The Cost of Silence: Quantifying the Prejudice of Non-Participation in Domain Name Disputes

Vinny Adjibi
verGeorgia Institute of Technology
School of Electrical and Computer Engineering
Atlanta, Georgia, USA
vinny.adjibi@gatech.edu

Kathryn Kleiman
American University
Washington College of Law
Washington, District of Columbia, USA
kleiman@wcl.american.edu

Michael Bailey
Georgia Institute of Technology
School of Cybersecurity and Privacy
Atlanta, Georgia, USA
mbailey@gatech.edu

Fabian Monrose
Georgia Institute of Technology
School of Electrical and Computer Engineering
Atlanta, Georgia, USA
fabian@gatech.edu

Abstract

For nearly three decades, the majority of trademark-based domain name disputes have been decided based solely on complainants' submissions, as most domain name owners (or respondents) do not participate in the proceedings. Although domain name dispute resolution policies typically include provisions that guarantee a fair outcome for respondents who fail to participate (or default), anecdotal evidence shows a large discrepancy in outcomes between default and contested disputes. For instance, under the Uniform Domain Name Dispute Resolution Policy (UDRP), the domain name dispute procedure adopted by the Internet Corporation for Assigned Names and Numbers (ICANN) in 1999 for "generic top-level domains" (gTLDs), such as .com, .org and .net, defaulting respondents are substantially more likely to lose than those who file a response. This disparity has prompted recurring concerns about potential structural imbalances in the dispute resolution process, yet the lack of large-scale empirical analyses has limited ICANN's ability to systematically evaluate these concerns and inform policy updates.

In this work, we present a large-scale comparative analysis of outcomes in contested and default domain name disputes. We analyze 42,306 UDRP cases initiated between July 2015 and November 2025, including 3,190 contested and 39,116 default proceedings. Using large language models and text embedding techniques, we extract and standardize the factual grounds cited by panels and cluster disputes resolved on substantively similar facts. Our analysis identifies 1,359 cases in which defaulting respondents lost disputes that were consistently won by participating respondents under comparable factual circumstances. We also observe systematic differences in the level of factual justification provided in default proceedings, suggesting variation in decision-writing practices that may bear on the transparency and perceived fairness of the process. The datasets, code artifacts, and prompts used for our analyses are publicly available at https://doi.org/10.5281/zenodo.19594587.

ICAIL 2026, Singapore

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

CCS Concepts

• **Information systems** → Document topic models; *Clustering*; • **Applied computing** → *Law*; • **Social and professional topics** → **Internet governance / domain names**; • **General and reference** → *Empirical studies.*

Keywords

Uniform Domain Name Dispute Resolution, Default Decisions

ACM Reference Format:
Vinny Adjibi, Kathryn Kleiman, Michael Bailey, and Fabian Monrose. 2026. The Cost of Silence: Quantifying the Prejudice of Non-Participation in Domain Name Disputes. In *Proceedings of 21st International Conference on Artificial Intelligence and Law (ICAIL 2026).* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Domain names occupy a unique position at the intersection of internet addressing and intellectual property. While corporations invest heavily to acquire and promote domain names as primary source identifiers—much like trademarks—the mechanisms to acquire them differ fundamentally. Unlike trademarks, which are territorial in nature and "only protected in the countries or regions where they are registered [47]," domain names are internet identifiers and part of the global Domain Name System (DNS). While trademarks require rigorous substantive examination, domain names can be registered instantly, at negligible cost, and without prior screening for conflicting rights [32]. In fact, the DNS's design allows the coexistence of identical strings across different namespaces, a feature further expanded by the recent introduction of new generic top-level domains [24]. For example, `nissan.com` and `nissan.co.jp` are technically distinct and owned by unrelated entities: the former by an individual named Nissan and the latter by the well-known automaker. However, as the protracted litigation over `nissan.com` [14] illustrates, this structural openness often leads to disputes when domain registration rights collide with trademark protections [19].

The existing out-of-court mechanisms used to decide such disputes differ significantly from typical arbitration proceedings [31]. They are initiated unilaterally by a trademark holder, triggering a strict timeline that requires the domain name owner, who may be unfamiliar with the Uniform Domain Name Dispute Resolution (UDRP) procedure, its dispute resolution forums, or the language of the complaint, to submit a formal response. In practice, however, the majority of respondents do not engage with this process. This high rate of default is particularly evident in the Uniform Domain Name Dispute Resolution Policy (UDRP), where respondents file a response in one out of every three disputes [13].

This absence has material consequences: earlier studies showed that defaulting respondents lose their disputes at a much higher rate than those who participate [13], raising questions about the structural equity of the dispute resolution process [20]. Despite these concerns, ICANN has largely retained existing UDRP rules governing respondent participation in the dispute process. This continuity may, in part, reflect the limited availability of empirical evidence assessing the scope of the alleged imbalance. A search for articles whose title or abstract contain the policy's name alongside the word "bias" on SCOPUS, Web of Science, and Google Scholar, returned only 6 of 58 entries involving some sort of empirical analysis. Of those, four were at least a decade-old [20, 30, 35, 40] and the two most recent studies did not study the potential impact of bias on decision outcomes [2, 13].

In this work, we measured the extent to which bias might have affected the domain name dispute resolution process, thereby providing empirical evidence to guide future reviews and reforms. We focused on bias against defaulting respondents, assessing if, when presented with the same factual evidence, panelists' decisions differ depending on whether the respondent participated in the dispute or not. We also compared the level of factual justification provided in default versus contested disputes. To conduct these analyses, we leveraged recent advances in natural language processing, combining large-language models and text-embedding techniques to extract the key facts used to justify the outcomes of past UDRP disputes. In total, we extracted the facts used by panels to justify their decisions in 3,190 contested and 39,116 default disputes started after the policy update of July 31, 2015. This extraction allowed us to create clusters of disputes in which the facts analyzed by the panels were semantically similar. Through a systematic evaluation of the generated clusters, we identified several types of instances in which respondents appear to have lost a dispute solely because they did not submit a response. Ultimately, our results suggest the need for effective interventions to increase respondent participation in the dispute process and improve the quality and transparency of panel decisions.

Our core contributions are as follows:

(1) We developed a scalable methodology to extract legal reasoning from dispute proceedings and applied it to a 10-year dataset of domain name disputes. As a modular approach, our technique can be extended to other legal tasks.
(2) We performed the first systematic review of the variation in decision-writing practices between default and contested UDRP disputes, finding significant discrepancies that may bear on the process's transparency and perceived fairness.
(3) We designed and implemented a methodology that abstracts the specifics of a dispute to create clusters of factually equivalent disputes. Those facts can serve as a checklist for improving panelists' decision-making or as a guide for respondents to prepare a strong defense.
(4) We conducted the first large-scale empirical analysis of the differences in outcomes between default and contested domain name disputes. We identified several issues that require policy updates to increase trust in the policy.

