

TECHNICAL NOTE N° IDB-TN-03108

Tracking the Reform Process from Newspaper Data

Rafaela Bastidas Ripalda
Stephen Hansen
John León-Díaz
Yabra Muvdi

Inter-American Development Bank
Office of Strategic Planning and Development
Effectiveness

March 2025



Tracking the Reform Process from Newspaper Data

Rafaela Bastidas Ripalda
Stephen Hansen
John León-Díaz
Yabra Muvdi

Inter-American Development Bank
Office of Strategic Planning and Development Effectiveness

March 2025



**Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library**

Tracking the reform process from newspaper data / Rafaela Bastidas-Ripalda, Stephen Hansen, John León-Díaz, Yabra Muvdi.

p. cm. — (IDB Technical Note ; 3108)

Includes bibliographical references.

1. Text data minig-Colombia. 2. Information measurement-Colombia. 3. Colombia-Politics and government-Press coverage. 4. Public administration-Press coverage-Colombia. I. Bastidas-Ripalda, Rafaela. II. Hansen, Stephen. III. León-Díaz, John. IV. Yabra, Muvdi. V. Inter-American Development Bank. Office of Strategic Planning and Development Effectiveness. VI. Series.

IDB-TN-3108

<http://www.iadb.org>

Copyright © 2025 Inter-American Development Bank ("IDB"). This work is subject to a Creative Commons license CC BY 3.0 IGO (<https://creativecommons.org/licenses/by/3.0/igo/legalcode>). The terms and conditions indicated in the URL link must be met and the respective recognition must be granted to the IDB.

Further to section 8 of the above license, any mediation relating to disputes arising under such license shall be conducted in accordance with the WIPO Mediation Rules. Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the United Nations Commission on International Trade Law (UNCITRAL) rules. The use of the IDB's name for any purpose other than for attribution, and the use of IDB's logo shall be subject to a separate written license agreement between the IDB and the user and is not authorized as part of this license.

Note that the URL link includes terms and conditions that are an integral part of this license.

The opinions expressed in this work are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.



Tracking the Reform Process from Newspaper Data*

Rafaela Bastidas Ripalda

Stephen Hansen

John León-Díaz

Yabra Muvdi

March 2025

Abstract

Countries often undertake structural reforms to boost productivity, reduce inequality, and improve public sector efficiency, yet there is limited data on the reform process to evaluate these efforts. We examine the extent to which large language models (LLMs) applied to media articles can fill this evidence gap, using a case study from Colombia. We find that suitably prompted LLMs produce output that aligns with human reading. The output reveals extensive coverage of the debate and discussion that precedes reform adoption or rejection, and can be used to track the evolution of specific reforms. We conclude that our methodology, once expanded, could produce a uniquely valuable dataset.

JEL Codes: C55, O29, P11, P41

Keywords: Structural Reform, Political Processes, Latin America, Large Language Models

*Corresponding author: johnjairol@iadb.org. Rafaela Bastidas Ripalda (rafaelab@iadb.org) is a consultant at the Office of Strategic Planning and Development Effectiveness (SPD) of the Inter-American Development Bank. Stephen Hansen (stephen.hansen@ucl.ac.uk) is a Professor of Economics at University College London, Research Fellow for the Institute for Fiscal Studies, and Fellow for Center for Economic Policy Research. John León-Díaz (johnjairol@iadb.org) is an economist at the Office of Strategic Planning and Development Effectiveness (SPD) of the Inter-American Development Bank. Yabra Muvdi (yabran.muvdi@gmail.com) is an independent researcher. The authors thank Francesca Castellani for her support in developing this project. During the execution of the project, Hansen and Muvdi worked under advisory contracts to the IADB.

1 Introduction

Structural reforms are regulatory and institutional changes aimed at boosting productivity, reducing inequality, and improving public sector efficiency. In recent years, these reforms have shifted towards more granular interventions focused on specific sectors, which contrasts with the broader economic paradigm shifts that characterize them in the 1990s ([Céspedes et al. 2023](#)).

This increased granularity has made it challenging to fully characterize the reform process, largely due to limitations in available data. Although several efforts have been made to capture these processes, they often rely on policy-based proxies to gauge reform efforts, which has limited the analysis to a narrower set of areas (e.g., [Lora \(2012\)](#), [Alesina et al. \(2024\)](#)).

A policy-based proxy is a measurable policy action or regulation that serves as an indirect indicator of a broader structural reform, such as using changes in average tariff rates to represent trade liberalization or amendments to labor protection laws for labor market reforms. While these proxies might be useful when direct measures are unavailable, they oversimplify the complexities behind the reform process or miss the different stages the reform goes through. For example, reforms are often hindered by technical and political challenges in their design, approval, and execution that these measures do not capture. According to the [OECD \(2010\)](#), the most successful reforms typically took several years to prepare and adopt, with implementation timelines often extending even further, and many followed earlier setbacks that were crucial in paving the way for more comprehensive initiatives in the future.¹

In this note, we present a novel methodological approach for systematically characterizing reform efforts across multiple areas and stages, overcoming existing limitations in the literature. Our approach uses newspaper data and text analysis tools to identify and categorize relevant reform information. A key advantage of using newspapers, compared to official gazettes, is their ability to capture not only successful reforms but also those that were dismissed or rejected. Unlike gazettes, which focus on final legal texts, newspapers provide insights into political debates, public reactions, and the reasons behind reforms being delayed or blocked, offering a more complete picture of the reform process. Although news structures may carry biases or value judgments, our methodology focuses on extracting objective information about reforms, such as their justification, timing, and topics, rather than their perceived relevance or popularity. The purpose of this note is to demonstrate the feasibility of our approach as a proof of concept,

¹An additional limitation of policy-based proxies is that, without understanding the context of a reform, gradual implementations can lead to multiple changes in the indicators, which may be mistakenly interpreted as separate reforms.

providing a foundation for future development and larger-scale implementation.

