

From Human Mapping to Machine Embedding: Uncovering Key Legal Drivers and Deterrents of ISDS Filing Frequencies

Sangchul Park*

Abstract

International investment agreements (IIAs), while intended to prop cross-border investment, have faced persistent criticism for potentially undermining the regulatory sovereignty of developing countries. Various mechanisms have been proposed as alternatives to traditional bilateral investment treaty (BIT) models, often with the goal of curbing investor-state dispute settlement (ISDS) filings. While existing research has uncovered the impact of nonlegal factors, such as macroeconomic crises, little has been done to systematically examine how legal provisions in either major model BITs or ISDS reform toolboxes influence ISDS filing patterns. To address this gap, this Article analyzes the interplay between (i) legal texts of 2,148 BITs and treaties with investment provisions (TIPs) and (ii) the occurrence of 1,060 ISDS cases. It builds on the United Nations Conference on Trade and Development (UNCTAD)'s IIA Mapping Project to assess the impact of key legal deterrents recommended by ISDS reform proponents, while leveraging large language models to identify the key legal drivers of ISDS filings. The outcome of Poisson regression appears to reveal that: (i) procedural provisions resembling those in the 2012 U.S. Model BIT are the strongest positive predictors of ISDS filings, outweighing the impact of economic crises, whereas substantive provisions such as investor treatment and expropriation clauses are not; (ii) the effectiveness of deterrent provisions remains inconclusive, suggesting that their ability to curb ISDS filings requires further scrutiny; and (iii) the assumption that IIAs between developed host countries and developing states are more prone to ISDS filings is unsubstantiated. These findings could contribute to ongoing discussions on BIT reform by highlighting the legal determinants of ISDS frequencies, with implications for policymakers seeking to balance investment protection with regulatory autonomy.

* Associate Professor, Seoul National University School of Law. This paper was funded by the 2023 Research Fund of the SNU Law Research Institute, donated by the SNU Law Foundation. I thank the CJIL Editorial Board members for organizing the symposium and for their thoughtful editing of this paper, including Mr. Ian G. Peacock's suggestion of zero-inflated models. This research project has benefitted from the Microsoft Accelerating Foundation Models Research (AFMR) grant program.

Table of Contents

I.	Introduction.....	221
II.	Methodology.....	223
	A. Data Collection.....	224
	B. Preprocessing.....	225
	C. Model	228
III.	Outcomes.....	231
	A. Outcome of Poisson Regression.....	231
	B. Robustness Checks	233
	C. Implications.....	235
	D. Limitations.....	236
IV.	Conclusion	236

I. INTRODUCTION

International investment agreements (IIAs) have remained a cornerstone of international investment protection, proliferating with at least 2,844 bilateral investment treaties (BITs) and 480 treaties with investment provisions (TIPs) as of March 1, 2025.¹ However, IIAs and their Investor-State Dispute Settlement (ISDS) mechanisms have faced persistent criticism for potentially undermining governments' regulatory sovereignty, particularly during economic or social crises, while lacking sufficient transparency and predictability. In response, the U.N. Commission on International Trade Law (UNCITRAL)'s Working Group III has been drafting proposals for ISDS reform,² while reputable institutions—including Chatham House, the Columbia Center on Sustainable Investment, and the Center for the Advancement of the Rule of Law in the Americas at Georgetown—have published reports and proposed toolkits addressing ISDS reform.³

While these efforts explore systemic changes, their specific recommendations frequently focus on curbing ISDS filings both substantially and procedurally.⁴ One reason is that a high volume of ISDS filings not only signals deeper systemic issues in the drafting or interpretation of IIAs but also contributes to regulatory chill and places significant resource burdens on states—particularly developing countries.⁵ Despite this focus on the frequency of ISDS filings,

¹ *Investment Policy Hub, International Investment Agreements Navigator*, UNCTAD <https://perma.cc/R9WD-HXNP> (last visited Mar. 1, 2025).

² UNCITRAL, *Possible Reform of Investor-State Dispute Settlement (ISDS): Draft multilateral Instrument on ISDS Reform*, U.N. Doc A/CN.9/WG.III/WP.246 (2025), <https://perma.cc/2P9A-NNV7>.

³ LAUGE N. SKOVGAARD POULSEN & GEOFFREY GERTZ, *REFORMING THE INVESTMENT TREATY REGIME: A 'BACKWARD-LOOKING' APPROACH* (2021), <https://perma.cc/2746-S7QV>; Nick Robins, *The Emergence of Sustainable Investing*, in *SUSTAINABLE INVESTING: THE ART OF LONG TERM PERFORMANCE* 3–18 (Cary Kronisky & Nick Robins eds., 2008). NAZLY D. GOMEZ & DANIEL R. JURADO, *NEW DIRECTIONS IN INTERNATIONAL INVESTMENT LAW: ALTERNATIVES FOR IMPROVEMENT* (2021), <https://perma.cc/7U7C-CZMZ>.

⁴ GOMEZ & JURADO, *supra* note 3, at 14–19 (proposing, as means of ISDS reform, procedural solutions such as (i) exhaustion of local remedies, (ii) fork in the road, (iii) counterclaims, (iv) keyholes, (v) regulation of third-party funding, and (vi) dismissal of frivolous claims, as well as substantive solutions such as (i) reduction of the reach of obligations and (ii) elimination of obligations). *See also* SKOVGAARD & GERTZ, *supra* note 3, at 1–8 (criticizing policymakers' emphasis on constraining or replacing ISDS only in future IIAs and proposing for alternative approaches, such as a plurilateral mechanism for “interpretative statements,” in which governments jointly clarify and define their positions on contentious clauses in their existing IIAs).