The remainder of this paper is organized as follows. Section 2 describes the UDRP procedure and Section 3 discusses the related work. Section 4 presents our dataset and in Section 5, we assess the differences in terms of winning rates (§5.1), number of facts analyzed by panels (§5.2), and decision outcomes (§5.3) between default and constested UDRP disputes. We conclude with recommendations for ICANN and its stakeholder groups in Section 6 and a discussion of our study's limitations in Section 7.

2 Background

The Uniform Domain Name Dispute Resolution Policy (UDRP) is the foremost mechanism used to resolve disputes arising from a trademark holder's objection to the bad-faith registration and use of a domain name similar to their trademark. It features a streamlined, efficient dispute resolution process that has been used in tens of thousands of disputes since the policy's adoption in 1999.

To start a dispute under the UDRP for generic top-level domains (gTLD), a trademark owner must file a complaint with an ICANN-accredited provider, mentioning the contentious domain name(s) and explaining why they consider that *(i)* the domain name(s) are identical or confusingly similar to their trademark; *(ii)* the domain name owner has no rights or legitimate interests in respect of the domain name(s); and *(iii)* the domain name(s) have been registered and are being used in bad faith. Upon receiving and accepting the complaint, the dispute resolution provider contacts the domain name's registrar of record to lock the domain name, and, if necessary, to seek the registrant's contact details, which may have been redacted in accordance with global data protection laws such as the European Union General Data Protection Regulation (GDPR). Upon receiving the registrant's name and contact information, the dispute resolution provider notifies the domain name owner about the complaint and gives them an opportunity to rebut the claimant's allegations within 20 days. As part of the mandatory registration agreement for gTLD domain names [25], every registrant agrees to be bound to the results of a UDRP decision by any ICANN-accredited provider, which can include transferring or cancelling the domain name. However, very few domain name owners have participated in the past. While the reasons for this low rate of participation remain unclear, some consider it an admission of wrongdoing on the part of the domain name owners [13, 42] while others contend it may be because respondents failed to receive the provider's notice or lacked enough time or resources to prepare a reasonable defense [20, 49].

Regardless of the respondent's participation status, the dispute resolution provider appoints a one- or three-member panel to review the dispute, with the process conducted entirely through online submissions. The panel's final decision, due within two weeks

of their appointment, depends on an analysis of the facts to determine if the three mandatory elements below are satisfied:

(1) The domain name is identical or confusingly similar to a mark in which the claimant has rights; *and*
(2) The respondent has no rights or legitimate interests in respect of the domain name; *and*
(3) The domain name has been registered and is being used in bad faith [26].

Upon reaching a decision, the panel must explain their reasoning in a document published on the provider's website at the end of the dispute resolution proceeding. Based on an analysis of past disputes, researchers estimate that over 90% of past disputes were decided in favor of claimants, resulting in the transfer or cancellation of the disputed domain name(s) [2, 6, 13]. In general, this high winning rate is used as an indication that the UDRP is mostly used for legitimate rather than baseless disputes. At the same time, it has also fueled a broad range of criticisms from some scholars who allege the existence of a systemic bias favoring trademark holders [20, 49], especially given earlier studies' observation that defaulting respondents were significantly more likely to lose a dispute than their non-defaulting counterparts [13]. Unfortunately, the lack of empirical analyses assessing the existence of such a bias may have stalled the enactment of relevant reforms likely to increase respondent participation.

3 Related Work

The decentralized, cross-border nature of the domain name industry and trademark-based disputes that the UDRP aims to manage has made the policy an interesting object of study for legal and technical scholars alike. Starting shortly after the policy's inception, a steady stream of research began focusing on assessing the quality of decisions rendered [34, 38, 41] and the implications of procedural factors on fairness and free speech [2, 20, 49]. One highly debated aspect of fairness is the issue of forum shopping, which refers to claimants' suspected tendency to file disputes with providers likely to increase their chance of winning the dispute. Although the existence of this practice has long been contentious among legal scholars [18, 20, 30, 35], a recent large-scale analysis showed that forum shopping likely affected a third of past UDRP disputes [2].

Researchers have also been broadly concerned with the fairness of decisions rendered in default disputes. In a two-decades-old analysis of 3,362 disputes, Mueller asserted that in up to 25% of disputes involving a default, respondents' non-participation in the process was the only basis relied upon by panelists to conclude that the domain name was registered in bad faith [36]. The author's analysis also revealed panelists' willingness to passively accept a claimant's allegations as true when respondents default. According to Mueller, those prejudicial actions partly contributed to the disproportionate rate at which defaulting respondents had lost their disputes.

In this work, we conducted a large-scale review of default dispute decisions using machine learning to empirically verify this claim. Although this line of inquiry is novel in the context of domain name disputes, similar analyses have been conducted in other areas of the law. Notably, researchers have used machine learning to assess the consistency of decisions rendered by judges deciding criminal [11, 45] or asylum cases [10, 11, 37]. Those studies evaluated the existence of unfair decisions by analyzing how non-legal factors, such as the defendant's nationality and the judge's identity, influence litigation outcomes. However, a recent study showed that these approaches might not be suitable for identifying unfairness, as they minimize the importance of the legal grounds on which adjudicators make their decisions [4]. According to Barale et al. [4], fairness is best evaluated using approaches that model why adjudicators rendered a certain decision, without the influence of procedural and contextual factors. Our work draws from those insights to develop a methodology that relies solely on the facts discussed by the domain name dispute panelists to assess the quality of decisions when respondents default.

4 Dataset

For a contemporary and rigorous investigation, we focused our analysis on single-domain UDRP disputes filed in the last decade, for which the panels made a decision only after evaluating the merits of the complaint. This last condition allows us to filter out two types of disputes: those in which the panel ordered a transfer solely because the respondent agreed to it, and those in which the complaint was rejected as out of scope or for other procedural issues. Finally, focusing on single-domain disputes prevents the analytical complexities of multi-domain complaints, where panels' tendency to discuss all domains simultaneously makes it difficult to isolate the facts for each domain name.

We initiated our data collection at the end of November 2025, scraping the websites of the five accredited dispute resolution providers in service at that time: *1)* the Internet Domain Resolution section of the Arbitration and Mediation Center of the World Intellectual Property Organization; *2)* the National Arbitration Forum; *3)* the Asian Domain Name Dispute Resolution Center; *4)* the Czech Arbitration Court Center for Internet Disputes; and *5)* the Canadian International Internet Dispute Resolution Center. Those websites typically provide, for each UDRP dispute ever managed by the provider, basic metadata such as the parties' identities, the list of disputed domains, the publication date, and a link to the published proceedings. More granular details, such as when the complaint was filed or whether the respondent has defaulted, are only available as free-form text in the published proceedings. Accordingly, we leveraged Adjibi et al.'s state-of-the-art named entity recognition (NER) model [2], which can accurately extract such information at scale, to determine the complaint submission date and the respondent's participation status for the 108,587 disputes we collected, some of which date back to 1999. Since that model only supports English documents, we used the highly performant FastText model [28] to identify each document's language and retained only the 98,989 English proceedings. After filtering out multi-domain disputes, we identified 42,871 complaints submitted on or after July 31, 2015, based on the NER model's output. We subsequently used the Gemini 2.5 Flash reasoning model to analyze the remaining documents and identify disputes in which panels made a decision without analyzing the merits of the dispute for any of the reasons mentioned earlier. As Figure 1 shows, those steps resulted in the identification of 3,190 contested and 39,116 default disputes studied in this work.