Our construction begins with a set of media outlets in Colombia encompassing several newspapers and magazines over the period from 1991 through 2011. As is standard in text analysis (e.g. [Baker et al. 2016](#)), we use a set of keywords to filter out articles unlikely to be related to reform. Next, we use a large language model to generate an initial classification of remaining articles as containing reform information and, if so, various dimensions of the reform such as the sector and stage of adoption. We then use human reading of the output to refine and extend the query. Overall, the accuracy of the model's output is quite high along all dimensions. The outcome of this process is the first key output: an article-level dataset that categorizes individual articles by type, stage, and other dimensions. From this dataset, we derive additional insights, such as the intensity and types of reform activity over time in Colombia, providing a comprehensive view of the reform landscape.

We then algorithmically cluster articles into groups that represent specific reforms, enabling us to track the full trajectory of individual reforms from early public discussion through to formal political debate. This process produces the second key output of the methodology: a reform-level dataset that maps individual articles to reforms. Additionally, we present case studies illustrating how our procedure captures two important episodes in Colombia: the 1992 fiscal reform and the 2011 higher education reform.

This methodological proposal is important for several reasons. From an academic perspective, the creation of a database that tracks reforms at various stages of their cycle serves as a public good, enabling researchers to address a wide range of questions. It offers a holistic view of the entire reform process, as it does not focus solely on approved reforms. Finally, it also helps overcome the limitations of using policy variables as proxies for the reform process and broadens the scope of analysis beyond the traditional focus of the literature on trade, the external sector, product markets, and labor; this extends to areas of reform in health, education, pensions, and environmental, among others.

For policymakers, it helps to gauge both the appetite around specific types of reforms, while also improving planning by offering estimates of the expected time for reform approval. Furthermore, it allows them to reference past cases and analyze the reasons why some reforms, even if approved, were later reversed. For international financial organizations, such as the Inter-American Development Bank (IADB), it offers a valuable tool to monitor their efforts in supporting reforms.

This note is structured as follows: section 2 characterizes the data, section 3 presents the general aspects of the methodology, section 4 presents the results applied to the case study of Colombia, section 5 presents final discussion and conclusions.

2 Data

The setting for our case study is Colombia. We obtain a sample of news articles from publicly available webpages from three major Colombian news outlets: El Tiempo (89%), Semana (7%), and Dinero (4%). This imbalance is due to the different nature of each one of the sources. El Tiempo is a daily newspaper covering a wide range of topics, while Semana and Dinero are weekly magazines focused primarily on political opinions and financial information, respectively.² In total, we have 91,749 articles that span the time period 1980-2011. Prior to 1991, though, we have relatively few articles. Starting in 1991, the sample contains at least 2,000 articles a year and in some years over 5,000 articles. For this reason, we limit attention in the results section below to the sample after 1991. The texts are in Spanish. In total there are 38,740,935 words. The mean article length is 422 words with a standard deviation of 447. The minimum article length is 18 and the maximum is 13,272.

Because the data contains articles about a variety of topics, the bulk of the content will be irrelevant for measuring the reform process in Colombia. We therefore create an initial filter that reduces the articles to a set that is plausibly relevant. To do so, we focus on articles that have at least one sentence that contains (i) one (or more) reform-related noun and (ii) one (or more) reform-related verb. Table 1 contains these nouns and verbs. We consider all conjugations of the verbs when applying the filter. The terms were chosen by us using our domain expertise. Of the 1,500,718 sentences in the data, 4,857 contain both a reform-related noun and verb. Examples of such sentences include the following, where keywords are in bold:

***Propongo** fomentar un debate sobre las razones y la conveniencia de la **propuesta** que el Concejo tiene en sus manos.*

*Diplomáticos dijeron que el jefe del Tesoro **proponía** limitar los superávits y déficits de cuenta corriente - la **medida** más amplia de comercio de bienes y servicios - a un 4 por ciento del Producto Interno Bruto.*

²The data was collected as part of a research project at Universidad de los Andes.

Nouns	Verbs
reforma	aprobar
ley	rechazar
regulación	modificar
decreto	debatir
resolución	discutir
medida	cambiar
proyecto	proponer
propuesta	formular
normativa	hundir
reglamentación	enmendar
ponencia	estudiar

Table 1: Keywords for Filtering Reform-Relevant Articles

*Los ponentes de la **reforma** tributaria y el Gobierno llegaron ayer a nuevos acuerdos para agilizar el trámite del proyecto, que hoy será **debatido** en las comisiones económicas del Congreso.*

In total, 3,721 articles contain such a sentence and this forms our set of articles for further analysis in the rest of the paper.

We have not benchmarked the performance of the keyword filter, and further adjustments could potentially reduce measurement error. However, the filter may be less restrictive than it seems, as a single reform is often reported by multiple news outlets, increasing the likelihood of inclusion. Additionally, this initial filter will be valuable for future algorithm optimization, especially given the download limitations of platforms compiling news. An alternative approach would be to input all articles into the large-language model, but this would likely require significantly more computational resources.

3 Measuring the Reform Process

3.1 Reform Process in Colombia

Before detailing the methodological approach to identifying reforms, it is important to provide a brief overview of the reform process in Colombia. This will help clarify the stages that the proposed methodology aims to capture. We consider three general types of reforms: those approved through the legislative process in Congress, those approved through presidential decrees, and, finally, regulatory changes approved by government agencies.

Colombia has a bicameral Congress, which consists of a House of Representatives and a Senate. Both chambers must approve a bill for it to become law. Major reforms that alter legal frameworks—such as labor, education, health, or tax reforms—are typically passed through Congress and formalized as laws. The legislative process is structured and thorough, involving several steps across both chambers. Laws can be proposed by government officials, members of Congress, or public initiatives, and they must undergo a series of debates, amendments, and votes in each chamber. After receiving approval from both chambers, the bill is sent to the President, who can either sign it into law or raise objections. Additionally, certain laws, especially those related to fundamental rights, require a more stringent process and a review by the Constitutional Court to ensure they comply with the Colombian Constitution.

Not all bills presented become laws, and original proposals often undergo modifications before approval. As a result, the approval of reforms can be a lengthy process, with a bill potentially progressing through various stages before completion. These stages include being drafted, debated in the chambers of Congress, awaiting approval by the chambers and the president, or undergoing a constitutional review by the courts, which can occasionally reverse the reform. Tracking these stages in our methodological approach will be crucial to accurately analyze and capture the progress and outcomes of legislative reforms.

