ data dependencies create substantial security exposure that is often underemphasized in AI ethics discussions. Second, AI misrepresentation constitutes a distinct issue area—companies overstating AI capabilities represent a meaningful category of AI-related harm. Third, the absence of AI-specific regulatory frameworks indicates that AI challenges are being addressed through adaptation of existing tools rather than purpose-built AI governance.

The hierarchical topic structure reveals that AI issues operate at multiple levels of specificity. Broad domains like ‘privacy and surveillance’ or ‘intellectual property’ subdivide into more granular challenges. This suggests that effective AI governance requires both high-level frameworks addressing general concerns and specific mechanisms targeting issue areas.

## 6. Limitations and Future Research

This section delineates the limitations of the study and outlines directions for future research. The limitations are presented explicitly to clarify the scope of inference, methodological constraints, and conditions under which the findings should be interpreted. The

discussion also translates these limitations into operational guidance for future empirical work on AI issues.

#### *6.1. Data Coverage and Selection Bias*

The study relies on litigation cases retrieved from the Westlaw database, which introduces inherent coverage limitations. Litigation captures only AI-related issues that have escalated to formal legal disputes, thereby excluding many AI failures resolved informally or not pursued through legal channels. As a result, the findings reflect high-severity and high-stakes AI challenges rather than the full universe of AI system problems.

In addition, Westlaw coverage may underrepresent cases settled confidentially, disputes in arbitration, or litigation in jurisdictions with limited digital reporting. These biases limit the generalizability of the findings and suggest that the identified AI challenges should be interpreted as conservative estimates of recurring problems.

#### *6.2. Preprocessing and Representation Limitations*

Text preprocessing decisions materially influence topic modeling outcomes. The removal of legal stop words and the application of document-frequency thresholds improve interpretability but may suppress infrequent yet substantively important AI issues. Lemmatization and n-gram construction introduce abstraction that can blur contextual nuance.

Moreover, the transformation of lengthy legal texts into bag-of-words representations discards syntactic structure and relational information. These representational limitations constrain the ability of topic models to capture causal reasoning, legal argumentation structure, or temporal sequencing of events.

#### *6.3. Statistical and Model-Based Limitations*

Unsupervised topic models such as LDA and NMF are sensitive to parameter selection, corpus composition, and initialization. Topic boundaries are probabilistic rather than deterministic, and coherence metrics provide only approximate indicators of semantic quality. Consequently, topic prevalence estimates should not be interpreted as precise measurements of issue frequency.

Topic interpretation involves subjective judgment, but this subjectivity arises from identifiable statistical properties of the models rather than arbitrary labeling. The triangulation of LDA, NMF, and manual verification mitigates but does not eliminate uncertainty in topic assignment.

#### *6.4. Transferability Across Legal Systems*

The findings of this study are derived exclusively from U.S. litigation and may not transfer directly to other legal systems. Common law jurisdictions such as the United Kingdom, Canada, and Australia share procedural and doctrinal features with the U.S. system, potentially supporting greater transferability of the identified AI issue categories. However, civil law jurisdictions in continental Europe, Asia, and Latin America operate under fundamentally different procedural frameworks, evidentiary standards, and regulatory structures that may produce distinct patterns of AI-related disputes. For example, the adversarial nature of U.S. litigation may surface AI issues through class actions and discovery processes that are less prominent in inquisitorial systems. Additionally, jurisdictions with comprehensive AI-specific regulations—such as the European Union’s AI Act—may channel AI disputes through administrative enforcement rather than private litigation, potentially altering which issues reach judicial resolution. The prominence of intellectual property disputes in this corpus may also reflect the particular strength of U.S. patent and trade secret protections, which vary considerably across jurisdictions. Consequently,

while the substantive AI challenges identified—cybersecurity vulnerabilities, algorithmic bias, and AI misrepresentation—are likely to arise globally, the relative frequency and legal framing of these issues may differ substantially across legal systems. Future research should replicate this methodology in other jurisdictions to assess the generalizability of these findings and identify jurisdiction-specific AI governance challenges.

#### *6.5. Future Research Directions and Operational Implications*

Future research can extend this work by integrating additional data sources, such as regulatory enforcement actions, arbitration records, and international case repositories, to reduce selection bias. Methodologically, incorporating supervised or semi-supervised approaches may improve precision when labeled data become available.