We consider this dataset adequate for a systematic comparison of default and contested UDRP disputes because it exhibits a moderate

positive correlation between the number of default and contested decisions rendered by a panelist (Pearson's $r = 0.58$, $p < 0.01$). In other words, the caseload of panelists is sufficiently balanced across both dispute types to allow meaningful comparisons.

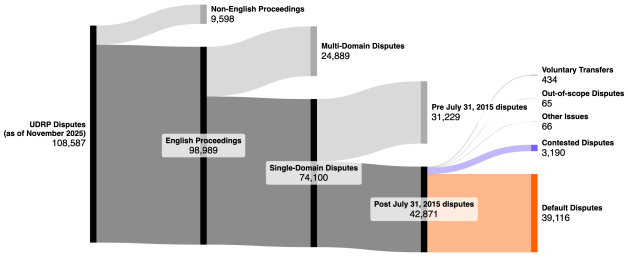

Figure 1: Data selection pipeline. Our analysis focuses on a comparison between contested (3,190) and default disputes (39,116), highlighted respectively in purple and orange.

5 Evaluation

Our study examines whether and to what extent non-participation in domain name disputes diminishes respondents' likelihood of retaining control of their domain names. To operationalize this objective, we formulate three research questions:

(1) *How do respondent win rates differ among defaulting respondents, self-represented respondents, and respondents represented by legal counsel?*
(2) *How does the evidentiary basis of panel decisions—measured by the number of facts cited—differ between default and contested disputes?*
(3) *When disputes involve substantively identical factual circumstances, to what extent do panel outcomes diverge between default and contested proceedings?*

5.1 Win Rates and Participation Status

> *Finding 1.* Defaulting respondents win nearly twenty times less often than legally represented respondents and roughly ten times less than self-represented respondents.

We use metadata from dispute providers' websites to estimate how often panelists decided in respondents' favor. Motivated by prior work suggesting that legal representation may improve respondent outcomes [2], we partition contested disputes into those with and without legal representation. A manual inspection revealed that decisions involving represented respondents typically include a standardized phrase naming the respondent, followed by "represented by" and the counsel's name. Using regular expressions derived from this pattern, we identify 970 contested disputes (30.40%) in which respondents were legally represented. Our analysis shows that legally represented respondents achieve the highest win rate (53.09%), compared to 26.40% for self-represented respondents and 2.71% for defaulting respondents. These results indicate that respondent participation—and especially legal representation—is strongly associated with substantially improved chances of prevailing in domain name disputes.

Given that several works have argued that respondents primarily default because they are "cybersquatters" lacking any viable defense [13, 42], we explore whether dispute outcomes vary with a respondent's prior involvement in domain name disputes. For that analysis, we distinguish between one-time respondents and repeat respondents, the latter being more likely to face allegations of systematic or malicious registration behavior. For repeat respondents, we only considered those with a consistent participation pattern throughout the study period (i.e., they either always defaulted, participated on their own, or were legally represented).

As shown in Figure 2, the gap in win rates between defaulting and participating respondents remains pronounced. Among participating respondents, those involved in multiple disputes exhibit a significantly higher win rate than those appearing in only a single one, regardless of whether they are legally represented. In contrast, one-time defaulting respondents only win 3.20% of their disputes, which is twice the win rate for repeat defaulters (1.43%). Taken together, these patterns are consistent with the hypothesis that experience with the dispute process—or repeated engagement—may modestly improve respondents' ability to defend their domain names, while repeated defaults may negatively influence panels' perceptions.

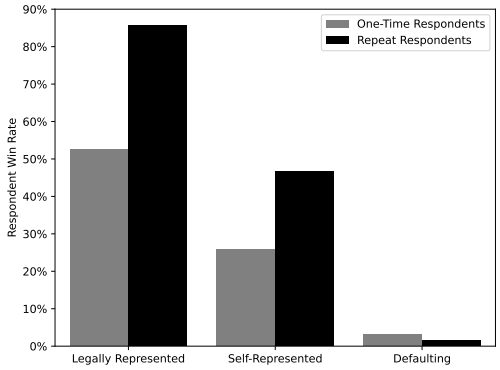

Figure 2: Proportion of disputes won by respondents depending on their participation status.

To evaluate the relationship between participation frequency and win probability, we employed binary logistic regression models, setting the dispute outcome as the dependent variable and the number of disputes a respondent had been involved in as the independent variable. We also included an intercept term to establish a baseline win probability and estimated the models' parameters using maximum likelihood estimation. This approach was used elsewhere to estimate the influence of dispute outcomes on provider selection [2, 30]; it allows us to estimate how each additional participation affects the respondent's chances of winning a dispute.

The results indicated a positive but not statistically significant relationship between participation frequency and winning probability for legally represented respondents ($\beta_1 = 0.713$, $p = 0.120$). For self-represented respondents, we noted a statistically significant

positive relationship ($\beta_1 = 0.548$, $p = 0.021$) with a baseline win rate of 26.24%. By exponentiating the coefficient β_1, we estimated that self-represented respondents improve their chances of winning a dispute by 73.05% with each new participation. In contrast, the group of defaulting respondents demonstrated a statistically significant negative relationship ($\beta_1 = -0.037$, $p < 0.001$), with each additional default reducing the odds of winning by approximately 3.65%. As illustrated in Figure 3, those results translate into a steep decline in respondents' probability to win a dispute the more they default. Those observations support our hypothesis regarding the relationship between the frequency of participation or default in disputes and decision outcomes under the UDRP.

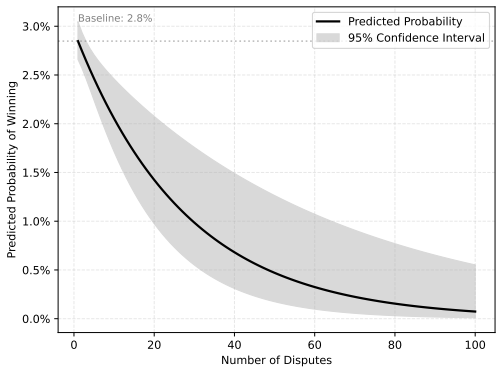

Figure 3: Predicted probability of a defaulting respondent winning a domain name dispute.

5.2 Justification of Default Decisions

> *Finding 2.* Panel decisions cite fewer justifications when respondents default than when they participate, particularly in disputes decided in the claimant's favor.

For transparency, the UDRP policy requires panels to include in their published decisions the factual evidence they considered and how it informed their judgment. This stipulation applies even when respondents default, as panels may still verify the claimant's assertions against available evidence, including domain ownership data provided by the domain registrar and evidence submitted by the claimant. In this section, we assess how the number of facts considered by panels differs between default and contested disputes.

To ensure we extract the facts in a manner that mirrors panel decision-making, we aligned our methodology with existing legal reviews, which indicated that panelists evaluate facts separately for each of the four requirements listed below [5, 38].