Alternatively, reforms can be implemented through various other mechanisms. First, reforms can be approved through presidential decrees without requiring legislative approval during a state of emergency. The president also has the power to issue regulatory decrees, which are necessary to implement and enforce the provisions contained in the laws and to establish the procedures for their compliance. For instance, in Colombia, free basic education was established through Decree 4807 of 2011. This decree aimed to regulate and develop existing laws, specifically the 1991 Constitution and the General Education Law (Law 115 of 1994), which recognize education as a fundamental right.

And second, reforms that involve technical or procedural changes, such as adjustments to regulatory frameworks, may be executed by relevant government agencies through administrative resolutions or regulatory adjustments, rather than through formal laws or decrees. For example, the Ministry of Health can issue administrative resolutions to address public health challenges, such as modifying vaccination protocols or establishing standards for hospital operations.

3.2 Methodology

Our main measurement challenge is mapping each article selected by our keyword filter to the various types and dimensions of the reform process described above. Although we anticipate fewer articles discussing reforms made by decrees or procedural changes by government institutions over a short period, we expect more extensive coverage of reforms progressing through Congress and the House of Representatives.

Our broad strategy is to prompt a pretrained large language model (LLM) to classify an article input along these dimensions. This strategy is known as *zero-shot learning* since it relies on the LLM's pre-existing knowledge to generate output.³ Even so, implementing the strategy still requires design choices. One must decide which LLM to adopt. Also, the output of LLMs can be sensitive to the formulation of the prompt, and developing prompts is typically a trial-and-error process requiring human oversight (a process known as *prompt engineering*). In this section, we describe how we go about selecting our final model.

There is an increasingly large number of LLMs, but as a starting point we limit attention to two: GPT-4 provided by OpenAI and Gemini provided by Google.⁴ We design initial prompts for each one to benchmark their performance and examine whether their outputs align with our measurement needs. The initial prompt for GPT-4 is in Appendix A.1 and for Gemini is in Appendix A.2. We refer the interested reader there for full details,⁵ but the main classification dimensions in the prompt are:

1. Whether the article mentions a reform
2. Whether the article refers to events or situations in Colombia
3. Whether the article refers to events or situations that are presently occurring (as opposed to in the past)
4. What kind of reform is mentioned. The available categories are 'fiscal', 'education', 'health', 'infrastructure', 'other', or 'not applicable'

³One alternative strategy—known as *few-shot learning*—would be to provide the model several examples of correctly classified articles before it generated output. Another would be to curate a more extensive labeled dataset and fine-tune the model to maximize predictive accuracy. We choose zero-shot learning since it is relatively easy to prototype, although in general one would expect gains from further model training.

⁴More precisely, we use GPT version gpt-4o-2024-08-06 and Gemini version gemini-1.5-flash.

⁵Although they do not form a part of our analysis below, we also prompt the model to give a specific name to a mentioned reform and to name individuals involved in the process.

5. The stage of the reform in the overall process. The available categories are ‘discussion’, ‘approval’, ‘reversed’, ‘other’, or ‘not applicable’

To assess the quality of each model and the initial prompt, we randomly sample 500 articles from the keyword-filtered set and pass each sampled article to both models. In 181 cases, both models agree that a reform is mentioned in the article.⁶ We choose this sample for human reading, and compare model output against our own labels for the above categories.

As a starting point, we find that 76% of the 181 articles mention a reform we consider relevant, which suggests that the joint application of the keyword filter and LLM prompts is still not sufficient to identify reform articles with a particularly high level of accuracy. We find that the measurement error arises from two main systematic sources. First, some articles classified as mentioning reform are opinion pieces in which the author advocates that a certain reform should be made. We do not consider these relevant since such articles discuss hypothetical rather than actual reform processes. Second, some reform mentions arise from local-level rather than national-level processes. As such, we modify our initial prompt accordingly.

Table 2 displays the results of the rest of our audit of the 181 selected articles. Table 2a presents both models’ accuracy at classifying whether an article discusses events in Colombia or elsewhere, and whether an article discusses present events. Both models are highly accurate at classifying news in Colombia, with lower accuracy scores for present-day events. Table 2b assesses the accuracy of the classification of the type of reform. We assess the accuracy of each category as a binary problem, i.e., we judge whether a type of reform is or is not mentioned against whether a model predicts a type of reform is or is not mentioned.⁷ Accuracy scores are generally high except for the ‘Other’ category. Our reading also revealed a relatively low incidence of the ‘Infrastructure’ category. As such, we modify the original prompt to remove the ‘Infrastructure’ category from reform type and instead introduce political and judicial reforms, respectively.⁸ Finally, Table 2c shows results for the stage of the reform process. Again, we find generally high accuracy scores. Our reading suggested, however, that ‘Discussion’ is too

⁶The reduction of the articles from 500 to 181 is largely because the keyword filtering is noisy, not because GPT-4 and Gemini have high disagreement rates. Of the 500 classified articles, the two models agree in 432 cases on whether a reform is mentioned or not.

⁷The prompt is designed to allow for each article to have multiple types of reform mentioned, so in principle this is not a multi-way classification problem in which only one category is correct.

⁸This does not imply the absence of infrastructure reforms but suggests that they may be discussed in a context different from what was initially outlined in our initial filters. Alternatively, it is possible that such news does not align with the interests of the newspapers included in the sample or that the level of discussion on this topic was relatively low during the study period.

Table 2: Accuracy comparisons for GPT-4 vs Gemini

Task	GPT Accuracy	Gemini Accuracy
Colombia	0.979	0.986
Present	0.822	0.705

(a) Accuracy comparison for domestic and contemporary reform mentions

Type	GPT Accuracy	Gemini Accuracy
Education	0.97	0.99
Fiscal	0.91	0.93
Health	0.95	0.99
Infrastructure	0.88	0.99
Other	0.63	0.65
N/A	0.84	0.84

(b) Accuracy comparison for reform type

Stage	GPT Accuracy	Gemini Accuracy
Discussion	0.71	0.62
Approval	0.82	0.84
Reversed	0.99	0.98
Other	0.92	0.92
N/A	1.00	1.00

(c) Accuracy comparison for adoption stage

expansive a category. Our original aim was to capture formal political debate inside official government institutions, but many ‘Discussion’-classified articles mention an earlier stage in which there is public discussion of a proposed reform but not yet formal political activity. As such, we modify the prompt to include additional stages in the reform process: ‘Discussion’ (drafts prior to legislative debate); ‘Debate’ (the reform has entered the legislative process); ‘Approval’ (the reform is formally adopted); ‘Rejected’ (the reform is explicitly dismissed and not approved); ‘Reversed’ (after adoption, another government entity reverses the reform).