⁵ For instance, (i) in 2019, Pakistan was required to pay USD 6 billion in compensation to an Australian mining company—an amount equivalent to the IMF bailout it received that year. *See Tethyan Copper Company Pty Limited v. Pakistan*, ICSID Case No. ARB/12/1, Award, 12 July 2019; Recited from SKOVGAARD & GERTZ, *supra* note 3, at 2; (ii) as of March 2025, Honduras is reportedly facing ISDS claims totaling USD 18 billion—some linked to energy transition—an amount that surpasses the country's entire annual budget. *See* Phoebe Western & Patrick Greenfield,

research remains sparse on whether and how legal provisions in major model BITs or ISDS reform toolkits influence ISDS filing patterns.⁶ Notably, a study of the frequency of ISDS filings focuses on macroeconomic factors; Bellak and Leibrecht's 2021 paper—whose insight this paper draws upon—identifies economic crises and the number of previous ISDS filings by investors in a specific host country as key determinants of ISDS filing frequencies.⁷ Berge's 2020 paper, meanwhile, finds that substantive obligations—including national treatment (NT), most-favored-nation treatment (MFN), fair and equitable treatment (FET), and compensation for expropriation—are the only significant legal predictor of ISDS occurrences, while greater flexibility or precision in treaty language does not significantly affect ISDS risks.⁸ Rather than assess the impact of specific substantive or procedural provisions, this study categorizes mapped IIA provisions into broader attributes and arrives at a conclusion that contrasts with Berge's 2020 finding.

In this context, this study analyzes the interplay between (i) the legal texts of 2,148 BITs and treaties with investment provisions that came into effect before 2019 and (ii) 1,060 ISDS cases filed before 2020, taking into account the availability of economic crisis data, one of the most critical control variables, up to 2019. A key challenge in this analysis is how to quantify treaty provisions. Significant efforts have been made to manually map IIA provisions, as exemplified by the U.N. Conference on Trade and Development (UNCTAD)'s IIA Mapping Project.⁹ This dataset has been instrumental in prior studies, including Berge's

Why Fear of Billion-Dollar Lawsuits Stops Countries Phasing Out Fossil Fuels, THE GUARDIAN (Mar. 6, 2025), <https://perma.cc/RGE9-5JW3>.

⁶ The empirical literature on ISDS has instead focused on (i) the impact of ISDS provisions on the amount of foreign directive investment. See Eric Neumayer & Laura Spess, *Do Bilateral Investment Treaties Increase Foreign Direct Investment to Developing Countries?*, 33 WORLD DEV. 1567 *passim* (2005); Peter Egger & Michael Pfaffermayr, *The Impact of Bilateral Investment Treaties on Foreign Direct Investment*, 32 J. COMP. ECON. 788 *passim* (2004); Andrew Kerner, *Why Should I Believe You? The Costs and Consequences of Bilateral Investment Treaties*, 53 INT. STUD. Q. 73 *passim* (2009); (ii) the impact of ISDS filings on IIA practices see Lauge N.S. Poulsen & Emma Aisbett, *When the Claim Hits. Bilateral Investment Treaties and Bounded Rational Learning*, 65 WORLD POLITICS 273 *passim* (2013); and (iii) determinants of ISDS outcomes see Daniel Behn, Tarald L. Berge & Malcolm Langford, *Poor States or Poor Governance? Explaining Outcomes in Investment Treaty Arbitration*, 38 NW. J. INT'L L. & BUS. 333 *passim* (2018); Julian Donaubauer, Eric Neumayer & Peter Nunnenkamp, *Winning or Losing in Investor-to-State Dispute Resolution: The Role of Arbitrator Bias and Experience*, 26(4) REV. INT'L ECON. 892 *passim* (2018).

⁷ Christian Bellak & Markus Leibrecht, *Do Economic Crises Trigger Treaty-Based Investor-State Arbitration Disputes?*, 24 J. INT'L ECON L. 127 *passim* (2021).

⁸ Tarald L. Berge, *Dispute by Design? Legalization, Backlash, and the Drafting of Investment Agreements*, 64 INT'L STUD. QUARTERLY 919 *passim* (2020).

⁹ UNCTAD, *supra* note 1.

2020 paper mentioned above and Thompson et al.'s 2019 study, which explores how actual ISDS cases influence calls for IIA modification or termination.¹⁰

While this Article also leverages UNCTAD's mapping data, it introduces novel analytical tools, incorporating the embedding capabilities of large language models (LLMs) to gain deeper insights into the legal drivers of ISDS filings. There are previous studies that analyzed the texts of IIAs. Notably, Baccini et al.'s 2015 study used cluster analysis to classify IIAs into three categories: the EU model, the NAFTA model, and the Southern model.¹¹ Allee et al.'s 2014 study found that stronger enforcement provisions in BITs correlate more with the bargaining power of the capital-exporting treaty partner than with the initiative of investment-seeking states.¹² Unlike the previous studies, which relied on manual coding, this study leverages the analytical capabilities of LLMs to engineer features that capture the substantive and procedural dimensions of IIA provisions. Poisson regression analysis, encompassing the newly engineered features, indicates that: (i) procedural provisions resembling those in the 2012 U.S. Model BIT are the most significant positive predictors of ISDS filings, surpassing the influence of economic crises and other variables, while substantive provisions do not show clear impact, contrary to Berge's findings; (ii) the effectiveness of deterrent provisions remains uncertain, suggesting that their role in reducing ISDS filings requires further investigation; and (iii) the widely held belief that IIAs between developed host countries and developing states are more susceptible to ISDS filings is not demonstrated.

II. METHODOLOGY

This Article seeks to identify the legal drivers and deterrents influencing the frequency of ISDS filings. To achieve this, it employs a multi-step approach: (i) extracting treatment, expropriation, and procedural provisions from IIAs; (ii) transforming these provisions into embeddings using LLMs and quantifying their similarity to the U.S. Model BIT 2012; and (iii) incorporating the resulting similarity scores—alongside other variables such as macroeconomic indicators and ISDS deterrent provisions—into a Poisson regression model designed to explain the occurrence of ISDS filings. To ensure replicability, a code script and dataset are publicly available at <https://github.com/replicable/isds>.

¹⁰ Alexander Thompson, Tomer Broude & Yoram Z. Haftel, *Once Bitten, Twice Shy? Investment Disputes, State Sovereignty, and Change in Treaty Design*, 73 INT'L. ORG. 859 *passim* (2019).