(1) **Trademark Standing and Similarity**, comprising:
  (1) **Establishing Rights in a Mark**, verifying the validity of the claimant's trademark.
  (2) **Identicality or Confusing Similarity**, assessing whether the domain is identical or confusingly similar to the mark.
(2) **Rights or Legitimate Interests**, determining if the respondent possesses a valid claim to the domain.
(3) **Bad Faith**, deciding if the domain name was registered and is being used in bad faith.

Figure 4 highlights how requirements are addressed in a representative decision. As shown, panelists typically organize their analysis into delineated sections, each introduced by a descriptive heading. Within these sections, the panel presents the relevant facts, explains its interpretation of those facts, and states its conclusion as to whether the corresponding requirement is satisfied.

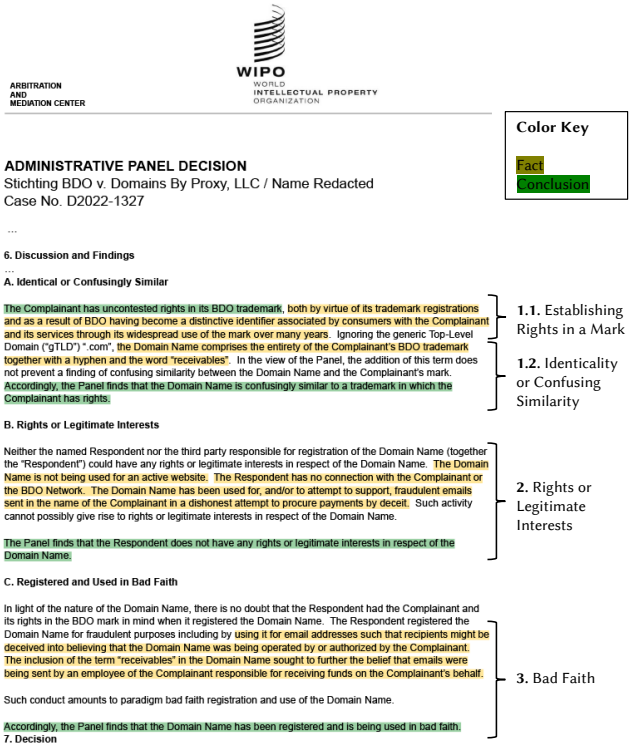

Figure 4: Annotated dispute proceeding highlighting the facts based on which the panel made a decision about the four dispute requirements and the panel's conclusions about them.

To extract this unstructured information from our large corpus of disputes, we devised a methodology that leverages the semantic reasoning and capabilities of instruction-tuned models to accurately isolate and extract relevant facts. Our methodology, represented in Figure 5, starts by converting the published decision HTML or PDF documents to Markdown using the `MarkItDown` Python package[1]. We chose this intermediary markup language because it preserves the hierarchical structure of documents, and has been shown to improve LLMs' understanding of lengthy legal documents [7].

We first prompted the Gemini 2.5 Flash reasoning model [21] to analyze the Markdown-formatted document and identify the document section where the panel discussed the requirement of interest. For example, in the case of the dispute shown in Figure 4, the model identifies the section titled *"A. Identical or Confusingly Similar"* when prompted to find the section relevant to requirements 1.1

[1] https://github.com/microsoft/markitdown

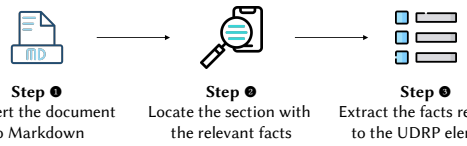

Figure 5: Methodology used to extract the relevant facts for each UDRP requirement.

and 1.2. Using only this relevant section, we prompted the model to return a list of the facts considered, along with the panel's conclusion on that requirement (i.e., whether the requirement was met or not). The intermediate step of extracting the section narrows the search space for facts, preventing the model from retrieving irrelevant data and mitigating the "lost-in-the-middle" performance degradation common in long-context LLMs [33].

Applying this methodology, we find that the vast majority of facts extracted by the model (98.61%) match the source text exactly, except for minor differences in capitalization or punctuation. For the remaining non-exact matches, we quantify textual similarity using the *Longest Common Subsequence* (LCS) similarity metric [44]. We compute LCS between each extracted fact and its source document and normalize the score by the length of the extracted text. LCS measures structural similarity between strings of unequal length by identifying the longest sequence of characters appearing in the same relative order in both texts. Excluding exact matches, the mean normalized LCS similarity is very high (98.38%). A manual review of a random subset of 20 non-exact matches reveals that the residual differences arise primarily from benign edits introduced by the language model, such as correcting typographical errors or replacing ambiguous pronouns with their explicit referents. These results confirm that the model followed the instructions and primarily returned quotations from the input documents.

We also analyzed cases in which the LLM did not extract any facts and found two potential reasons for that. In some cases, panelists effectively decided on some requirements without justifications in the appropriate section, while other instances of empty facts related to disputes where the panel decided to abstain from analyzing a requirement because another requirement was not satisfied. This latter observation is consistent with prior work showing that, given the conjunctive nature of the requirements, panels often stop their evaluation once a requirement is not met [38]. To prevent those instances from skewing our analysis, we only considered entries with empty facts if the panel had concluded that the requirement was met or if that requirement was the only unmet requirement in the dispute. Based on those criteria, we discarded 11 disputes.

Using the remaining dataset, we conducted a series of Mann–Whitney U tests to evaluate whether the number of facts cited by panels differs between default and contested disputes. For each dispute requirement, we test the null hypothesis H_0 that the distribution of facts extracted from default disputes is identical to that of contested disputes. At a 99.99% confidence level, we reject H_0 in favor of the alternative hypothesis H_1 that contested disputes are supported by significantly more cited facts. Although this disparity could be attributed to panels receiving submissions from only one party in default disputes, two considerations challenge this explanation.

First, we observe that panels cite substantially more facts in a specific subset of default decisions—namely, those in which the complainant does not prevail. To establish this, we conduct additional Mann–Whitney U tests comparing the distribution of facts extracted from default disputes won by complainants to those they lost. At a 99.99% confidence level, these tests show that panels discuss significantly more facts when ruling against complainants for requirements *1.1. Establishing Rights in a Mark* and *2. Rights or Legitimate Interests*. As illustrated in Figure 6, these results demonstrate that, even in the absence of any respondent submission, panels expand their factual discussion when denying a complaint. This pattern is consistent with the possibility that panels provide more extensive justification when denying a complaint brought by the cost-bearing complainant than when ruling against a respondent who did not participate in the proceeding.

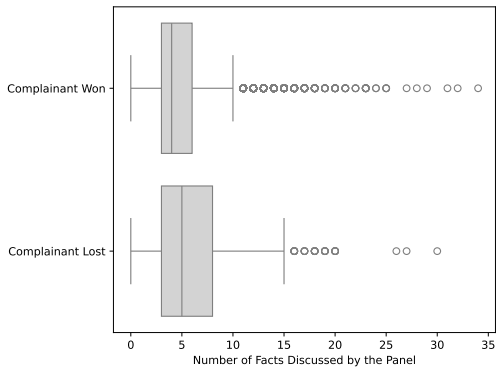

Figure 6: Number of facts considered by panels in default decisions when establishing if the respondent has rights or legitimate interests in the domain name.