From an operational perspective, the findings suggest that organizations deploying AI systems should prioritize governance mechanisms addressing data management, model transparency, and intellectual property controls—areas that consistently generate litigation. Policymakers may use litigation-derived evidence to target oversight resources toward AI challenges that have already demonstrated real-world consequences.

## **7. Conclusions and Future Research**

This study examined artificial intelligence issues and challenges through a systematic analysis of U.S. litigation, using machine learning–based text analytics as an empirical lens. Rather than evaluating legal doctrine, the analysis focused on identifying recurring AI system challenges that have generated documented real-world conflict. By grounding the analysis in litigation records, the study provides evidence-based insight into which AI issues have progressed beyond theoretical concern into practical consequence.

Across multiple analytical stages, the findings show that AI-related litigation concentrates around a limited and recurring set of challenges, including data governance failures, intellectual property disputes, cybersecurity vulnerabilities, and the operational impacts of automated decision-making. These challenges reflect systemic governance and organizational issues rather than isolated technical malfunctions, underscoring the need for integrated approaches to AI risk management.

Methodologically, the study demonstrates the value of combining exploratory text analysis, multiple unsupervised topic modeling techniques, and structured manual verification. The convergence of results across Latent Dirichlet Allocation, Non-negative Matrix Factorization, and manual taxonomy construction strengthens confidence in the robustness and interpretability of the identified AI issue categories. At the same time, the analysis illustrates the importance of transparency and triangulation when applying unsupervised methods to complex legal texts.

The conclusions of this study are intentionally bound. The findings apply to AI-related litigation within the U.S. legal system and reflect the characteristics of adversarial judicial processes and available case reporting. The results should therefore be interpreted as identifying high-severity AI challenges that have triggered formal disputes, rather than as a comprehensive inventory of all AI risks.

Future research can extend this work along several dimensions. Empirically, incorporating additional data sources—such as regulatory enforcement actions, arbitration records, and international litigation databases—would enable comparative analysis across institutional and jurisdictional contexts. Methodologically, integrating supervised or semi-supervised learning approaches may improve precision as labeled datasets mature. Substantively, future studies can examine how identified AI challenges evolve over time and how governance interventions influence litigation patterns.

Taken together, this study contributes an empirically grounded perspective on AI issues and challenges, demonstrating how litigation-based evidence can complement ethics-oriented and policy-driven approaches. As AI systems continue to proliferate across high-stakes domains, empirically identifying the challenges that have already produced real-world consequences remains essential for effective governance, accountability, and risk management.

**Author Contributions:** Methodology, W.R. and T.K.; Software, T.K.; Validation, A.S.; Formal analysis, T.K.; Writing—original draft, W.R. and A.S.; Writing—review & editing, W.R. and A.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to privacy.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A. Manual Verification Process

To enhance transparency regarding the manual verification procedures employed in this study, this appendix summarizes the systematic process used to validate topic model outputs and case categorizations.

**Topic Labeling Verification:** Following unsupervised topic extraction via LDA and NMF [5,6], each topic was independently reviewed by two researchers. For each topic, reviewers examined the top 15 weighted terms and a random sample of 10 high-loading documents. Reviewers proposed topic labels and brief descriptions based on semantic coherence and substantive interpretability [42,43]. Where initial labels diverged, reviewers discussed the underlying documents and converged on consensus labels. This dual-review process aimed to reduce subjective bias in topic interpretation.

**Case Categorization Verification:** For the nine-category manual taxonomy, a stratified sample of 50 cases (approximately 15% of the corpus) was independently coded by two researchers. Each case was assigned to one or more of the nine issue categories based on the substantive legal claims and AI-related facts described in the case summary. Inter-rater agreement was calculated using Cohen's kappa, yielding  $\kappa = 0.78$ , indicating substantial agreement. Discrepancies were resolved through discussion and reference to full case documents where necessary.

**Triangulation Process:** The convergence between manual categorization and algorithmic topic discovery was assessed by mapping each manually assigned category to corresponding LDA and NMF topics [41]. Strong alignment between methods (as reported in Table 8) provided validation that the identified AI issue areas reflect genuine patterns in the data rather than artifacts of any single analytical approach.

**Limitations of Manual Verification:** While these procedures enhance reliability, manual verification remains subject to reviewer expertise and judgment. The sample sizes used for inter-rater reliability assessment, while standard in qualitative research, represent a subset of the full corpus. Additionally, topic interpretation inherently involves subjective judgment about which semantic dimensions are most salient [42]. These limitations are acknowledged to maintain transparency about the verification process.