¹¹ Leonardo Baccini, Andreas Dür & Yoram Haftel, *Imitation and Innovation in International Governance: The Diffusion of Trade Agreement Design*, in TRADE COOPERATION: THE PURPOSE, DESIGN AND EFFECTS OF PREFERENTIAL TRADE AGREEMENTS (Andreas Dür & Manfred Elsig eds., 2015).

¹² Todd Allee & Clint Peinhardt, *Evaluating Three Explanations for the Design of Bilateral Investment Treaties*, 66 WORLD POLITICS 47 *passim* (2014).

A. Data Collection

As noted, the UNCTAD Investment Policy Hub provides access to the IIA Navigator, which includes 2,844 BITs and 480 TIPs, totaling 3,194 agreements. From this dataset, I extracted 2,148 effective treaties after applying the following exclusions: (i) 108 treaties that are not yet in force (terminated treaties, however, are included); (ii) 17 overlapping samples, primarily related to the dissolution of the Soviet Union; (iii) 151 IIAs that came into effect after 2019, as I failed to obtain consistent economic crisis data after this period; (iv) 749 IIAs lacking UNCTAD's mapping data (however, I manually mapped 11 IIAs that had triggered five or more ISDS cases but were omitted in the UNCTAD's database, including the North American Free Trade Agreement (NAFTA) and the Energy Charter Treaty (ECT), and included them in the dataset); and (v) 146 IIAs without ISDS provisions, including 12 treaties incorrectly coded by UNCTAD as containing ISDS provisions and 10 treaties for which the presence of ISDS provisions cannot be definitively verified due to the unavailability of an English version. Additionally, the UNCTAD Investment Policy Hub offers the Investment Dispute Settlement Navigator that provides information on 1,332 known treaty-based ISDS cases,¹³ from which I extracted 1,060 cases which were filed before 2020.

For these samples, UNCTAD mapping data is available. However, to conduct a more detailed analysis of substantive provisions—including NT, MFN, FET, and expropriation as well as procedural ISDS provisions, I developed and implemented an automated method to identify relevant treaty articles. While this approach involved a keyword search and cosine similarity of embeddings, initially focusing on article headings, and when unavailable, analyzing paragraph content,¹⁴ I manually reviewed and corrected misclassifications to sort out irrelevant articles. While the UNCTAD Investment Policy Hub provides scanned treaty versions, these often trigger high levels of optical character recognition errors. To overcome

¹³ UNCTAD, *supra* note 1.

¹⁴ Cosine similarity (also known as the normalized dot product) between two vectors in an embedding space measures their semantic closeness, typically yielding higher values when the vectors have large components in the same dimensions, which indicates that they have similar meanings. *See* DANIEL JURAFSKY & JAMES H. MARTIN, *SPEECH AND LANGUAGE PROCESSING* 110–11 (3d ed. 2025), <https://perma.cc/R27F-ZC6N> (last visited Apr. 27, 2027). The cosine similarity between two vectors v and w can be computed as:

$$\text{cosine}(v, w) = \frac{v \cdot w}{|v||w|} = \frac{\sum_{i=1}^N v_i w_i}{\sqrt{\sum_{i=1}^N v_i^2} \sqrt{\sum_{i=1}^N w_i^2}}$$

This metric is particularly useful for retrieving specific provisions from lengthy agreements based on their semantic relationship to keywords such as treatment, expropriation, and arbitration.

this issue, my code matched and extracted text file versions from the Electronic Database of Investment Treaties (EDIT), managed by the World Trade Institute.¹⁵

B. Preprocessing

From each IIA, three key provisions are extracted: (i) treatment of investors, consolidated in the order of FET, NT, and MFN¹⁶; (ii) expropriation; (iii) procedures for submission of an ISDS claim to arbitration, including consent to arbitration and conditions and limitations on consent. Other substantive obligations, such as free transfer and prohibition of performance requirements, were excluded due to their relative homogeneity and lesser significance in ISDS disputes. Additionally, while procedural provisions related to the constitution of arbitral panels and the conduct of arbitration may also be significant, they were excluded from this analysis due to (i) their absence in many IIAs and (ii) the technical challenges of reliably and consistently isolating them from single-article, all-inclusive ISDS provisions within certain IIAs found in many IIAs. To minimize potential bias from this omission, the analysis focuses on comparing relevant procedural content with Article 24 (Submission of a Claim to Arbitration), Article 25 (Consent of Each Party to Arbitration), and Article 26 (Conditions and Limitations on Consent of Each Party) of the 2012 U.S. Model BIT, while excluding other procedural articles such as Article 27 (Selection of Arbitrators), Article 28 (Conduct of the Arbitration), and Article 29 (Transparency of Arbitral Proceedings).

Each IIA's three provisions were transformed into contextual embeddings for numerical analysis of their semantic meaning. This process employs OpenAI's GPT-3.5 vector embedding model, text-embedding-3-small, which produces 1,536-dimensional embeddings from textual inputs.¹⁷ My code then calculates the cosine similarity of each of the three provisions to those in the 2012 U.S. Model BIT, which might stand for the sophistication of the provision or customary international practices.¹⁸ Cosine similarity between two vectors in an embedding

¹⁵ Wolfgang Alschner, Manfred Elsig & Rodrigo Polanco, *Introducing the Electronic Database of Investment Treaties (EDIT): The Genesis of a New Database and Its Use*, 20 *WORLD TRADE REV.* 73, 73–94 (2021).

¹⁶ FET provisions are often located in articles titled “Promotion and Protection of Investors” or similar. Such articles were included in the analysis, except when they exclusively address the facilitation and admission of investment without explicitly providing fair, equitable, or non-discriminatory treatment of investors.

¹⁷ An embedding is a vector of floating-point numbers. Vectors representing tokens—often morphemes that make up words—are placed (“embedded”) in a vector space such that the distance between two vectors reflects their relatedness: smaller distances indicate higher relatedness, while larger distances indicate lower relatedness. See *Vector Embeddings*, OPENAI, <https://perma.cc/9MCK-9H25> (last accessed Mar. 29, 2025).