In short, the human audit establishes that LLMs produce output regarding the reform process that generally aligns with human reading in spite of having no prior training data. Furthermore, the audit allows us to refine the original prompt to both reduce sources of measurement error and add additional detail. In terms of which model to select, both are comparable in terms of accuracy. We opt for GPT-4 primarily for practical reasons: it allows for structured output in which the response to a prompt directly populates a tabular dataset which we can immediately use for analysis.

The final prompt to GPT-4 is in Appendix [A.3](#). We pass each of the 3,721 keyword-filtered articles

through the model. We limit attention to articles which are classified as:

1. Mentioning a reform
2. Discussing events in Colombia
3. Discussing present events
4. Not an opinion article or interviews
5. Pertaining to national rather than local issues

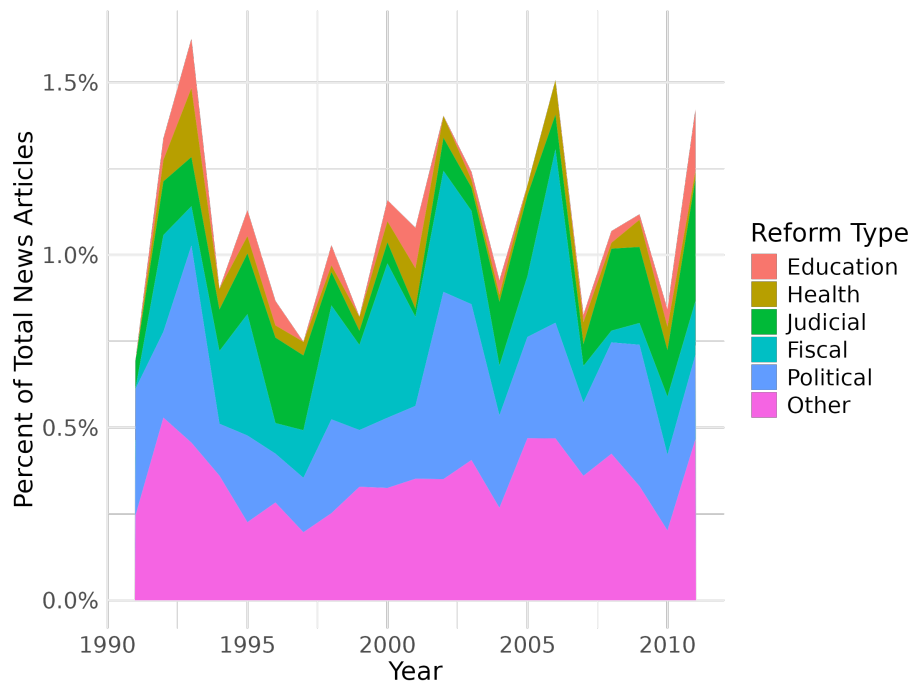
In total, 985 articles satisfy these criteria, or slightly more than 1% of the raw number of original articles. These articles, along with their classifications, are available at https://drive.google.com/file/d/1Sw1KOYCRnsJa8ax0luDlzD_1MVDKpYt2/view?usp=drive_link.

4 Descriptive Results

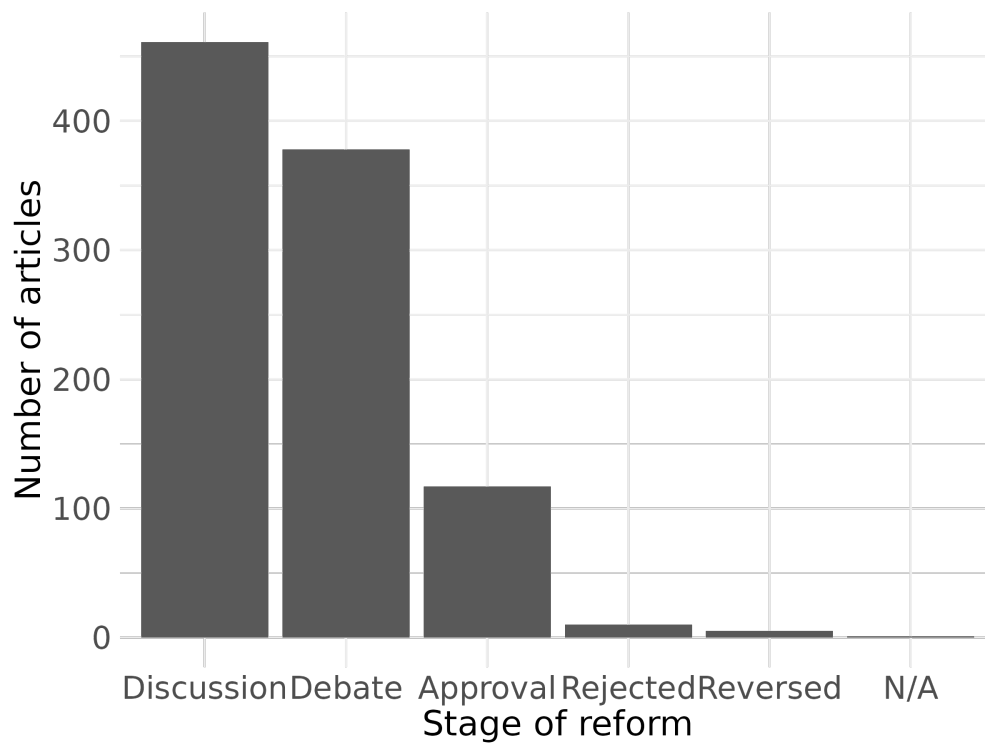
We now turn to describing basic properties of the reform process in Colombia. Recall from Section 2 that we observe relatively few news articles prior to 1991. In this section, we therefore focus on the reform articles observed from this year onwards—972 in total.