The second consideration draws on the WIPO Overview 3.0 [46], a document that panelists frequently cite in their decisions. It clarifies that the first UDRP element, encompassing requirements 1.1 and 1.2, is intended solely to establish whether the complainant has standing to bring a dispute. Under this framework, the analysis of these requirements is expected to apply irrespective of respondent participation. Accordingly, the presence of a statistically significant difference in the number of facts cited for these requirements between default and contested disputes is difficult to reconcile with panelists' responsibilities. While this disparity does not by itself establish a denial of due process, it indicates variation in the level of justification applied in default proceedings. Such variation may, in turn, affect perceptions of fairness and transparency, potentially undermining confidence in the UDRP dispute resolution process.

5.3 Consistency of Default Decisions

> *Finding 3.* Panelists ordered the transfer or cancellation of 1,359 domains belonging to respondents who did not participate in the dispute, when, in similar circumstances, respondents who participated won 100% of their disputes.

We now assess how non-participation in the dispute process is associated with decision outcomes by identifying factually similar disputes and comparing results across default and contested proceedings. We define *factually similar* disputes as those in which the facts cited by panels when evaluating a given requirement form a cluster of semantically similar excerpts. Figure 7 illustrates this concept using two representative disputes. In both examples, the panel's analysis of the bad-faith requirement relies primarily on the respondent's offer to sell the domain name at a high price. Despite this similarity, the outcomes differ: the respondent who filed a response (*Example 1*) prevailed, whereas the defaulting respondent (*Example 2*) did not. Building on this illustration, this section systematically examines how frequently such outcome divergences occur when disputes are resolved on comparable factual grounds.

5.3.1 Method. To establish semantic relationships between disputes, we leveraged `BERTopic` [22], a topic modelling algorithm that has successfully been deployed in prior work to analyze legal documents [3, 16]. BERTopic features a modular architecture that sequentially combines an *embedding model* to vectorize sentences, a *clustering algorithm* to group semantically similar vectors, and a *representation model* to extract keywords characterizing each cluster. In this work, we employed `Voyage-Large-3` for the embedding stage, selected for its superior performance in both legal and general domain retrieval tasks [8, 43]. We utilized the model's 256-dimensional output truncation to optimize computational efficiency while retaining the semantic richness required for accurate clustering. For the clustering algorithm, we used HDBSCAN due to its robust ability to identify clusters of varying densities and automatically determine the optimal number of clusters while filtering out noise [9]. In our configuration, we set the minimum cluster size to 2—the lowest possible threshold. This parameter choice was designed to maximize the granularity of our topic model, ensuring that even small thematic patterns were captured as distinct topics rather than being discarded as outliers.

To ensure that clustering reflects substantive patterns rather than dispute-specific identifiers, we apply an abstraction step prior to clustering. Named entities such as trademarks, company names, and proper nouns can otherwise dominate similarity estimates and result in clusters driven by lexical overlap rather than shared reasoning. Accordingly, we generate controlled abstractive summaries of the extracted facts following established prompt engineering practices for precise summarization [1]. This normalization step removes case-specific identifiers and emphasizes the underlying narrative structure used by panels in their evaluations. For example, in Figure 4, a fact referencing the mark "BDO" is abstracted into the requirement-level summary: "*The complainant establishes its trademark rights through registrations and long-standing use.*"

5.3.2 Metrics. We evaluate the quality and fidelity of the generated summaries using BERTScore [50], a commonly used metric for assessing machine-generated text in the absence of human-labeled ground truth. Prior work has shown that BERTScore generally correlates better with human judgments than traditional n-gram–based metrics such as BLEU for text generation tasks [15, 23]. Across all pairs of extracted facts and their corresponding summaries, we observe an average BERTScore of 0.77, which is well within the accepted range for accurate abstractive summaries [15].

Cluster: The respondent offered the disputed domain for sale.

Example 1: Respondent Won

"Given the Panel's finding on the issue of 'rights or legitimate interests,' the Panel concludes that the disputed domain name was not registered and is not being used in bad faith. See John Fairfax Publ'n Pty Ltd. v. Domain Names 4U, D2000-1403 (WIPO Dec. 13, 2000) (finding legitimate interests and no bad faith registration where the respondent is a seller of generic domain names.) While Respondent is offering the disputed domain name for sale for an amount far in excess of what it cost to obtain and register, that, apparently, is the business model followed by domain name resellers, such as Respondent. "

Example 2: Respondent Lost

"Respondent (who did not reply to Complainant's contentions) has not presented any plausible explanation for its use of Complainant's mark. In accordance with paragraph 14(b) of the Rules, the Panel shall draw such inferences from Respondent's failure to reply as it considers appropriate. Accordingly, the Panel finds that Respondent did not have a legitimate use in mind when registering the disputed domain name.
Indeed, as already noted, Respondent offers the disputed domain name for sale for a price in excess of out-of-pocket costs. Under Policy ¶ 4(b)(i), there is generally a finding of bad faith registration and use where the respondent uses the resolving webpage to offer the domain name for sale. See Airbnb, Inc. v. Super Privacy Service LTD c/o Dynadot, FA 1821386 (Forum Jan. 10, 2019) ('Complainant argues that Respondent registered and uses the <airbnbb.com> domain name in bad faith by offering it for sale. The Panel agrees and finds that Respondent registered and uses the disputed domain name in bad faith under Policy ¶ 4(b)(i).'); see also Staples, Inc. v. lin yanxiao, FA1505001617686 (Forum June 4, 2015) ('Respondent's offering to sell the disputed domain name to a third party (in this case, the general public) supports a finding of bad faith registration and use.'); see also Bank of Am. Corp. v. Nw. Free Cmty. Access, FA 180704 (Forum Sept. 30, 2003) ('Respondent's general offer of the disputed domain name registration for sale establishes that the domain name was registered in bad faith under Policy ¶ 4(b)(i).'); see also Am. Anti-Vivisection Soc'y v. "Infa dot Net" Web Serv., FA 95685 (Forum Nov. 6, 2000) (finding that 'general offers to sell the domain name, even if no certain price is demanded, are evidence of bad faith'). Thus the Panel finds that Respondent registered and uses the disputed domain name in bad faith per Policy ¶ 4(b)(i)."

Figure 7: Excerpts of two factually similar disputes that led to different outcomes. The highlighted texts indicate the relevant facts regarding bad faith extracted from each dispute.