¹⁸ See *Glamis Gold Ltd. v. United States*, UNCITRAL, Award ¶¶ 602–03 (June 8, 2009) (referencing model BITs as one of the “very few authoritative sources” for customary international practices).

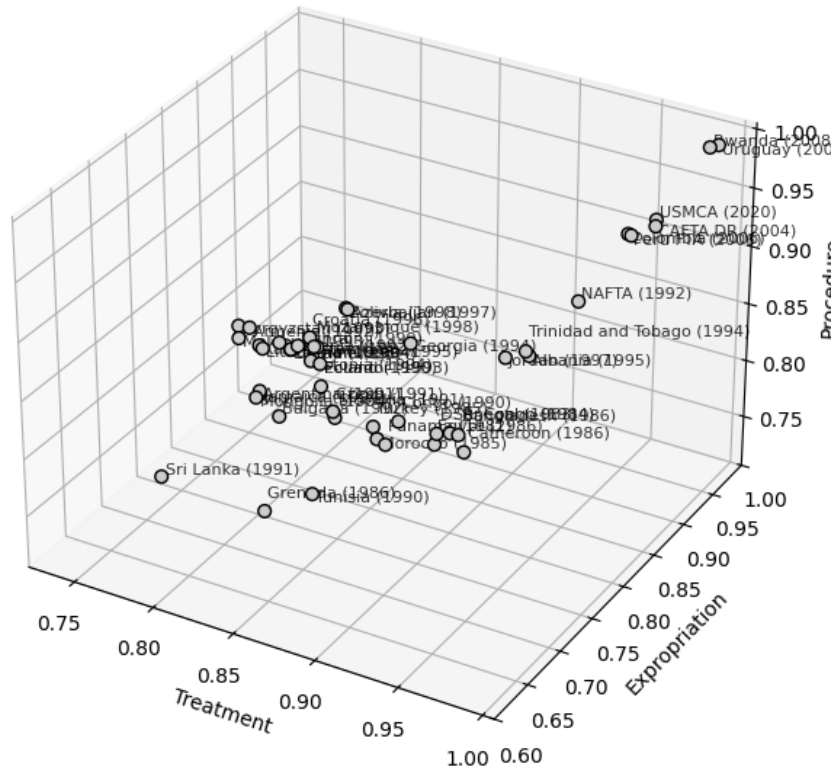
space reflects how often the corresponding tokens co-occur, indicating their semantic closeness. For sentence embeddings, the cosine similarity is computed as the weighted average of the similarities between the embeddings of the tokens that make up the sentence. This score is highly effective in capturing the semantic similarity between two provisions.

An example of this methodology—comparing NAFTA with the 2012 U.S. Model BIT—is presented in Table 1.

	NAFTA	2012 U.S. Model BIT	Similarity
Treatment	Article 1105 (Minimum Standard of Treatment) (<i>omitted</i>)	Article 5 (Minimum Standard of Treatment) (<i>omitted</i>)	0.940
	Article 1102 (National Treatment) (<i>omitted</i>)	Article 3 (National Treatment) (<i>omitted</i>)	
	Article 1103 (Most-Favored-Nation Treatment) (<i>omitted</i>)	Article 4 (Most-Favored-Nation Treatment) (<i>omitted</i>)	
Expropriation	Article 1110 (Expropriation and Compensation) (<i>omitted</i>)	Article 6 (Expropriation and Compensation) (<i>omitted</i>)	0.900
Procedure	Article 1116 (Claim by an Investor of a Party on Its Own Behalf) (<i>omitted</i>)	Article 24 (Submission of a Claim to Arbitration) (<i>omitted</i>)	0.875
	Article 1120 (Submission of a Claim to Arbitration) (<i>omitted</i>)	Article 25 (Consent of Each Party to Arbitration) (<i>omitted</i>)	
	Article 1122 (Consent to Arbitration) (<i>omitted</i>)	Article 26 (Conditions and Limitations on Consent of Each Party) (<i>omitted</i>)	

[TABLE 1] COMPARING NAFTA TO THE 2012 U.S. MODEL BIT

Figure 1 illustrates how the embedding-based similarity scores work for IIAs that the U.S. has entered into, when the similarity scores for treatment, expropriation, and procedure are used as the x , y , and z coordinates, respectively. The U.S.–Uruguay BIT (2005) and the Rwanda–U.S. BIT (2008) exhibit particularly high levels of similarity (98.7% and 98.5% on average), whereas the U.S.–Sri Lanka BIT (1991) shows a lower level of similarity (78.5% on average). Notably, the United States–Mexico–Canada Agreement (USMCA) (2020) (based on the U.S.–Mexico terms) and NAFTA (1992) exhibit average similarity scores of 94.93% and 90.50%, respectively.



[FIGURE 1] THE SIMILARITY OF ACTUAL U.S. BITS TO THE 2012 U.S. MODEL BIT

The code puts all the 3,194 IIAs and applies k -means clustering to provide insight into the convergence and divergence of IIAs in the global investment landscape. Figure 2 presents the results of k -means clustering, showing that IIAs are effectively clustered into three groups,¹⁹ along three axes: similarity to the 2012 U.S. Model BIT in terms of treatment, expropriation, and procedure. Group 1 consists of modern FTAs that are heavily influenced by NAFTA, whereas Group 3 includes IIAs that diverge significantly from NAFTA in terms of investor treatment and Group 2 is in the mezzanine. The samples most adjacent to centroids of Group 1, 2, 3 – Colombia – U.K. BIT (2010), China – U.K. BIT (1986), and France – Panama BIT (1982), respectively.

¹⁹ The silhouette score, which evaluates how well k -means clustering groups data points is highest (0.3862) when the number of clusters (k) is three. See Peter J. Rousseeuw, *Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis*, 20 COMPUT. APPL. MATH. 53 *passim* (1987). The scores are 0.2708, 0.2978, 0.2993, 0.2501, 0.2615, and 0.2553 for $k = 2, 4, 5, 10, 20, 50$, respectively.