## References

1. Jobin, A.; Ienca, M.; Vayena, E. The global landscape of AI ethics guidelines. *Nat. Mach. Intell.* **2019**, *1*, 389–399. [CrossRef]
2. Giarmoleo, F.V.; Ferrero, I.; Rocchi, M.; Pellegrini, M.M. What ethics can say on artificial intelligence: Insights from a systematic literature review. *Business and Society Review* **2024**, *129*, 258–292. [CrossRef]
3. Pew Research Center. How the U.S. public and AI Experts View Artificial Intelligence. 2025. Available online: <https://www.pewresearch.org/internet/2025/04/03/how-the-us-public-and-ai-experts-view-artificial-intelligence/> (accessed on 21 December 2025).
4. Stanford HAI. AI Index Report 2025, Chapter 8: Public Opinion. 2025. Available online: <https://hai.stanford.edu/ai-index/2025-ai-index-report> (accessed on 21 December 2025).
5. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
6. Lee, D.D.; Seung, H.S. Learning the parts of objects by non-negative matrix factorization. *Nature* **1999**, *401*, 788–791. [CrossRef] [PubMed]
7. Molitor, D.; Raghupathi, W.; Saharia, A.; Raghupathi, V. Exploring Key Issues in Cybersecurity Data Breaches: Analyzing Data Breach Litigation with ML-Based Text Analytics. *Information* **2023**, *14*, 600. [CrossRef]
8. Raghupathi, W.; Molitor, D.; Raghupathi, V.; Saharia, A. Identifying Key Issues in Climate Change Litigation: A Machine Learning Text Analytic Approach. *Sustainability* **2023**, *15*, 16530. [CrossRef]
9. Livermore, M.A.; Riddell, A.; Rockmore, D. The Supreme Court and the judicial genre. *Ariz. Law Rev.* **2017**, *59*, 837–901. [CrossRef]
10. Arksey, H.; O'Malley, L. Scoping studies: Towards a methodological framework. *Int. J. Soc. Res. Methodol.* **2005**, *8*, 19–32. [CrossRef]
11. Khan, S.; Khan, H.U.; Nazir, S. Systematic analysis of ethical challenges in artificial intelligence. *IEEE Access* **2021**, *9*, 71758–71775.
12. Ipsos. The Ipsos AI Monitor 2024. 2024. Available online: <https://www.ipsos.com/en-us/ipsos-ai-monitor-2024> (accessed on 21 December 2025).
13. Schwartz Reisman Institute. *Global Public Opinion on Artificial Intelligence (GPO-AI)*; University of Toronto: Toronto, ON, Canada, 2024. Available online: <https://srinstitute.utoronto.ca/news/public-opinion-ai-survey-24> (accessed on 21 December 2025).
14. Melbourne Business School. Global Study Reveals Trust of AI Remains a Critical Challenge. 2025. Available online: <https://mbs.edu/news/global-study-reveals-trust-of-ai-remains-a-critical-challenge> (accessed on 21 December 2025).
15. Brookings/GRAIL. AI SHARE: What the Public Thinks about AI and the Implications for Governance. 2025. Available online: <https://www.brookings.edu/articles/what-the-public-thinks-about-ai-and-the-implications-for-governance/> (accessed on 21 December 2025).
16. Gallup/SCSP. Americans Prioritize AI Safety and Data Security. 2025. Available online: <https://news.gallup.com/poll/694685/americans-prioritize-safety-data-security.aspx> (accessed on 21 December 2025).
17. Alan Turing Institute; Ada Lovelace Institute. How do people feel about AI? A Nationally Representative Survey of Public Attitudes to Artificial Intelligence in Britain. 2023. Available online: <https://www.turing.ac.uk/research/research-projects/understanding-public-attitudes-ai> (accessed on 21 December 2025).
18. IEEE Spectrum. What U.S. Members Think About Regulating AI. 2024. Available online: <https://spectrum.ieee.org/regulating-ai-ieee-survey> (accessed on 21 December 2025).
19. Buolamwini, J.; Gebru, T. Gender shades: Intersectional accuracy disparities in commercial gender classification. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency, New York, NY, USA, 23–24 February 2018; pp. 77–91.
20. Brookings Institution. Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harm. 2019. Available online: <https://www.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/> (accessed on 21 December 2025).
21. Angwin, J.; Larson, J.; Mattu, S.; Kirchner, L. *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks*; ProPublica: New York, NY, USA, 2016.
22. Chouldechova, A. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data* **2017**, *5*, 153–163. [CrossRef] [PubMed]
23. Kleinberg, J.; Mullainathan, S.; Raghavan, M. Inherent trade-offs in the fair determination of risk scores. In Proceedings of the 8th Innovations in Theoretical Computer Science Conference, Berkeley, CA, USA, 9–11 January 2017.
24. Obermeyer, Z.; Powers, B.; Vogeli, C.; Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **2019**, *366*, 447–453. [CrossRef]
25. ScienceDirect. Bias and Ethics of AI Systems Applied in Auditing: A Systematic Review. 2024. Available online: <https://www.sciencedirect.com/science/article/pii/S2468227624002266> (accessed on 21 December 2025).
26. Partnership on AI. AI Incident Database. 2024. Available online: <https://incidentdatabase.ai/> (accessed on 21 December 2025).
27. McGregor, S. Preventing repeated real world AI failures by cataloging incidents: The AI Incident Database. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020.
28. MIT AI Risks Repository. *AI Incident Tracking and Classification System*; MIT FutureTech: Cambridge, MA, USA, 2024.