To begin, Figure 1a plots the total share of reform articles from 1991 through 2011. There is considerable time series variation in the share. The year with peak share is 1993, after which the share of reforms drops before rising again in the mid-2000s and towards the end of the sample. The Figure also breaks down the total share into the individual types of reforms that the model classifies. Reform types are ordered by their frequency in the overall sample. Thus, the ‘Other’ category accounts for the largest total fraction (0.31). Still, the remaining five categories make up the bulk of reform articles. Political and fiscal reforms accounts for most of these, but their overall importance declines over time. Education and health reforms are rarer, but in some years these too generate media coverage.

Figure 1b breaks down reform articles into the various stages of the process. We have ordered the stages left-to-right according to the timeline we expect individual legislative reforms to follow in their adoption process. Interestingly, we observe that the number of articles is declining along this sequence of steps. One interpretation is that individual reforms are ‘born’ in the Discussion stage and either ‘die’ at that stage or pass on to the next Debate stage where they again either die or pass to the Approval stage. These dynamics would naturally explain decreasing discussion



(a) Share of Reform Articles by Reform Type: 1991-2011



(b) Frequency of Articles by Process Stage

Figure 1: Summary of Reform Articles by Type and Stage

in each stage along the adoption path. Interestingly, the formal legislative process of rejecting reforms (Rejected) or other institutions' invalidating them (Reversed) is rarely mentioned in the news media. This suggests that most reforms are abandoned without formal government action. These findings highlight a limitation of databases like official gazettes, which record only enacted reforms and may overlook much of the reform activity. At least according to media attention, most political activity surrounding reform is spent in stages that precede formal political adoption, an area that warrants further research.

4.1 Reform-level analysis

The aggregate statistics discussed so far are the byproduct of individual reforms' separate processes. Although our final GPT-4 prompt (Appendix A.3) includes a prompt for the name of a specific reform, this does not produce a unique reform ID that can be immediately used to follow individual reforms. Instead, we develop a clustering algorithm whose goal is to group articles that discuss the same reform. The procedure is as follows:

1. Order all reform articles by date
2. Treat the earliest article as an initial 'base' article and call its month $t = 1$.
3. Proceed in sequence through articles in the following six months (i.e. months $t = 1, \dots, 6$) of the same reform type as the base article. For each article, prompt GPT-4 to identify whether the new article discusses the same reform as the base article.
 - (a) If yes, consolidate the article with the base article, and generate a new base article by asking GPT-4 to summarize their joint content.
 - (b) If no, proceed to the next article.
4. If no article is found that relates to the initial base article, assign the initial base article a unique reform ID. Otherwise, continue searching for related articles of the same reform type in the next six months (i.e. months $t = 7, \dots, 12$). If none are found, assign all related article in months $t = 1, \dots, 6$ a unique reform ID. Otherwise, continue searching for related articles in months $t = 13, \dots, 18$. Proceed similarly until no related article is found in a given six-month window.

Finally, we remove all articles with a unique reform ID and repeat the above procedure. We continue until no articles remain to be clustered. The resulting grouping of articles into reforms is

available at https://drive.google.com/file/d/1_rouc8XL03PL8C1zxZryfbZvJVumt2QC/view?usp=drive_link.

There are 650 unique articles clustered because we remove articles in the ‘Other’ reform type category.

Articles per Reform ID	1	2	3	4	5	6	7	8	9	10+
Frequency	319	43	16	7	5	4	2	2	1	7

Table 3: Tabulation of Number of Articles per Reform ID

Table 3 tabulates how many articles are present in each unique reform ID. Most reform IDs are discussed in only one article. This may reflect that, as discussed above, certain reforms in Colombia can be adopted by decree without a meaningful process. Also, our sample of news articles may not be detailed enough to capture the full scope of media attention to individual reforms. At the same time, there are many instances in which the same reform is discussed across multiple articles and just over half of articles are part of a multi-article reform ID. To obtain a sense of the chains of articles our procedure identifies, we present two examples. The first, presented in Table 4, tracks the journey of the government’s proposal to reform higher education in 2011 that, among other changes, would allow for a greater role for profit-making educational institutions.⁹ This is a notable episode in Colombian politics¹⁰. In the immediate aftermath of initial proposal made to university rectors by the government, students and teachers organized a series of nationwide protests in April. These were largely peaceful and attended by thousands of participants. In spite of the opposition of the university sector, the government continued through the Summer to insist that the reform was necessary to maintain the financial health of tertiary education. The situation continued to escalate, and students organized a nationwide strike at the beginning of the academic year in an effort to derail the proposal. Finally, in the Fall of 2011, the government abandoned its reforms in the face of the disruption the student strikes were causing.

The second example, presented in Table 5, tracks the process of the government’s proposed fiscal reform in 1992.¹¹ Among other measures, the government of César Gaviria sought to increase value added tax by three percentage points, from 12% to 15%, and to increase legal

⁹In total there are six articles in this reform ID, all of which we display here.

¹⁰See for example <https://web.archive.org/web/20110411024759/http://elespectador.com/noticias/nacional/articulo-261478-santos-afrenta-primera-protesta-de-sindicatos-maestros-y-estudia>

¹¹In total there are seven articles in this reform ID, but we here present five for space considerations.