A key hypothesis to test is that the frequency of ISDS filings in a given calendar year follows a Poisson distribution,²⁰ influenced by several categories of variables: $Crisis_{ti}$, $Party_{ti}$, $Terminate_{at}$, $USBITSim_a$, and $Deter_a$.

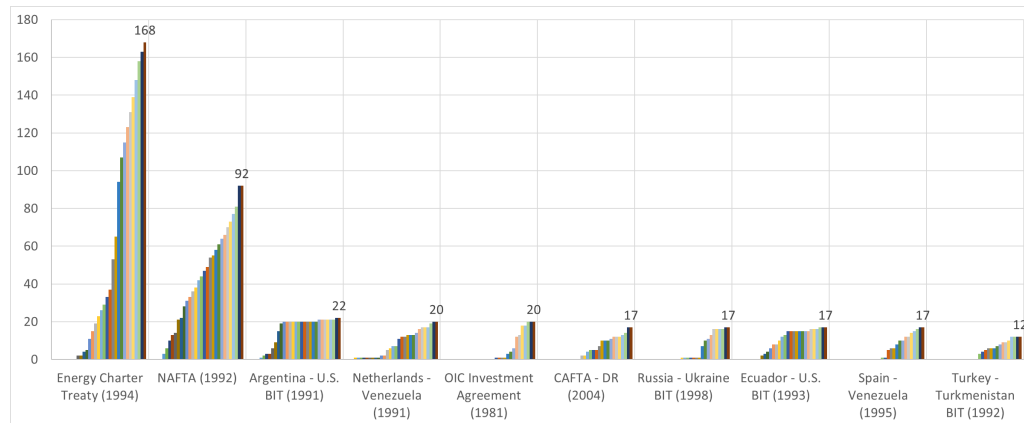
²⁰ J.A. Nelder, *Log Linear Models for Contingency Tables: A Generalization of Classical Least Squares*, 23(3) J. R. STAT. SOC., SERIES C (APPL. STAT.) 323 *passim* (1974).

$$\theta^T X_{ati} = \theta_0^T + \theta_1^T \text{Crisis}_{ti} + \theta_2^T \text{Party}_{ti} + \theta_3^T \text{Terminate}_{at} \\ + \theta_4^T \text{USBITSim}_a (+\theta_5^T \text{Deter}_a)$$

where α represents a specific IIA (out of 2,148 in total), t denotes a specific year between 1993 to 2019 (following the IIA's effective date), and i denotes a particular IIA. The dataset comprises 39,524 treaty-dyadic yearly samples, which comprise the following variables:

1. *Filing_{at}*: Dependent variable

This represents the number of ISDS filings for a specific IIA in a given year. Figure 3 presents the top 10 IIAs with the highest cumulative number of ISDS filings during 1993–2024.



[FIGURE 3] TOP 10 IIAS WITH THE HIGHEST NUMBER OF ISDS FILINGS

2. *Crisis_{ti}*: Economic crises

As noted, the selection of this variable is inspired by Bellak and Leibrecht (2021).²¹ However, unlike their study, which uses country-dyadic data, this analysis is conducted at the treaty level, as this Article matches ISDS cases with relevant IIAs, rather than with a specific country. For multilateral IIAs, *Crisis_{ti}* is coded as 1 if at least one of the participating parties is experiencing an economic crisis.

To capture economic crises, I utilized three dummies indicating banking, currency, and debt crisis coded by Nguyen et al. (2022), covering the period of

²¹ Bellak & Leibrecht, *supra* note 7, at 138–41.

1993 to 2019.²² Due to the lack of crisis data beyond 2019, IIAs that came into effect after that year were excluded from the analysis.

3. *Party_{ti}*: Other treaty party characteristics

This vector consists of two variables: (i) whether at least one party was a developed country and at least one party was a developing country in a given year, based on the International Monetary Fund's classification of advanced economies and emerging and developing economies²³; and (ii) the total number of ISDS claims filed by investors within the host country prior to the given year, which Bellak and Leibrecht (2021) also identified as one of the key factors affecting ISDS filing frequency.²⁴

4. *Terminate_{at}*: Years after termination of the treaty

Apart from IIAs that were replaced by another agreement, already terminated IIAs each year are assigned a value of 1 for this variable, rather than being entirely excluded from the sample.

5. *USBITSim_a*: Similarity to the 2012 U.S. Model BIT

This vector comprises three scores representing similarity to the 2012 U.S. Model BIT, in terms of treatment, expropriation, and the initiation of ISDS claims.

6. *Deter_a*: Deterrent provisions

This vector consists of dummy variables representing potentially deterrent provisions as coded by UNCTAD's IIA Mapping Project, including: (i) IIA scope exclusions (taxation, subsidies, procurement, others, and applicable only to post-BIT events), (ii) MFN exceptions (economic integration, taxation treaties, procedural issues, regulatory), (iii) expropriation exceptions (regulatory and compulsory licenses), (iv) social protections (health and environment, labor standards, right to regulate, corporate social responsibilities, and not lowering of standards), (vi) exception self-judging, (vii) ISDS scope exclusions (scope of claims, ISDS limitation, policy exclusion, taxation or prudential, limited remedies, and limiting amicus curiae), and (viii) ISDS filing restrictions (case-by-case consent,

²² Thanh C. Nguyen, Vitor Castro & Justine Wood, *A New Comprehensive Database of Financial Crises: Identification, Frequency, and Duration*, 108 ECON. MODEL. 1, 12–15 (2022).

²³ INT'L MONETARY FUND, WORLD ECONOMIC OUTLOOK DATABASE: GROUPS AND AGGREGATES INFORMATION (2023), <https://perma.cc/A6EK-NHAD> (last accessed Mar. 29, 2025).

²⁴ Bellak & Leibrecht, *supra* note 7, at 143–44 (“This variable captures the awareness of foreign investors from a country of the possibility to bring a case against a particular host country before an arbitration council.”).

fork in the road, no U turn, limitation period, local remedies first, and voluntary ADR). While UNCTAD provides a detailed codebook for each variable,²⁵ Appendix I provides its summary.