29. OECD. *AI Incidents and Hazards Monitor: Methodology and Classification Framework*; OECD Digital Economy Papers: Paris, France, 2024.
30. Chalkidis, I.; Androutsopoulos, I.; Aletras, N. Neural legal judgment prediction in English. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 4317–4323.
31. Quid. Stanford AI Index Report Partnership. 2024. Available online: <https://www.quid.com/2024-stanford-ai-index-report> (accessed on 21 December 2025).
32. European Union Fundamental Rights Agency. *Bias in Algorithms: Artificial Intelligence and Discrimination*; FRA Publications: Washington, DC, USA, 2022.
33. Busch, F.; Hoffmann, L.; Xu, L.; Zhang, L.J.; Hu, B.; García-Juárez, I.; Toapanta-Yanchapaxi, L.N.; Gorelik, N.; Gorelik, V.; Rodriguez-Granillo, G.A.; et al. Multinational attitudes toward AI in health care and diagnostics among hospital patients. *JAMA Netw. Open* **2025**, *8*, e2514452. [[CrossRef](#)]
34. Young, A.T.; Amara, D.; Bhattacharya, A.; Wei, M.L. Patient and general public attitudes towards clinical artificial intelligence: A mixed methods systematic review. *Lancet Digit. Health* **2021**, *3*, E599–E611. Available online: [https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(21\)00132-1/fulltext](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(21)00132-1/fulltext) (accessed on 21 December 2025).
35. Raghavan, M.; Barocas, S.; Kleinberg, J.; Levy, K. Mitigating bias in algorithmic hiring: Evaluating claims and practices. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, 27–30 January 2020; pp. 469–481.
36. Rutgers University. National AI Opinion Monitor (NAIOM). 2024. Available online: <https://comminfo.rutgers.edu/news/survey-reveals-concerns-and-adoption-trends-around-ais-rising-influence> (accessed on 21 December 2025).
37. Ashley, K.D. *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*; Cambridge University Press: New York, NY, USA, 2017.
38. Grimmer, J.; Stewart, B.M. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Anal.* **2013**, *21*, 267–297. [[CrossRef](#)]
39. Lauderdale, B.E.; Clark, T.S. Scaling politically meaningful dimensions using texts and votes. *Am. J. Political Sci.* **2014**, *58*, 754–771. [[CrossRef](#)]
40. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv* **2013**, arXiv:1301.3781. [[CrossRef](#)]
41. Baumer, E.P.; Mimno, D.; Guha, S.; Quan, E.; Gay, G.K. Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *J. Assoc. Inf. Sci. Technol.* **2017**, *68*, 1397–1410. [[CrossRef](#)]
42. Chang, J.; Gerrish, S.; Wang, C.; Boyd-Graber, J.L.; Blei, D.M. Reading tea leaves: How humans interpret topic models. *Adv. Neural Inf. Process. Syst.* **2009**, *22*, 288–296.
43. O’Callaghan, D.; Greene, D.; Carthy, J.; Cunningham, P. An analysis of the coherence of descriptors in topic modeling. *Expert Syst. Appl.* **2015**, *42*, 5645–5657. [[CrossRef](#)]
44. Greene, D.; O’Callaghan, D.; Cunningham, P. How many topics? stability analysis for topic models. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 498–513.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.