Date	Stage	Summary	Text Extracts	Article ID
2011-04-07	Discussion	Más de cinco mil personas en Cartagena participaron en una marcha pacífica en contra de la reforma a la educación superior, organizada por trabajadores, estudiantes y profesores de la Universidad de Cartagena.	Eliécer González, presidente del sindicato de trabajadores de la Universidad de Cartagena, dijo que el objetivo de la marcha se cumplió en un ciento por ciento. “Desde acá también rechazamos esa desafortunada reforma a la educación superior y todo el estamento universitario estuvo de acuerdo en la protesta pacífica de hoy”, señaló.	2618
2011-04-07	Discussion	Miles de estudiantes y sindicatos en Colombia protestan contra la reforma a la Ley 30 de educación superior y otras demandas laborales.	A esta hora, la manifestación se concentra en la Plaza de Bolívar, en donde los estudiantes rechazan la propuesta de reforma a la Ley 30 de educación superior.	85878
2011-04-08	Discussion	Miles de personas protestaron pacíficamente en Colombia contra la reforma de la ley de educación superior, con algunos incidentes aislados de violencia.	De forma pacífica, y solo con hechos aislados de violencia en Pereira, Bogotá y Bucaramanga, miles de personas salieron a las calles de las principales ciudades del país para oponerse al proyecto del Gobierno para reformar la ley de educación superior.	1934
2011-07-06	Discussion	El Gobierno colombiano está preparando un proyecto de reforma al sistema de educación superior que incluye la creación de universidades con ánimo de lucro, enfrentando oposición de rectores.	Sin embargo, el Gobierno se mantiene en su propuesta de crear instituciones con ánimo de lucro, como una forma de ampliar cobertura, pues “es necesario que ingresen más recursos a financiar la educación superior en el país; adoptaremos todas las medidas que impidan que se presenten aspectos no deseables producto de esta decisión”, dijo Campo.	287
2011-10-21	Debate	El artículo discute el paro estudiantil en Colombia en protesta contra la reforma de la educación superior, destacando el impacto económico y las demandas de los estudiantes para retirar el proyecto del Congreso.	En medio de un ‘pupitrero’ en contra de la reforma de la educación superior, que se extendió a varias universidades del país, la ministra de Educación, María Fernanda Campo, hizo un nuevo llamado a los líderes estudiantiles para que participen en la mesa de diálogo y terminar así con el paro indefinido que realizan desde la semana pasada.	646
2011-11-12	Rejected	Los estudiantes universitarios debaten si continuar el paro tras el retiro del Congreso de la reforma a la educación superior por parte del gobierno.	Líderes estudiantiles debaten el sábado acerca de si mantienen o no un paro nacional de universidades, luego que el gobierno oficializara el retiro del Congreso de un proyecto de ley de reforma a la educación superior, principal causa de la protesta.	86102

Table 4: Articles Discussing 2011 Education Reform

Date	Stage	Summary	Text Extracts	Article ID
1992-04-20	Debate	El artículo discute la propuesta de la Reforma Tributaria en Colombia que introduce penas de cárcel para evasores de impuestos, generando controversia sobre su aplicación y constitucionalidad.	En muchos países del mundo hablar cárcel por evasión de impuestos puede sonar normal. Pero en un país como Colombia, donde la evasión tributaria es el pan de cada día en cualquier operación comercial, el anuncio de que los evasores pueden convertirse en presos ha producido escozor. La polémica inicial sobre el nivel del IVA está siendo rápidamente desplazada por el debate sobre el carcelazo.	89168
1992-05-14	Debate	El Partido Conservador no ha llegado a un acuerdo sobre su posición frente a la reforma tributaria en debate en el Congreso, oponiéndose al aumento del IVA.	De todas formas, la posición del conservatismo es de abierta oposición al aumento del IVA del 12 al 15 por ciento como lo propusieron los ponentes liberales del proyecto.	8101
1992-05-22	Debate	El artículo discute la crisis política en Colombia debido a la oposición de varios bloques políticos a la reforma tributaria propuesta por el presidente César Gaviria, especialmente el aumento del IVA.	En un corrillo, senadores liberales hablaban ayer en la mañana de un Presidente sitiado. Se referían, por supuesto, al presidente César Gaviria, a propósito del bloqueo planteado por su mismo partido y otras organizaciones políticas a la reforma tributaria. El problema, en síntesis, es que si hoy se votara la propuesta del Gobierno, probablemente sería derrotada.	17332
1992-06-08	Discussion	El Gobierno de Colombia está considerando incluir congresistas en el gabinete para impulsar la reforma tributaria, a pesar de restricciones constitucionales.	El Gobierno está pensando en llamar a algunos congresistas para formar parte del gabinete en la próxima crisis ministerial.	87484
1992-07-02	Approval	El artículo discute la participación de diferentes fuerzas políticas en el gabinete de César Gaviria tras la aprobación de la reforma tributaria en Colombia.	Tanto la AD M - 19 como los conservadores y el Movimiento de Salvación Nacional fueron severos críticos de la reforma tributaria y votaron contra el reajuste del IVA, que finalmente fue aprobado con votación principalmente de liberales y de la Nueva Fuerza Democrática.	82274

Table 5: Articles Discussing 1992 Fiscal Reform

penalties associated with tax evasion.¹² This generated opposition from a wide range of political parties, including Gaviria's own. However, Gaviria ultimately built a coalition to support the measure which was approved.

These examples illustrate the promise of the database to shed light on the granular details of the reform process, and to identify key bottlenecks for adoption.

5 Discussion and Conclusion

In this report, we have described an approach for extracting information from media articles about the reform process. We use human reading combined with a large language model to classify articles as pertaining to reform or not; the type of reform being undertaken; and the stage of the adoption process being discussed, among other dimensions. We illustrate our approach on a sample of Colombian news articles and show promising results. The output of the model is generally accurate when compared with human readings. The output also reveals that extensive public debate occurs prior to a reform's being adopted or rejected. Finally, we apply a clustering procedure that organizes articles at the level of individual reforms, which allows one to enter into the granular details of their specific trajectories. Overall, the results suggest the potential value of scaling the approach to build a more comprehensive database that covers a longer time period and more countries.

At the same time, it is important to acknowledge the limitations of the proof-of-concept and highlight areas where further investment is required for the methodology to achieve its full potential. Most basically, the media data we use is limited in scope. Our sample is dominated by a single newspaper, *El Tiempo*, which covers a range of topics unrelated to politics and business. The data collection process was somewhat *ad hoc* and relied on scraping tools. These factors necessarily imply we have a limited sample of articles from which to measure the reform process. Expanding the set of newspapers, focusing on more specialized content, and using APIs or other data curation services would all contribute to a broader information base. Moreover, smaller reforms that are not considered newsworthy may go unreported in newspapers. As a result, the methodology might primarily capture major reforms with significant impacts on broader segments of society; testing this hypothesis would require validation in a context with access to a wider range of news sources.

¹²This episode is set within the broader history of Colombian fiscal reform in [Sánchez and Espinosa \(2005\)](#).