III. OUTCOMES

A. Outcome of Poisson Regression

The results of the Poisson regression predicting the counts of ISDS filings are presented in Table 2.

	MODEL WITHOUT <i>Deter_a</i>	MODEL WITH <i>Deter_a</i>
(INTERCEPT)	-9.3946547*** (< 2e-16)	-10.617567*** (< 2e-16)
CRISIS: BANKING	0.3477968*** (1.59e-05)	0.435705*** (1.96e-07)
CRISIS: CURRENCY	0.1575510 (0.204553)	0.243165. (0.052653)
CRISIS: DEBT	-0.0315204 (0.702595)	0.073466 (0.397282)
PARTY: DEVELOP	-0.2422656*** (0.000140)	0.001301 (0.986160)
PARTY: PRIOR CLAIMS	0.0039947*** (1.87e-05)	0.007070*** (1.58e-13)
TERMINATE	-0.0915554 (0.272499)	-0.103608 (0.224036)
USBITSIM: TREATMENT	-2.5288473*** (0.000794)	1.622045 (0.102204)
USBITSIM: EXPROPRIATION	-2.2438151*** (2.94e-08)	-2.757705*** (1.53e-05)
USBITSIM: PROCEDURE	13.7917405*** (< 2e-16)	9.964697*** (< 2e-16)
IIA: TAXATION		-0.223596* (0.042177)
IIA: SUBSIDIES		2.154914*** (< 2e-16)
IIA: PROCUREMENT		-2.159584*** (2.79e-15)
IIA: OTHERS		0.311248* (0.013689)
IIA: POST-BIT		0.548526*** (3.89e-06)
MFN: ECONOMIC		-0.843211*** (2.29e-11)
INTEGRATION		0.227083* (0.036309)
MFN: TAXATION		0.754468*** (3.14e-07)
MFN: PROCEDURAL ISSUES		-1.914092*** (8.68e-10)
EXPROPRIATION:		
REGULATORY		0.941618*** (2.05e-05)
EXPROPRIATION:		
COMPULSORY LICENSES		0.314018* (0.018089)
SOCIAL: HEALTH AND		
ENVIRONMENT		0.124098 (0.550898)
SOCIAL: LABOR STANDARDS		1.224714*** (< 2e-16)
SOCIAL: RIGHT TO REGULATE		

²⁵ UNCTAD, UNCTAD IIA MAPPING PROJECT, <https://perma.cc/3YE7-EU4Q> (last accessed Mar. 29, 2025).

SOCIAL: CORPORATE RESPONSIBILITY	-1.271816* (0.010943)
SOCIAL: NOT LOWERING STANDARDS	-0.295417 (0.201128)
SECURITY: EXCEPTION SELF-JUDGING	-0.257900 (0.195030)
ISDS SCOPE: TREATY CLAIMS	-0.103657 (0.207556)
ISDS SCOPE: LIMITATION	0.358940*** (0.000573)
ISDS SCOPE: POLICY	-0.044751 (0.800290)
ISDS SCOPE: TAXATION OR PRUDENTIAL	-0.199341 (0.289735)
ISDS SCOPE: LIMITED REMEDIES	0.731173*** (0.000930)
ISDS SCOPE: ALLOWING AMICUS CURIAE	0.928169* (0.017298)
ISDS FILING: CASE-BY-CASE CONSENT	0.106267 (0.472885)
ISDS FILING: FORK IN THE ROAD	-0.927350** (0.002669)
ISDS FILING: LIMITATION PERIOD	-1.071619*** (2.19e-06)
ISDS FILING: LOCAL REMEDIES FIRST	-0.206317 (0.237112)
ISDS FILING: VOLUNTARY ADR	0.244765** (0.001312)
<i>N</i>	39,524
AIC	9805.6
DISPERSION	1.823704
MAX VIF	1.3808 (USBITSIM: PROCEDURE)
	8.394632 (IIA: PROCUREMENT)
[TABLE 2] OUTCOMES OF POISSON REGRESSION PREDICTING COUNTS OF ISDS FILINGS ²⁶	

Overall, procedural provisions, prior claims, and banking crisis appear to be the strongest predictors of ISDS filings.

The similarity of procedural provisions to the 2012 U.S. Model BIT emerges as a major driver of ISDS filings, reinforcing the hypothesis that procedural elements in treaties play a crucial role in dispute occurrences. The coefficients of 13.792 and 9.965 for the procedural similarity score are notably high, translating to a 976,764- and 21,269-fold increase ($e^{13.792}$ and $e^{9.965}$) in ISDS filing frequencies,

²⁶ The coefficient of the model is to be interpreted as the increase in the logarithm of the count of ISDS claims when each independent variable increases by 1.

respectively. The primary reason may be that the variable's variance is significantly lower than that of other 0-or-1 dummy variables, as shown in Figure 2. Table 2 indicates that both models exhibit a decent degree of dispersion (1.8237 and 1.3442, respectively), suggesting that variance does not deviate from mean.

In contrast, expropriation provisions (and treatment provisions in the case of the model without *Deter_a*) exhibit a surprisingly negative correlation with ISDS filings, challenging prior literature that highlights substantive obligation standards as key determinants.

Among party-specific factors, prior ISDS claims serve as a predictor, indicating that countries with a history of ISDS cases are more likely to face future claims, though its coefficient is not particularly strong. Meanwhile, developed-developing party status has a surprisingly negative impact in the model without *Deter_a* and no significant impact in the model with *Deter_a*, undermining the assumption that IIAs between developed host countries and developing states are inherently more prone to ISDS filings.

The effectiveness of deterrent provisions in reducing ISDS claims remains inconclusive. While some provisions—such as procurement exceptions (-2.16), regulatory carve-outs for expropriation (-1.91), corporate social responsibility (CSR) (-1.27), filing limitation periods (-1.07), fork-in-the-road clauses (-0.93), and economic integration exceptions for MFN (-0.84)—are associated with lower ISDS risk, the broader impact of deterrent mechanisms is not clearly substantiated. In particular, subsidy/grant exceptions (+2.15), right to regulate (+1.22), exclusion of compulsory licenses from expropriation (+0.94), exclusion of ISDS provisions found in other treaties (+0.75), limited remedies (+0.73), post-BIT provisions (+0.55), etc. unexpectedly correlate with more ISDS filings, while they are designed to restrict ISDS filings or awards.

B. Robustness Checks

As noted, the coefficients of 13.792 and 9.965 for the procedural similarity score are notably high. However, as both models show a reasonable level of dispersion (1.8237 and 1.3442, respectively), these unusually high coefficients would not signal overdispersion—where the variance exceeds the mean—in the models. As such, it is not necessary to implement negative binomial distribution models.

A key robustness challenge lies in the overwhelming number of zero counts in ISDS filings: out of 39,524 treaty-dyadic yearly observations, 38,763 (98.1%) record no ISDS filings, while only 632, 73, and 28 observations record one, two, and three filings, respectively. This justifies the use of a zero-inflated Poisson (ZIP) regression model, which first uses a logistic regression to model the probability of

excess zeros and then applies a Poisson regression to model count outcomes for observations not predicted to be structural zeros.²⁷ Table 3 shows its outcome.

	MODEL WITHOUT $Deter_a$	MODEL WITH $Deter_a$
(INTERCEPT)	-1.036583 (0.13445)	-10.619204*** (8.46e-05)
CRISIS: BANKING	0.038923 (0.78041)	0.380994* (0.010475)
CRISIS: CURRENCY	-0.056606 (0.82031)	0.651357** (0.002915)
CRISIS: DEBT	-0.162466 (0.28447)	0.806135*** (4.93e-05)
PARTY: DEVELOP	-0.429502** (0.00428)	-0.645591*** (0.000975)
PARTY: PRIOR CLAIMS	-0.001853 (0.34385)	-0.007406*** (0.000480)
TERMINATE	-0.305293 (0.16325)	-0.059928 (0.705238)
USBITSIM: TREATMENT	-2.980004* (0.02992)	7.702902** (0.002470)
USBITSIM: EXPROPRIATION	-7.840975*** (8.35e-09)	0.799259 (0.638882)
USBITSIM: PROCEDURE	13.294663*** (< 2e-16)	3.246007. (0.063159)
IIA: TAXATION		-2.106189*** (1.80e-14)
IIA: SUBSIDIES		1.849253*** (9.61e-06)
IIA: PROCUREMENT		-3.281795*** (9.64e-12)
IIA: OTHERS		1.119760*** (0.000100)
IIA: POST-BIT		0.194069 (0.508992)
MFN: ECONOMIC		-0.947298*** (0.000543)
INTEGRATION		0.162344 (0.491965)
MFN: TAXATION		2.286047*** (3.59e-08)
MFN: PROCEDURAL ISSUES		-0.952592* (0.047612)
EXPROPRIATION:		2.727114*** (3.77e-09)
REGULATORY		
EXPROPRIATION:		
COMPULSORY LICENSES		
SOCIAL: HEALTH AND		-1.252208*** (0.000101)
ENVIRONMENT		
SOCIAL: LABOR STANDARDS		0.143434 (0.768106)
SOCIAL: RIGHT TO REGULATE		1.347133*** (1.67e-08)
SOCIAL: CORPORATE		-0.244752 (0.771486)
RESPONSIBILITY		
SOCIAL: NOT LOWERING		-0.190675 (0.657703)
STANDARDS		
SECURITY: EXCEPTION SELF-		0.417826 (0.381988)
JUDGING		
ISDS SCOPE: TREATY		0.240664 (0.145280)
CLAIMS		
ISDS SCOPE: LIMITATION		0.872235*** (7.94e-05)
ISDS SCOPE: POLICY		-0.214258 (0.399828)
ISDS SCOPE: TAXATION OR		0.172492 (0.618988)
PRUDENTIAL		

²⁷ Diane Lambert, *Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing*, 34 *TECHNOMETRICS* 1 *passim* (1992).

ISDS SCOPE: LIMITED REMEDIES	0.287729 (0.359550)
ISDS SCOPE: ALLOWING AMICUS CURIAE	0.665365 (0.440267)
ISDS FILING: CASE-BY-CASE CONSENT	-0.02385 (0.946945)
ISDS FILING: FORK IN THE ROAD	-2.256180*** (0.000322)
ISDS FILING: LIMITATION PERIOD	-1.772655*** (2.78e-06)
ISDS FILING: LOCAL REMEDIES FIRST	0.402943 (0.258133)
ISDS FILING: VOLUNTARY ADR	-0.035774 (0.848018)
<i>N</i>	39,524
	39,524

[TABLE 3] OUTCOMES OF ZERO-INFLATED POISSON REGRESSION
PREDICTING COUNTS OF ISDS FILINGS

The results of the model with $Deter_a$ are largely consistent with those of the Poisson regression, with the exception that the macroeconomic crisis variables and prior claims lose their explanatory power. In the model with $Deter_a$, though the statistical significance of similarity in procedural provisions is reduced, it still shows a positive correlation with higher ISDS filing counts.

C. Implications

This reinforces previous findings that ISDS filing occurrences are closely linked to banking crises, as evidenced by the exceptionally high number of filings faced by Argentina and Venezuela during the Chávez-Maduro era. Setting aside the political turmoil that accompanied these crises and contributed to increased filings, this trend may prompt states to reassess the impact of IIAs on their ability to navigate financial crises effectively.

While aligning substantial obligations with global standards—such as the 2012 U.S. Model BIT—appears to be either neutral or even ISDS-deterrent, potentially due to enhanced clarity and predictability, the alignment of procedural provisions is strongly and significantly correlated with ISDS occurrences. States interested in ISDS reform may find it more effective to focus on procedural provisions.

The impact of deterrent provisions on reducing ISDS filings remains inconclusive. Given that ISDS reform discourses are still relatively new and these provisions lack a long historical track record, further research is needed before drawing definitive conclusions.

D. Limitations

This study has several limitations. First, the method used to extract treatment, expropriation, and procedural provisions from IIAs may influence the calculated similarity scores and, consequently, the outcomes of the regression models. Second, the *Deter_a* variables rely heavily on human-coded data from the UNCTAD IIA Mapping Project, meaning any inaccuracies in the coding could affect the reliability of the regression results. Lastly, the study does not account for potential reverse causality: countries facing higher ISDS risk might be more likely to adopt U.S.-style treaties or incorporate stronger deterrent provisions.

IV. CONCLUSION

Rather than simply embracing the hype surrounding AI, this Article critically examines its practical role in augmenting traditional empirical research, demonstrating how machine learning and text embeddings can enhance quantitative legal analysis. By leveraging AI-driven methodologies, this approach enables more precise data clustering, offering deeper insights into patterns of convergence and divergence across jurisdictions. Additionally, AI-assisted text analysis facilitates the creation of novel independent variables, allowing for more nuanced and comprehensive empirical assessments of ISDS dynamics.

The findings reinforce the importance of procedural treaty design in driving ISDS filings while challenging conventional assumptions about the role of substantive treatment provisions and developed-developing country dynamics. Furthermore, AI-driven legal analytics can empower policymakers by providing empirical clarity on the impact of deterrent provisions, helping refine IIA harmonization efforts at the global level and guiding more informed national strategies for negotiating IIAs and managing ISDS risks. This Article underscores AI's potential not as a replacement for legal expertise, but as a powerful tool to complement traditional methods, offering policymakers a data-driven foundation for evidence-based treaty design and dispute resolution strategies.

CODEBOOK

VARIABLES	TYPE	CODING CRITERIA ²⁸
CRISIS: BANKING	Binary	• 1 if at least a party to the IIA experienced each type of macroeconomic crisis during a given year. 0 otherwise.
CRISIS: CURRENCY	Binary	
CRISIS: DEBT	Binary	
PARTY: DEVELOP	Binary	• 1 if at least a party to the IIA was a developed country and at least a party to the IIA was a developing country in a given year. 0 otherwise. • The total number of previous ISDS claims filed by investors within the host country prior to the given year
PARTY: PRIOR CLAIMS	Continuous	
TERMINATE	Binary	• 1 if the IIA is already terminated as of the given year. 0 otherwise.
USBITSIM: TREATMENT	Continuous	• Cosine similarity of treatment, expropriation, procedural provisions between the IIA and the 2012 U.S. Model BIT (between 0 and 1).
USBITSIM: EXPROPRIATION	Continuous	
USBITSIM: PROCEDURE	Continuous	
IIA: TAXATION	Binary	• 1 if the IIA excludes from its scope (i) taxation matters, (ii) subsidies and grants, (iii) government procurement, and (iv) any other subject matters, respectively. 0 otherwise.
IIA: SUBSIDIES	Binary	
IIA: PROCUREMENT	Binary	
IIA: OTHERS	Binary	
IIA: POST-BIT	Binary	• 1 if the IIA applies only to investments having been made after its entry into force. 0 otherwise.
MFN: ECONOMIC INTEGRATION	Binary	• 1 if the MFN does not cover advantages by virtue of (i) economic integration agreement, and (ii) double taxation treaties, respectively. 0 otherwise.
MFN: TAXATION	Binary	
MFN: PROCEDURAL ISSUES	Binary	• 1 if the IIA expressly states that the MFN provision does not apply to ISDS provisions found in other treaties. 0 otherwise.
EXPROPRIATION: REGULATORY	Binary	• 1 if the IIA carves out from expropriation (i) regulatory measures of general application and (ii) compulsory licenses, respectively. 0 otherwise.
EXPROPRIATION: COMPULSORY LICENSES	Binary	
SOCIAL: HEALTH AND ENVIRONMENT	Binary	• 1 if the IIA uses the terms related to each subject matter. 0 otherwise.

²⁸ For variables other than CRISIS, PARTY, TERMINATE, and USBITSIM, *see* UNCTAD, *supra* note 25, at 7–19.

SOCIAL: LABOR STANDARDS	Binary	
SOCIAL: RIGHT TO REGULATE	Binary	
SOCIAL: CORPORATE RESPONSIBILITY	Binary	
SOCIAL: NOT LOWERING STANDARDS	Binary	
SECURITY: EXCEPTION SELF-JUDGING	Binary	• 1 if the essential security exception is formulated as self-judging. 0 otherwise.
ISDS SCOPE: TREATY CLAIMS	Binary	• 1 if only claims alleging a breach of the IIA can be submitted to ISDS. 0 otherwise.
ISDS SCOPE: LIMITATION	Binary	• 1 if not all the IIA provisions are subject to ISDS. 0 otherwise.
ISDS SCOPE: POLICY	Binary	• 1 if the IIA excludes a particular policy area from its scope. 0 otherwise.
ISDS SCOPE: TAXATION OR PRUDENTIAL	Binary	• 1 if the disputing parties or the tribunal must refer certain matters for joint determination. 0 otherwise.
ISDS SCOPE: LIMITED REMEDIES	Binary	• 1 if the IIA specifies the types of remedy that a tribunal can award. 0 otherwise.
ISDS SCOPE: ALLOWING AMICUS CURIAE	Binary	• 1 if the IIA does not regulate submissions from <i>amicus curiae</i> . 0 otherwise.
ISDS FILING: CASE-BY-CASE CONSENT	Binary	• 1 if parties do not provide prior consent to ISDS. 0 otherwise.
ISDS FILING: FORK IN THE ROAD	Binary	• 1 if the IIA requires the investor to choose between domestic courts and ISDS. 0 otherwise.
ISDS FILING: LIMITATION PERIOD	Binary	• 1 if the IIA bans ISDS filings after a certain period. 0 otherwise.
ISDS FILING: LOCAL REMEDIES FIRST	Binary	• 1 if the IIA obliges an investor to undergo local remedy before filing ISDS. 0 otherwise.
ISDS FILING: VOLUNTARY ADR	Binary	• 1 if the IIA mentions the possibility of conciliation or mediation. 0 otherwise.