5.3.3 Results. We apply our clustering methodology to the generated summaries, yielding between 1,129 and 2,121 clusters per dispute requirement. Figure 8 presents illustrative examples. To assess clustering quality, we compute the average cosine similarity between all pairs of disputes within each cluster. This metric aligns with our downstream objective of forming *compact* groups of factually similar disputes, rather than maximizing separation between clusters. Across requirements, the mean within-cluster cosine similarity ranges from 87.92% to 95.73% (see Table 1 in the appendix), indicating that clustered disputes exhibit a high degree of semantic coherence. While this measure does not establish that clusters correspond to legally meaningful categories, it provides

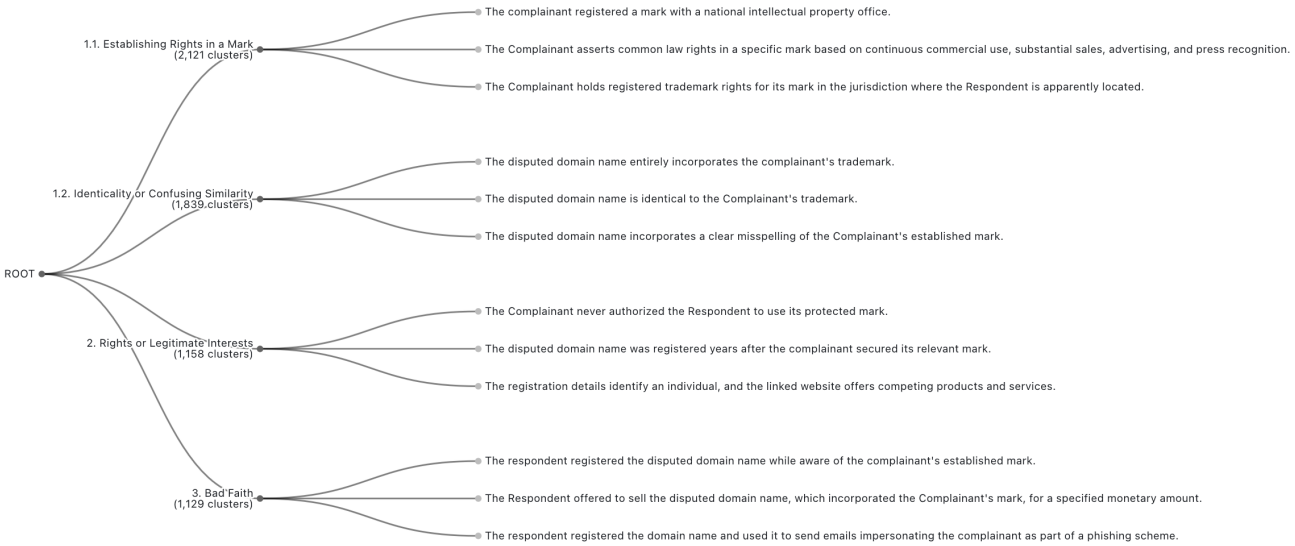

Figure 8: Top 3 largest clusters identified relative to each of the four dispute requirements analyzed. Those clusters represent ideas expressed in UDRP disputes filed between 2015 and 2025.

evidence that the clustering procedure consistently groups disputes with closely related factual descriptions.

Of the 6,247 semantic clusters identified, 1,084 (17.35%) contain both default and contested disputes, indicating that substantively similar factual circumstances were evaluated in both procedural settings. Within this subset, 676 (62.36%), covering 4,135 disputes, have identical respondent win rates regardless of participation status. In these cases, outcomes are consistent across default and contested proceedings when panels rely on the same underlying facts (see, for example, clusters (a) and (b) in Figure 9 where both groups of respondents either lost or won all their disputes).

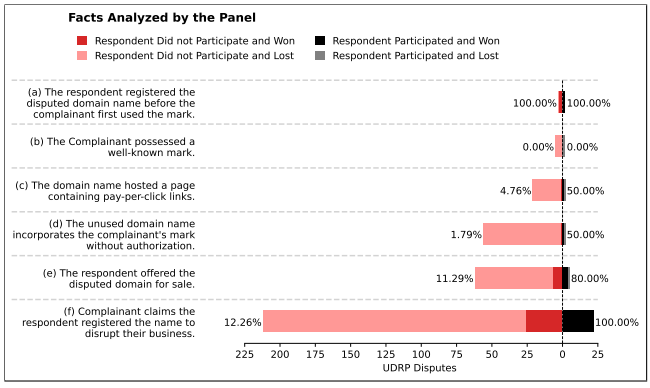

Figure 9: Differences in outcomes between default and contested disputes for a sample of representative clusters. The percentage value next to the bars represents the win rate for respondents in disputes assigned to that cluster.

We further identified 356 clusters comprising 4,497 disputes (577 contested and 3,920 default) in which participating respondents prevail more frequently than defaulting respondents. Across these clusters, respondents who participated won, on average, 79.50 percentage points more often than those who did not participate. Cluster (e) in Figure 9 illustrates such a scenario where respondents who participated won 4 of 5 (80.00%) disputes, while defaulting respondents won just 7 out of 62 (11.29%).

In 241 of the 356 clusters, participating respondents achieved a 100% win rate; yet, within these same factually comparable groups, defaulting respondents incurred 1,359 losses. Cluster (f) in Figure 9 is one such example, where participating respondents achieved a 100% success rate compared to just 12.26% (26/212) for those who defaulted. These findings document substantial divergence in outcomes associated with participation status when panels rely on similar factual grounds.

While the data do not establish the causes of these divergences, they highlight patterns that may inform ongoing discussions on participation incentives, procedural safeguards, education, and accuracy of dispute resolution outcomes.

5.3.4 External Validation. As an external point of validation, we examined whether disputes exhibiting outcome divergence in our analysis had attracted commentary in domain name–focused online forums. Specifically, we searched for discussions of the 1,359 disputes where defaulting respondents lost despite being factually similar to disputes won by participating respondents (for example, the 186 default losses in cluster (f) shown in Figure 9). The sources reviewed included CircleID [12] and Domain Name Wire [17], longstanding publications covering the domain name ecosystem, as well as the Internet Commerce Association [27], an advocacy organization representing domain name investors.

This search yielded 22 articles discussing panel decisions in 16 disputes. To characterize the tone of those articles, we apply a reasoning model to infer whether the authors agreed or disagreed with the panel outcomes. We found that commentators agreed with the panel decision in 12 disputes and disagreed in 4 four others. The presence of controversial outcomes among these commentaries provides independent confirmation that some of the divergences identified in our analysis are also noted by policy observers. At the same time, the limited number of publicly discussed cases (i.e., less than 2% of the 1,359 disputes) highlights the challenges of relying on anecdotal review and underscores the value of scalable, systematic methods like ours for examining UDRP decisions.

6 Recommendations

Overall, our findings highlight areas where existing procedures may warrant closer examination, particularly with respect to the handling of default disputes. One possible avenue for further consideration is the development of a systematic mechanism for reviewing a subset of panel decisions, with the goal of assessing the consistency and completeness of the reasoning provided across UDRP disputes. Such a review process could support multiple objectives, including identifying patterns in decision-writing practices, informing panelist training and guidance materials, and surfacing policy provisions that may benefit from clarification or refinement.

Recent commentary [48] has similarly noted the potential value of mechanisms for evaluating panelist performance, while also observing the absence of a standardized framework for conducting such assessments. In this context, the methods introduced in this work could be viewed as one input into a broader review process, for example, by helping to flag decisions whose reasoning differs substantially from that observed in factually similar cases, and which may merit additional scrutiny.

Our results also point to the role of respondent participation as an empirical factor associated with dispute outcomes. As noted earlier, while the causes of non-participation are likely heterogeneous, prior work suggests that some respondents default due to limited time and resources, lack of effective notice, or absence of familiarity with the process [49]. For these cases, UDRP policymakers may wish to explore approaches aimed at lowering participation barriers, such as clearer guidance on evidentiary expectations or tools that assist respondents in preparing submissions. Emerging AI-assisted systems could potentially support these goals by helping respondents identify relevant evidence and structure their responses, without altering the substantive standards applied by panels.

More broadly, given the observed differences in outcomes associated with legal representation and participation, initiatives that improve accessibility and procedural understanding may contribute to more consistent decision-making and increased confidence in the dispute resolution process. Any such measures would benefit from careful evaluation to balance efficiency, fairness, and the diverse interests of stakeholders within the domain name ecosystem.

7 Limitations

Our analysis is subject to several limitations. First, while semantic clustering allows us to perform comparisons across large numbers of disputes, similarity in extracted factual excerpts does not guarantee full legal equivalence across cases; panels may weigh facts differently or consider contextual elements not captured in the excerpts. Second, although we use state-of-the-art automated text extraction and summarization techniques — that are further validated through multiple checks — these techniques may omit nuances [7, 29] present in narrative reasoning. Third, clustering quality is evaluated using distributional similarity metrics that assess coherence rather than legal correctness, and therefore should be interpreted as indicative rather than definitive. In particular, BERTScore does not directly measure semantic completeness or factual faithfulness.

More broadly, this study examines associations between participation status, reasoning depth, and outcomes, but does not identify causal mechanisms, which we leave to future work. And, our analysis is limited to single-domain disputes decided in English, excluding cases that may exhibit different patterns. Extending this approach to multi-domain and non-English proceedings remains an important direction for future research, especially for settings where current techniques require new text representations [39]. We nevertheless hope these findings are helpful to the broader community and help inform ongoing discussions of procedural design.

8 Conclusion

Differences in outcomes between default and contested cases have long been a topic of discussion within the domain name community, particularly with respect to the operation of existing dispute resolution mechanisms. In this work, we examine whether panel decisions are consistent when disputes involve substantively similar factual circumstances, independent of respondent participation. Using scalable natural language processing techniques, we analyze large collections of UDRP decisions to compare outcomes and decision rationales across procedural settings. Our analysis indicates that respondent outcomes vary systematically with participation and representation status, that default decisions tend to cite fewer factual grounds, and that disputes presenting similar facts can nonetheless result in different outcomes. These observations provide empirical context for ongoing ICANN and community-led policy development efforts by documenting how participation and decision-writing practices are associated with observed outcomes.

We view our findings as important input to broader discussions on procedural safeguards, transparency, and the effective operation of the UDRP, rather than as prescriptive conclusions regarding the requirement for any specific policy changes.

References

[1] Griffin Adams, Alex Fabbri, Faisal Ladhak, Eric Lehman, and Noémie Elhadad. 2023. From sparse to dense: GPT-4 summarization with chain of density prompting. In *Proceedings of the 4th New Frontiers in Summarization Workshop*. Association for Computational Linguistics, Singapore, 68–74.

[2] Vinny Adjibi, Athanasios Avgetidis, Manos Antonakakis, Alberto Dainotti, Michael Bailey, and Fabian Monrose. 2026. Repairing trust in domain name disputes practices: Insights from a quarter-century's worth of squabbles. In *Proceedings of the 33rd Annual Network and Distributed System Security Symposium*. Internet Society, San Diego, CA, USA, 18 pages.

[3] Moses Ananta, Rahayu Utari, Amany Akhyar, and Gusti Ayu Putri Saptawati. 2024. Integrating BERTopic and large language models for thematic identification of Indonesian legal documents. In *2024 11th International Conference on Advanced Informatics: Concept, Theory and Application (ICAICTA)*. IEEE, Singapore, 1–6.

[4] Claire Barale, Michael Rovatsos, and Nehal Bhuta. 2025. *When fairness isn't statistical: The limits of machine learning in evaluating legal reasoning.*

[5] Amy Bender, Simon A. Maeder, and Megan Kirk. 2003. *UDRP opinion guide*. Retrieved Jan 04, 2026 from https://cyber.harvard.edu/udrp/opinion/

[6] L. Karl Branting, Craig Pfeifer, Bradford Brown, Lisa Ferro, John Aberdeen, Brandy Weiss, Mark Pfaff, and Bill Liao. 2021. Scalable and explainable legal prediction. *Artificial Intelligence and Law* 29, 2 (2021), 213–238.

[7] Christian Braun, Alexander Lilienbeck, and Daniel Mentjukov. 2025. *The hidden structure – improving legal document understanding through explicit text formatting*.

[8] Umar Butler, Abdur-Rahman Butler, and Adrian Lucas Malec. 2025. *The massive legal embedding benchmark (MLEB)*.

[9] Ricardo J. G. B. Campello, Davoud Moulavi, and Joerg Sander. 2013. Density-based clustering based on hierarchical density estimates. In *Advances in Knowledge Discovery and Data Mining*. Springer, Berlin, Heidelberg, 160–172.

[10] Daniel L. Chen and Jess Eagel. 2017. Can machine learning help predict the outcome of asylum adjudications?. In *Proceedings of the 16th Edition of the International Conference on Articial Intelligence and Law (ICAIL '17)*. Association for Computing Machinery, New York, NY, USA, 237–240.

[11] Daniel L. Chen and Markus Loecher. 2025. Mood and the malleability of moral reasoning: the impact of irrelevant factors on judicial decisions. *Journal of Behavioral and Experimental Economics* 116 (2025), 1–11.

[12] CircleID. 2026. *CircleID*. Retrieved Jan 15, 2026 from https://circleid.com/

[13] Derrick L. Cogburn, Theodore Andrew Ochieng, and Haiman M. Wong. 2023. Towards an understanding of global 'private ordering' in ICANN: Text mining 23 years of Uniform Domain-Name Dispute-Resolution Policy (UDRP) decisions. *Journal of Cyber Policy* 8, 2 (2023), 186–217.

[14] Wikipedia contributors. 2026. *Nissan Motors v. Nissan Computer*. Retrieved Dec 30, 2025 from https://en.wikipedia.org/wiki/Nissan_Motors_v._Nissan_Computer

[15] Goutam Datta, Nisheeth Joshi, and Kusum Gupta. 2022. Analysis of automatic evaluation metric on low-resourced language: BERTScore vs BLEU score. In *Speech and Computer*. Springer International Publishing, Cham, 155–162.

[16] Krish Didwania, Durga Toshniwal, and Amit Agarwal. 2024. Unveiling themes in judicial proceedings: A cross-country study using topic modeling on legal documents from India and the UK. In *Joint Ontology Workshops (JOWO)*. CEUR-WS, Enschede, The Netherlands., 12 pages.

[17] Domain Name Wire. 2026. *Domain Name Wire*. Retrieved Jan 15, 2026 from https://domainnamewire.com/

[18] M. Scott Donahey, Tomlinson Zisko Morosoli, and L. L. P. Maser. 2001. The UDRP: Fundamentally fair, but far from perfect. *Electronic Commerce and Law Reports* 6 (2001), 6 pages.

[19] Assafa Endeshaw. 2000. The threat of domain names to the trademark system. *Journal of World Intellectual Property* 3, 3 (2000), 323–342.

[20] Michael Geist. 2002. Fair.com: An examination of the allegations of systemic unfairness in the ICANN UDRP. *Brooklyn Journal of International Law* 27, 3 (2002), 903–938.

[21] Gemini Team. 2025. *Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities*.

[22] Maarten Grootendorst. 2022. *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*.

[23] Michael Hanna and Ondřej Bojar. 2021. A fine-grained analysis of BERTScore. In *Proceedings of the Sixth Conference on Machine Translation*. Association for Computational Linguistics, Online, 507–517.

[24] ICANN. 2015. *New generic top-level domains 2012 program*. Retrieved Jan 13, 2026 from https://newgtlds.icann.org/en

[25] ICANN. 2024. *Registrar accreditation agreement*. Retrieved Dec 21, 2025 from https://itp.cdn.icann.org/en/files/accredited-registrars/registrar-accreditation-agreement-21jan24-en.pdf

[26] ICANN. 2024. *Uniform Domain Name Dispute Resolution Policy*. Retrieved Jan 29, 2026 from https://www.icann.org/resources/pages/policy-2024-02-21-en

[27] Internet Commerce Association. 2026. *Home - InternetCommerce.org*. Retrieved Jan 15, 2026 from https://internetcommerce.org/

[28] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of tricks for efficient text classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*. Association for Computational Linguistics, Valencia, Spain, 427–431.

[29] Adrielson Ferreira Justino, Antônio Fernando Lavareda Jacob Junior, and Fábio Manoel França Lobato. 2025. A comparative study of BERT models for semantic retrieval of Brazilian legal precedents. In *Symposium on Knowledge Discovery, Mining and Learning*. Sociedade Brasileira de Computação, Brazil, 65–72.

[30] Jay P. Kesan and Andres A. Gallo. 2004. The market for private dispute resolution services – an empirical re-assessment of ICANN-UDRP performance. *Michigan Telecommunications & Technology Law Review* 285, 11 (July 2004), 285–380.

[31] Konstantinos Komaitis. 2005. Pandora's box is finally opened: The uniform domain name dispute resolution process and arbitration. *International Review of Law, Computers & Technology* 19, 1 (2005), 99–116.

[32] Jessica Litman. 2000. The DNS wars: Trademarks and the internet domain name system essay. *Journal of Small and Emerging Business Law* 4, 1 (2000), 149–166.

[33] Nelson F. Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. 2024. Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics* 12 (2024), 157–173.

[34] Pavel Loutocky. 2015. Are we getting good decisions by top-level domain name dispute resolution providers? *Masaryk University Journal of Law and Technology* 9, 1 (2015), 111–128.

[35] Milton Mueller. 2001. Rough justice: A statistical assessment of ICANN's uniform dispute resolution policy. *The Information Society* 17, 3 (2001), 151–163.

[36] Milton Mueller. 2002. *Success by default: A new profile of domain name trademark disputes under ICANN's UDRP*. Syracuse University School of Information Studies.

[37] Sebastiano Antonio Piccolo, Panagiota Katsikouli, Thomas Gammeltoft-Hansen, and Tijs Slaats. 2023. On predicting and explaining asylum adjudication. In *Proceedings of the 19th International Conference on Artificial Intelligence and Law (ICAIL '23)*. Association for Computing Machinery, New York, NY, USA, 217–226.

[38] Alpana Roy and Althaf Marsoof. 2016. A critical and comparative review of auDRP and UDRP domain name decisions. *The Journal of World Intellectual Property* 19, 5-6 (2016), 203–237.

[39] Daniel da Silva Junior, Daniel de Oliveira, and Aline Paes. 2025. Evaluating text representations for unsupervised legal semantic textual similarity in Brazilian Portuguese. *Discover Data* 3, 1 (2025), 23 pages.

[40] David A. Simon. 2012. An empirical analysis of fair use decisions under the uniform domain-name dispute-resolution policy. *Boston College Law Review* 53, 1 (2012), 65–130.

[41] Ian L. Stewart. 2001. The best laid plans: How unrestrained arbitration decisions have corrupted the uniform domain name dispute resolution policy note. *Federal Communications Law Journal* 53, 3 (2001), 509–532.

[42] Wayan Oger Vihikan, Meladel Mistica, Inbar Levy, Andrew Christie, and Timothy Baldwin. 2021. Automatic resolution of domain name disputes. In *Proceedings of the Natural Legal Language Processing Workshop 2021*. Association for Computational Linguistics, Punta Cana, Dominican Republic, 228–238.

[43] Voyage AI. 2025. *Voyage-3-Large: The new state-of-the-art general-purpose embedding model*. Retrieved Jan 04, 202 from https://blog.voyageai.com/2025/01/07/voyage-3-large/

[44] Robert A. Wagner and Michael J. Fischer. 1974. The string-to-string correction problem. *J. ACM* 21, 1 (Jan. 1974), 168–173.

[45] Yuzhong Wang, Chaojun Xiao, Shirong Ma, Haoxi Zhong, Cunchao Tu, Tianyang Zhang, Zhiyuan Liu, and Maosong Sun. 2021. *Equality before the law: Legal judgment consistency analysis for fairness*.

[46] WIPO. 2017. *WIPO overview of WIPO panel views on selected UDRP questions, third edition ("WIPO overview 3.0")*. Retrieved Dec 31, 2025 from https://www.wipo.int/amc/en/domains/search/overview3.0/

[47] World Intellectual Property Organization. 2025. *How to protect a trademark?* Retrieved Jan 26, 2026 from https://www.wipo.int/en/web/trademarks/protection

[48] World Intellectual Property Organization and Internet Commerce Association. 2025. *Final report of the WIPO-ICA udrp review project team*. Last accessed on Jan 10, 2026.

[49] Monika Zalnieriute. 2020. Reinvigorating human rights in internet governance: the UDRP procedure through the lens of international human rights principles. *The Columbia Journal of Law & the Arts* 43, 2 (Feb. 2020), 197–235.

[50] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. *BERTScore: evaluating text generation with BERT*.

A Clustering Evaluation

Table 1 presents the evaluation results of the clusters generated for the four dispute requirements. The intra-cluster similarity score represents the average of the pairwise cosine similarity between every pair of elements in each cluster.

Table 1: Results of the clustering algorithm

#	Requirement	Clusters	Similarity (%)
1.1.	Establishing Rights in a Mark	2,121	95.73%
1.2.	Identicality or Confusing Similarity	1,839	95.96%
2.	Rights or Legitimate Interests	1,158	91.29%
3.	Bad Faith	1,129	87.92%