The natural language processing tools we use also have scope for improvement. The initial keyword filters we use to screen out irrelevant content are a reasonable starting point, but have not been validated by human reading. To reduce the computational burden, we opted for strict search rules to limit the number of articles processed by LLMs. An alternative approach would be to relax the rules, for example to only require an article to contain an instance of a reform noun in Table 1, and let an LLM determine what content is relevant or not. This would be another way of maximizing the use of information available from raw media data.

Our use of an off-the-shelf LLM was also governed by some resource limitations. A larger corpus of media articles could fine-tune the model to better understand the particular structure of Spanish-language newspaper articles. In the same spirit, a larger effort to generate human ‘ground-truth’ labels for articles would provide training data we could use to improve our model’s classification accuracy. Finally, there is scientific and policy value in creating a replicable pipeline. While GPT-4 is clearly powerful and simple to use, it can only be accessed by OpenAI’s API and may deliver different outputs even when queried under identical conditions from the user’s perspective.¹³ It can also not be easily fine-tuned by the end user. Using open-source models comes with a higher start-up cost but increases transparency and adaptability.

While limited in scope, our case study demonstrates the potential of this methodology to reveal valuable insights into the reform process. Moving forward will require additional investment, but these efforts could significantly enhance our understanding of reform dynamics. A deeper knowledge base would allow researchers to explore new questions, help policymakers gauge interest in specific reforms within a country, and provide international financial organizations with key information on past reform efforts.

¹³This is not merely a result of the model sampling different words from a fixed distribution in response to the same query. Even with the so-called ‘temperature’ of the model set equal to zero, which in principle should always draw the modal outcome, one can observe different outcomes.

References

- Alesina, A., Furceri, D., Ostry, J. D., Papageorgiou, C., and Quinn, D. P. (2024). Structural Reforms and Elections: Evidence from a World-Wide New Dataset. *Journal of the European Economic Association*, 22(4):1936–1980.
- Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring Economic Policy Uncertainty*. *The Quarterly Journal of Economics*, 131(4):1593–1636.
- Céspedes, L. F., León-Díaz, J., Orduz, M., and Bustos, S. (2023). From Discussion to Action: Characterizing Areas of Reform in Latin America and the Caribbean. IDB Publications (Working Papers) 13118, Inter-American Development Bank.
- Lora, E. (2012). Structural reform in Latin America: What has been reformed and how it can be quantified (updated Version). Research Department Publications 4809, Inter-American Development Bank, Research Department.
- OECD (2010). *Making Reform Happen*.
- Sánchez, F. and Espinosa, S. (2005). Impuestos Y Reformas Tributarias En Colombia, 1980-2003. Documentos CEDE 2824, Universidad de los Andes, Facultad de Economía, CEDE.

Appendix

A Model Prompts

A.1 Initial GPT-4 prompt for human evaluation

Sistema:

Eres un asistente experto en análisis de artículos de prensa, especializado en identificar y extraer información sobre distintos tipos de reformas en Colombia en formato JSON. Tu trabajo requiere un alto nivel de precisión, ya que la exactitud en los detalles es crucial.

Asistente:

A continuación, se te proporcionará un artículo de prensa. Tu tarea es identificar y extraer información clave sobre reformas en Colombia. Proporciona la información solicitada en formato JSON, asegurándote de que cada clave corresponde a una característica especificada y su valor es tu respuesta.

1. **“es_reforma”**: Indica si el artículo menciona alguna reforma. Usa 1 para “Sí” y 0 para “No”.
2. **“es_colombia”**: Indica si el artículo se refiere a eventos o situaciones que ocurren en Colombia. Usa 1 para “Sí” y 0 para “No”.
3. **“es_presente”**: Indica si el artículo se refiere a eventos o situaciones que están ocurriendo actualmente. Usa 1 para “Sí” y 0 para “No”.
4. **“tipo_reforma”**: Clasifica la reforma mencionada en una de las siguientes categorías: “fiscal”, “educación”, “salud”, “infraestructura”. Responde “otra” si la reforma discutida es claramente distinta a las categorías propuestas. Responde “no aplica” si el artículo no menciona una reforma. Usa solo estas categorías. No es posible responder con categorías diferentes. Por último, si en el artículo se mencionan varias categorías de reformas, proporcióñalas en una lista separada por comas.
5. **“otra_nombre”**: Si la reforma fue categorizada como “otra”, da un nombre a la categoría de la reforma. En caso contrario responde “no aplica”.
6. **“etapa_reforma”**: Determina la etapa en que se encuentra la reforma, eligiendo entre las siguientes categorías: “discusión”, “aprobación”, “reversada”, “otra”, “no aplica”.

7. **“nombre_reforma”**: Si se menciona en el artículo, proporciona el nombre exacto de la reforma, tal como “reforma tributaria” o “reforma de la ley 100”.
8. **“actores_involucrados”**: Proporciona una lista de los actores relevantes mencionados en el artículo que están involucrados en la reforma, separándolos por comas.
9. **“resumen”**: Proporciona un resumen conciso de una oración sobre el contenido principal del artículo.

Ejemplos del output correcto:

```
{
  "es_reforma": 1,
  "es_colombia": 1,
  "es_presente": 1,
  "tipo_reforma": "salud",
  "otro_nombre": "no aplica",
  "etapa_reforma": "discusión",
  "nombre_reforma": "reforma del sistema de salud",
  "actores_involucrados": "Ministro de Salud, Congreso",
  "resumen": "El artículo discute la nueva reforma del sistema de salud que está
↪ en etapa de discusión en el Congreso."
}

{
  "es_reforma": 1,
  "es_colombia": 1,
  "es_presente": 1,
  "tipo_reforma": "fiscal, educación",
  "otro_nombre": "no aplica",
  "etapa_reforma": "aprobación",
  "nombre_reforma": "",
  "actores_involucrados": "Ministro de Hacienda, Ministro de Educación,
↪ Congreso",
}
```

```
"resumen": "El artículo analiza las reformas fiscal y educativa que acaban de  
↪ ser aprobadas en el Congreso."  
}
```

Artículo:

A.2 Initial Gemini prompt for human evaluation

Eres un asistente experto en análisis de artículos de prensa, especializado en identificar y extraer información sobre distintos tipos de reformas en Colombia en formato JSON. Tu trabajo requiere un alto nivel de precisión, ya que la exactitud en los detalles es crucial.

A continuación, se te proporcionará un artículo de prensa. Tu tarea es identificar y extraer información clave sobre reformas en Colombia. Proporciona la información solicitada en formato JSON, asegurándote de que cada clave corresponde a una característica especificada y su valor es tu respuesta.

1. **"es_reforma"**: Indica si el artículo menciona alguna reforma. Usa 1 para "Sí" y 0 para "No".
2. **"es_colombia"**: Indica si el artículo se refiere a eventos o situaciones que ocurren en Colombia. Usa 1 para "Sí" y 0 para "No".
3. **"es_presente"**: Indica si el artículo se refiere a eventos o situaciones que están ocurriendo actualmente. Usa 1 para "Sí" y 0 para "No".
4. **"tipo_reforma"**: Clasifica la reforma mencionada en una de las siguientes categorías: "fiscal", "educación", "salud", "infraestructura". Responde "otra" si la reforma discutida es claramente distinta a las categorías propuestas. Responde "no aplica" si el artículo no menciona una reforma. Usa solo estas categorías. No es posible responder con categorías diferentes. Por último, si en el artículo se mencionan varias categorías de reformas, proporcióñalas en una lista separada por comas.
5. **"otra_nombre"**: Si la reforma fue categorizada como "otra", da un nombre a la categoría de la reforma. En caso contrario responde "no aplica".
6. **"etapa_reforma"**: Determina la etapa en que se encuentra la reforma, eligiendo entre las siguientes categorías: "discusión", "aprobación", "reversada", "otra", "no aplica".

7. **“nombre_reforma”**: Si se menciona en el artículo, proporciona el nombre exacto de la reforma, tal como “reforma tributaria” o “reforma de la ley 100”.
8. **“actores_involucrados”**: Proporciona una lista de los actores relevantes mencionados en el artículo que están involucrados en la reforma, separándolos por comas.
9. **“resumen”**: Proporciona un resumen conciso de una oración sobre el contenido principal del artículo.

Ejemplos del output correcto:

```
{
  "es_reforma": 1,
  "es_colombia": 1,
  "es_presente": 1,
  "tipo_reforma": "salud",
  "otro_nombre": "no aplica",
  "etapa_reforma": "discusión",
  "nombre_reforma": "reforma del sistema de salud",
  "actores_involucrados": "Ministro de Salud, Congreso",
  "resumen": "El artículo discute la nueva reforma del sistema de salud que está
↪ en etapa de discusión en el Congreso."
}

{
  "es_reforma": 1,
  "es_colombia": 1,
  "es_presente": 1,
  "tipo_reforma": "fiscal, educación",
  "otro_nombre": "no aplica",
  "etapa_reforma": "aprobación",
  "nombre_reforma": "",
  "actores_involucrados": "Ministro de Hacienda, Ministro de Educación,
↪ Congreso",
}
```

```
"resumen": "El artículo analiza las reformas fiscal y educativa que acaban de  
↪ ser aprobadas en el Congreso."  
}
```

Artículo:

A.3 Final GPT-4 prompt for generating reform process database

Sistema:

Eres un asistente experto en análisis de artículos de prensa, especializado en identificar y extraer información sobre reformas en Colombia en formato JSON. Tu trabajo requiere un alto nivel de precisión, ya que la exactitud en los detalles es crucial.

Asistente:

A continuación, se te proporcionará un artículo de prensa. Tu tarea es identificar y extraer información clave. Proporciona la información solicitada en formato JSON, asegurándote de que cada clave corresponde a una característica especificada y su valor es tu respuesta.

1. **es_reforma:** Indica si el artículo menciona alguna reforma. Usa 1 para "Sí" y 0 para "No".
2. **es_colombia:** Indica si el artículo se refiere a eventos o situaciones que ocurren en Colombia. Usa 1 para "Sí" y 0 para "No".
3. **es_presente:** Indica si el artículo se refiere a eventos o situaciones que están ocurriendo actualmente. Usa 1 para "Sí" y 0 para "No".
4. **es_opinion_entrevista:** Indica si el artículo es un artículo de opinión o una entrevista. Usa 1 para "Sí" y 0 para "No".
5. **es_nacional:** Indica si el artículo discute una reforma a nivel nacional. Usa 1 para "Sí" y 0 para "No".
6. **tipo_reforma:** Clasifica la reforma mencionada en una de las siguientes categorías: "fiscal", "educación", "salud", "política" o "judicial". Responde "otra" si la reforma discutida es claramente distinta a las categorías propuestas. Responde "no aplica" si el artículo no menciona una reforma. Usa solo estas categorías. No es posible responder con categorías diferentes. Si en el artículo se mencionan varias categorías de reformas, proporcióñalas en una lista separada por comas (por ejemplo: ["salud", "política"]).

7. **otro_nombre:** Si la reforma fue categorizada como “otra”, da un nombre a la categoría de la reforma. En caso contrario responde “no aplica”.
8. **intensidad_reforma:** Indica la intensidad con la que el artículo habla sobre el tipo de reforma identificado. Usa un número entre 0 y 100 para medir la intensidad, donde 100 significa que todo el artículo habla sobre esa reforma y 0 significa que ninguna parte del artículo habla sobre la reforma.
9. **etapa_reforma:** Determina la etapa en que se encuentra la reforma, eligiendo entre las siguientes categorías: “discusión” (se está hablando sobre una reforma pero todavía no hay debates legislativos), “debate” (la reforma se encuentra en alguna instancia del proceso legislativo), “aprobación” (la reforma fue aprobada por el ente competente), “rechazada” (la reforma fue rechazada o no prosperó), “reversada” (después de haber sido aprobada, la reforma fue reversada por algún ente), o “no aplica” (el artículo no habla sobre reformas).
10. **nombre_reforma:** Si se menciona en el artículo, proporciona el nombre exacto de la reforma, tal como “reforma tributaria” o “reforma de la ley 100”.
11. **actores_involucrados:** Proporciona una lista de los actores relevantes mencionados en el artículo que están involucrados en la reforma, separándolos por comas.
12. **resumen:** Proporciona un resumen conciso de una oración sobre el contenido principal del artículo.

Artículo: