Available online at www.ijournalse.org

Emerging Science Journal

(ISSN: 2610-9182)

Vol. 8, No. 1, February, 2024

Artificial Intelligence for Impact Assessment of Administrative Burdens

Victor Costa [1], Pedro Coelho [1], Mauro Castelli [1*]

[1] NOVA Information Management School (NOVA IMS), Universidade NOVA de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal.

Abstract

This study proposes the use of Artificial Intelligence (AI) to automatize part of the legislative impact assessment process. In particular, the focus of this study is the automatic identification of administrative burdens from legislative documents. The goal of impact assessment for administrative burdens is to apply an evidence-based approach toward compliance costs generated by regulation. Employing advanced Natural Language Processing (NLP) techniques based on a transformer architecture, a system was specifically developed and tested using Portuguese legislation. The experimental phase involved the system's ability to accurately and comprehensively identify administrative burdens. Experimental results demonstrated the system's effectiveness, showing its suitability for supporting the legislative impact assessment process by automating a time-consuming task. To the best of our knowledge, this is the first attempt concerning the use of AI for automatizing the identification of administrative burdens. The proposed system may provide governments and policymakers with a tool to speed up the legislative impact assessment process, thereby streamlining decision-making processes. Moreover, the use of AI can make the legislative impact assessment process less subjective, thus increasing its transparency and making citizens more confident about the impartiality of the process that leads to new legislation.

Keywords:

Impact Assessment;
Administrative Burdens;
Artificial Intelligence;
Natural Language Processing;
Transformers; BERT.

Article History:

Received:	20	October	2023
Revised:	11	January	2024
Accepted:	21	January	2024
Published:	01	February	2024

1- Introduction

Artificial Intelligence (AI) has radically changed our daily habits, and its impact is evident in e-commerce [1], engineering [2], medicine [3], and many more domains [4]. To obtain a competitive advantage, companies need suitable systems that can extract insights from data, and AI represents the ideal solution for this task [5]. Despite its successful application in different disciplines and private companies, AI is still not popular in public governance. A recent study [6] highlighted the existence of exploratory and conceptual research concerning the application of AI in this domain, thus suggesting that the use of AI in this area is still in its infancy. In particular, existing literature shows a lack of practical applications of AI in this area that, on the other hand, can substantially benefit from AI [7]. For instance, AI can modify traditional forms of policy-making, public procurement, and service provision in favor of more objective and data-oriented processes. For instance, AI-based systems may improve the quality of public services, transparency and accountability [8], and foster citizens' trust [9]. On the other hand, the use of AI represents a challenge for public institutions. The violation of citizens' privacy and the lack of fairness in the use of AI for public governance may hamper citizens' trust in public institutions [10, 11]. Anyway, the limited use of AI in public governance seems to be related to one main factor: compared to the private sector, where AI is nowadays widely used, existing literature highlights that there is less knowledge concerning AI opportunities in the context of the public sector [12, 13].

Among the opportunities offered by AI in the context of public governance, the generation of accurate forecasts and the experimentation of various policy options have seen increasing interest [14]. In this context, one of the most

[*] **CONTACT**: mcastelli@novaims.unl.pt

DOI: http://dx.doi.org/10.28991/ESJ-2024-08-01-019

challenging tasks of governments is the legislative process that, from the identification of an initial public policy objective, results in the definition of a new law. The complexity of this process is determined by several factors, including the impossibility of the legislator to foresee particular scenarios and the difficulty in evaluating the economic impact of a new law on the different sectors of the economy [15]. Thus, while regulation is a pillar of any modern society, providing the "rules of the game" for citizens, businesses, and governments, its impact might also reflect negatively upon a community. In the United States of America (USA), for instance, Carey [16] presented estimates of regulatory costs amounting from $ 260 billion to over $ 2 trillion. Another study suggests a worse scenario, stating that regulation has hampered an additional U$4 trillion in annual growth for the national economy [17].

Within the range of possible costs created by regulation, administrative burdens are a particularly controversial one [18]. These are bureaucracy-related demands imposed on society that may translate into significant hours invested in paperwork and similar activities. Resorting to compliance measures within regulation without proper assessment of their fundamental necessity has been shown to create several negative outcomes by George et al. [19]. The authors also highlight that such negative impacts are stable across sectors and administrative traditions, making it a universal issue. In 2015, for instance, 9.78 billion hours of government-imposed paperwork activities were registered in the USA alone [20].

As a global reference on best practices for regulatory design, the Organization for Economic Co-Operation and Development (OECD) defends that an evidence-based approach towards regulation can ensure better-quality government intervention [21]. Furthermore, the OECD states legislative impact assessment (simply "impact assessment" hereafter) as one of the most important regulatory tools available to governments, providing crucial information to decision-makers on whether and how to regulate to achieve public goals [21].

Within the umbrella of impact assessment and specific to analyzing administrative burdens, the Standard Cost Model (SCM) is a constantly employed method, especially among European countries and the European Union [18]. It measures the stock of administrative burdens stemming from regulation as monetary costs. Such information helps legislators decide whether to adopt or change laws based on quantitative insights. Therefore, incorporating this anticipatory mindset of analyzing regulation before its implementation may avoid legislative "mistakes". Nonetheless, these calculations are currently carried out manually, forcing impact assessment teams to choose a select number of regulations to analyze [21].

In the context of impact assessment, AI can play an important role. Recent advances in natural language processing [22, 23] allow extracting insights from textual documents, and AI models can be trained by taking advantage of the vast amount of data that is nowadays available. In particular, the definition of the transformer architecture in 2017 [24] revolutionized the area of natural language processing (NLP), making it possible to analyze long textual documents and achieve state-of-the-art results compared to the existing machine learning techniques. The transformer neural network aims to solve sequence-to-sequence tasks while handling long-range dependencies. Different from the encoder-decoder architectures, in the transformer architecture, the input sequence can be analyzed in parallel, thus reducing the time needed to train the model. Moreover, by relying on multi-headed attention layers [25], it overcomes previous limitations that allowed us to successfully analyze only short textual sequences. In this light, AI can support human experts in the tedious and time-consuming process of identifying administrative burdens.

Despite the opportunities offered by AI to support the legislative process, its usage is limited to other areas of public governance, like strengthening the immigration process control system and personalizing the digital service experience [26]. Thus, there is a need for an AI-based system that can support the legislative process and the economic assessment of new law proposals.

To answer this call, this paper investigates the feasibility of AI as a solution to scale the legislative impact assessment process. The proposed solution exploits recent advances in the area of NLP [23]. It employs a Deep Learning [27] model capable of interpreting legal texts and recognizing the patterns that communicate the presence of administrative burdens. Once mapped, the filtered-out text passages are presented to subject-matter experts so they can review and calculate costs accordingly.

To the best of our knowledge, this is the first attempt to support the process of identifying administrative burdens through AI. Therefore, the research aims to serve as a proof-of-concept on how AI may contribute to the implementation of impact assessment by scaling manual routines. An experimental evaluation was carried out in partnership with the Portuguese Technical Unit for Regulatory Impact Assessment to illustrate the solution's feasibility. As an overarching goal, this work should foster the adoption of evidence-based assessment on a greater stock of regulations.

The paper is organized as follows: Section 2 recalls recent works in which AI techniques were proposed for addressing problems in the area of public governance. Section 3 formally defines the AI task to be carried out. Also, this section reviews the data annotation process and the challenges faced with training data. Section 4 presents the methodology used for developing the experimental AI models. Section 5 outlines the results obtained, aiming to demonstrate the suitability of the proposed system. Finally, Section 6 concludes the paper by summarizing the main findings.

2- Related Works

This section reviews recent contributions concerning the application of AI techniques in the area of public governance. Finally, it presents existing initiatives concerning legislative impact assessment.

The development of technologies and the availability of mobile devices have created new opportunities for exploiting the vast amount of data collected every day. In particular, increased data availability allows for improving the quality of AI systems that are heavily data-dependent [28]. On the one hand, AI brings along challenges such as its ethical implications and the lack of models' explainability [29]. Due to these issues, particularly critical for governments, AI is mainly used in the private sector, and companies are considering the application of AI to extract insights from their data [5]. Conversely, as discussed in a recent survey [6], the use of AI in the public sector is still in its infancy.

Kuziemski & Misuraca [26] analyzed different case studies concerning the use of AI for the public sector in Canada, Finland, and Poland. After examining the legal and policy instruments associated with the case studies, he highlighted the need for a common framework to evaluate the potential impact of AI in the public sector. Moreover, the same paper discusses the effects of automated decision support systems and the crucial role of governments in the digital society to ensure that the potential of technology is harnessed while avoiding its negative effects. While Kuziemsky considered relevant applications of AI in public governance, the case studies are related to tasks not related to the legislative process. In particular, the paper considered the use of AI for strengthening the immigration process control system in Canada, the optimization of employment services in Poland, and the personalization of the digital service experience in Finland. In Portugal, the Portuguese Quality Institute (IPQ) developed, in 2020, an AI-based system to support its daily routines and predict the maintenance interval of measuring instruments.

In all the aforementioned tasks, AI systems are used to replace human experts in repetitive tasks or simple processes that are optimized based on the particular user. In other words, the considered use cases are limited to simple situations in which a learning algorithm can easily learn how to perform the task at hand. Similarly, Castelli et al. [30] proposed an AI-based framework supporting law-enforcement organizations in the process of analyzing crime data to extract useful knowledge for decision-makers. The application of AI in this context has received greater attention in recent years, thanks to the fast growth of urban populations and the development of smart cities. The first AI applications were limited, with most of the systems dealing with software tools for enabling cooperative information police departments [31] and the detection and analysis of specific events, such as vehicle movements and crowd dynamics [32, 33].

To the best of our knowledge, the use of AI to support the legislative process did not receive attention from the scientific community, and this fact is also corroborated by a recent work by Zuiderwijk et al. [6], in which the authors proposed a systematic literature review concerning the use of artificial intelligence in public governance. Based on their findings, they highlighted that empirical research on AI for public governance is needed. In fact, from the analysis of recent literature, one can observe that a relevant effort was dedicated to the main challenges of implementing artificial intelligence in public governance [10, 34–36] and to the definition of guidelines and research agendas for the adoption of AI by governments [9, 12, 37]. On the other hand, the use of AI focuses on simple applications related to chatbots or the prediction of some phenomenon by exploiting the available historical data [38]. Thus, based on the analysis of the scientific literature, one can observe that the use of AI to support the legislative process is still in an embryonic stage, with a clear need for practical contributions. At the European level, the existing studies/projects focus on suggesting strategies to enhance the analytical abilities of the national legislation offices [39]. The final goal is to lead to the adoption of appropriate methodological tools and processes for regulatory impact assessment to improve the country's overall regulatory performance. In this vein, two projects proposed for Croatia [39] and Germany [40] highlighted the need for a modern and automatic regulatory impact assessment by pointing out the potentiality of artificial intelligence.

Similarly, a report by Deloitte ("Kostbar study") showed that, in Germany, companies are subject to a considerable number of laws and regulations, with insurance and mechanical engineering companies spending between 4 and 7 percent of their annual personnel and material expenses to comply with the federal relevant regulations. These regulatory expenses can make or break a company's profitability. For this reason, the study highlighted the importance of reducing these costs by adopting NLP techniques that may simplify the existing regulations. In particular, when asked how legislation should be designed to reduce regulatory burdens, companies highlighted the importance of pursuing better regulation through advanced data science methods. However, despite the existence of these studies and initiatives, no further steps were taken toward the implementation of AI-based systems.

Karpen [41] described the current methods of regulatory impact assessment at European and Member State levels, and, in both cases, there is no evidence of AI-based systems supporting this task. Other initiatives highlighting the suitability of AI for improving public governance and regulatory impact assessment exist [42], but, to the best of our knowledge, no country has moved towards the implementation of automatic systems for supporting this task. This is mainly due to the lack of expertise as well as the difficulty in foreseeing the benefits provided by AI in this area.

Moving from the European to the global level, the situation does not change. A recent report by the World Bank titled "Global Indicators of Regulatory Governance: Worldwide Practices of Regulatory Impact Assessments" reported a list of countries that improved their regulatory impact assessment practices and how they achieved this result. Interestingly, all the developments were achieved by modifying existing assessment procedures, and there is no evidence of AI-based systems supporting the task at hand.

3- Tasks and Challenges

In the Portuguese scenario, impact assessment experts set out to identify the different Information Obligations (IOs) present within regulation. These are requirements imposed by law that determine companies must either collect, store, and/or submit data and information to public entities (or third parties). One example of an IO is a retail store being required to offer within its facilities a way for clients to report their dissatisfaction with a product or service — analogous to a “complaints book”.

From the identified IO(s) in a regulation, experts translate administrative burdens into monetary values by mapping the IO(s) with a set of standardized costs. These are previously estimated together with national private companies, where the idea is to calculate the time spent and, consequently, the financial cost of complying with the identified obligations.

3-1- AI Task Definition

Given a legislative document as input, the main objective of the proposed solution is to output passages that indicate a high probability of containing administrative burdens.

In more detail, the proposed AI-based system answers the need for understanding text at scale by employing natural language processing techniques. Concretely, the objective is to automatize the analysis of legislative documents, aiming at identifying paragraphs containing information obligations. To achieve this objective, NLP model(s) will be trained to identify information obligations within legislative documents, thus posing a text classification problem.

Recent advances in Deep Neural Networks, such as the Transformer architecture [24] and large, pre-trained language models [43-46] are considered the state-of-the-art in NLP and, therefore, will be considered for building the envisioned system. The core idea is to take advantage of a Transfer Learning setting [47], where knowledge previously learned by the model from a vast amount of generic data is repurposed and fine-tuned for a specific downstream task.

Additionally, Gururangan et al. [48] have demonstrated that adapting the pre-trained model (trained on a generic corpus) to the downstream task domain—i.e., instead of directly applying it “out-of-the-box"—can be beneficial due to the differences in the vocabulary between generic data and the specific domain of the task. Specifically, Chalkidis et al. [49] demonstrated notable enhancements through domain adaptation in legal-related downstream tasks, affirming the existence of distinct distributions between legislative texts and common web corpora. Therefore, following such recommendations from the scientific literature, the same adaptation will be investigated by further pre-training a Portuguese language model in the legal domain. Since the considered pre-trained model implements a BERT-like architecture [43], the custom Portuguese legal BERT model will be adapted through masked language modeling [50].

3-2- Data Annotation of Administrative Burdens

Machine learning models require annotated (or “labeled) data to learn a predictive task. In the context of this study, deep learning models are developed to address a supervised learning task (the identification of information obligations lying in a legal text). In this paradigm, models learn from a labeled dataset, which encompasses input data (e.g., text documents) and corresponding output labels (in the considered applications, specific information obligations). These labels act as the ground truth, enabling the model to discern intricate patterns and associations between input features and output labels. Essentially, without labeled data, the model's ability to generalize and make precise predictions would be severely compromised.

Recently, multiple techniques that aim at mitigating the need for labeled data—be it in terms of the amount of data or the amount of time invested by the annotator—have been receiving greater attention from the scientific community. Weak supervision [51], for instance, builds training sets through labeling functions—e.g., distilled experts’ annotation logic that works as computer programs (functions)—that label an entire dataset instantly (although with less precision than a manual process). Another technique is Active Learning [52, 53], which minimizes the amount of manual labeling needed to achieve satisfactory results by choosing samples that lie close to the model’s decision boundary and asking the human annotator to label them correctly.

While these techniques can be investigated to collect more data, none of them eliminates the need for some “expert-approved” labeled data: a high-quality supervised sample that shall serve as ground truth for the AI models' development. Therefore, this section discusses the protocol followed to create a ground-truth dataset for training the NLP models.

The data gathered to support the experimental evaluation of the proposed text classification task was acquired from a random sample of Portuguese laws*. Specifically, 41 documents were annotated by two impact assessment experts, where each relevant text passage identified within the corpora was labeled with a positive or negative label — i.e. containing or not a burden. The time experts took to annotate a full document oscillated on a scale of hours. Extreme cases in terms of document size, such as with Decree-Law 91/2018†, actually required days for a single document. The annotators' experience helps illustrate the challenges of manually identifying administrative burdens.

Further on, challenges exist from the perspective of training data. All annotated documents combined resulted in 322 examples of text passages containing administrative burdens. At the same time, the combined corpus amounted to a total of 19,465 text passages, which depicts a considerable imbalance between passages with and without administrative burdens—on average, only 1.68% of documents' passages contained IOs.

This scarcity of labeled data is not a challenge specific to this research. In fact, many scholars defend that AI's true bottleneck is not the lack of performant models anymore but the lack of high-quality training data instead [51, 54]. Not only it is time-consuming to acquire all the necessary data, but it can also be unfeasibly expensive to have subject matter experts dedicated to the annotation process. Therefore, several techniques to mitigate the issue were employed and are generalized within this study to serve as a template for other impact assessment scenarios.

4- Research Methodology

This section presents the steps taken to develop the NLP model capable of identifying information obligations. Furthermore, it reviews how the encountered data challenges—class imbalance and label scarcity—informed the decisions behind the implemented methods. Figure 1 presents a systemic view of the process. The continuation of this section describes all the steps characterizing Figure 1.

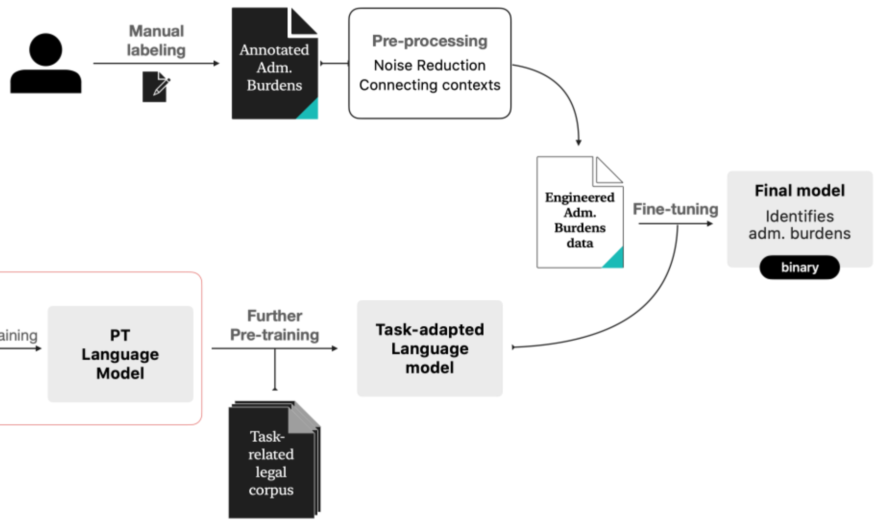

Figure 1. Flowchart of the model's development process

4-1- Training Data

As mentioned in Section 4.2, 41 documents (unique identifiers are listed in the Appendix I) were annotated with administrative burdens. To build the training data, all passages within those documents were combined into one dataset. Each sample is a sentence accompanied by a binary target indicating the presence or absence of an administrative burden. Table 1 presents the target feature's class counts, highlighting the label imbalance.

Table 1. Target feature's class count of the administrative burdens annotated data

Class	Count	%
With administrative burden	322	1.68
Without administrative burden	19143	98.32

Two methods, described in sections 4.1.1 and 4.1.2, were employed with the sole purpose of alleviating label imbalance while also improving the data quality. These techniques were tailored specifically to the present research. However, both follow patterns that are shared among regulatory documents' structures and, therefore, can serve as a blueprint for other researchers.

* Portuguese regulation is publicly available at https://dre.pt/

† https://dre.pt/dre/detalhe/decreto-lei/91-2018-116936932

4-1-1- Noise Reduction

The Noise Reducing module focuses on removing passages from the input corpus that were previously known to not contain administrative burdens. It follows a series of rules defined together with impact assessment experts, leveraging their prior knowledge of the laws' structure. It was deliberately designed as an automated, algorithmic technique—i.e., instead of a manual cleaning of the data—to serve as a viable option to be included in a real scenario. The resulting algorithm's functioning is summed up in the following:

- Automatically identifies and discards titles, subtitles, numbers-only, and symbols-only passages, among other content-poor text snippets.
- Documents' introductions are also disregarded since administrative burdens are only stated within the regulation's articles.
- In the case of a regulation altering an existing legal document, the module is capable of identifying only the new passages and ignoring outdated ones.
- Finally, and most significantly, it extracts entire articles that indicate the absence of administrative burdens through their underlying semantic context, e.g., articles referring to the law's general logistics.

As a means of comparison, during the experimental evaluation, the Noise Reducing module was able to alleviate target imbalance, increasing the average percentage of passages with administrative burdens within documents from 1.68% to 5.89%.

4-1-2- Connecting Split Contexts

The Portuguese legislative documents, as with regulatory texts from many entities, have a reoccurring format of splitting a concept (or message) between a paragraph's initial statement and continuing it through subparagraphs. Figure 2 illustrates how the initial statement's logic is interrupted and continued on the subparagraph with the marker "c". The issue to observe in these situations is that the algorithm (discussed in the next section) treats each one of the split passages as an independent input, disrupting the semantic composition of the intended message. Critical to this research's end goal is that many instances of annotated burdens have their information obligation's "trigger" and subject within the statement, with the information element being required within one of the subparagraphs. Figure 2 is also an example of a split administrative burden: "must comply" defines the obligation's trigger, but the word "list" characterizes it as demanded information.

Vessels must comply with the following:
a) …
b) …
c) A list must be posted, next to the communications equipment, with the contact details of the entities to be used in case of emergency;

Figure 2. Example of an administrative burden being split between initial statement and subparagraph

This motivated us to combine subparagraphs with their corresponding statements, preserving the input's complete context and the administrative burden. Using the same example, the transformed input in Figure 2 reads as: "*Vessels must comply with the following: a list must be posted next to the communications equipment with the contact details of the entities to resort to in case of emergency*."

4-2- Algorithm

In a classical supervised learning paradigm, as with our text classification task, Machine Learning models still heavily rely on labeled data for training. Again, this translates into a time-consuming and, many times, expensive process of manually annotating text, as observed with the labeling of administrative burdens. Transfer Learning [47] offers a response to mitigate such a limitation, leveraging data from other domains to train models that present improved generalization ability. Such models can serve as a starting point to be later customized for specific tasks.

In recent years, the field of NLP has capitalized on transfer learning techniques by using vast amounts of freely available text from the web to yield pre-trained models that significantly improved upon state-of-the-art results for several NLP tasks. Devlin et al. [43], Howard and Ruder [55], Peters et al. [56], and Radford et al. [57] provided empirical confirmation on the benefit of training an initial neural network on large, unlabeled corpora to be later on customized (or fine-tuned, as commonly referred to) on downstream supervised tasks. Only now is a fraction of the labeled data required to achieve satisfactory results.

As a general rule, the majority of pre-trained models recently published to serve transfer learning purposes follow two main concepts: an architecture based on the Transformer [24] and the use of large generic corpora for pretraining. This research exploits such performance improvement by using a BERT pre-trained model [43] as "base knowledge" and later fine-tuning it on the administrative burdens' labels. Figure 3 illustrates the concept.

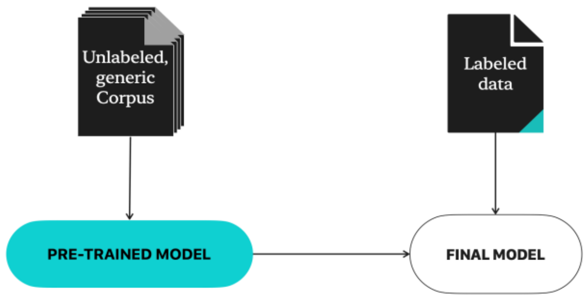

Figure 3. Conceptual visualization of Transfer Learning within NLP

Within the scope of this project, the pre-trained model used was developed by *Unicamp* (State University of Campinas, Brazil) in partnership with the University of Waterloo in Canada and is available in the open repository Hugging Face*. An interesting aspect to note about this pre-trained model (and others that follow the same logic) is the vast amount of data used during training, a corpus composed of 2.7 billion tokens, and the size of the model itself—around 110 million parameters.

4-2-1- Further Pre-training

Moving from the state-of-the-art performance that transfer learning and transformer models achieved on multiple NLP tasks, Gururangan et al. [48] have demonstrated that adapting the pre-trained model to the downstream task's domain can be beneficial. Additionally, Chalkidis et al. [49] have also shown improvement in law-related downstream tasks, reinforcing the notion of differing distributions among legislative texts and common web corpora.

Still, Gururangan et al. [48] also reinforce that further pre-training on a smaller corpus, yet directly related to the task, has proven effective for achieving improved downstream results. Referred to as task-adaptive pre-training, as opposed to the slightly more general domain-adaptive pre-training, the unlabeled dataset is drawn from the task's distribution.

This research follows the path of task adaptation by further pre-training the Portuguese BERT model on 262 legal documents extracted from the same source as the supervised data. In this step, the exact same documents that were intended for fine-tuning have not been selected, but both sets go through the same pre-processing. Therefore, along with a shared origin, the pre-processing techniques that surface text passages relevant for administrative burdens—e.g., removing documents' introductions and specific non-related articles—are applied to both. Performances of the initial BERT's base model and the BERT Task-adapted model are reported in Section 5.

4-3- Experimental Setup

For the purpose of task adaptation, the Portuguese BERT model was further pre-trained on 262 documents during 100 epochs with a learning rate equal to 5e-5 and the AdamW optimizer, i.e., the Adam algorithm with weight decay fix [58]. Whole-word masking was applied, with a masking probability of 0.15. The tokenized sentences were concatenated to form batches of 512 tokens, as per BERT Base's maximum length accepted. 15% of the task-related corpus was held out to evaluate the models' cross-entropy loss in predicting randomly masked tokens (simply, MLM Loss).

Following task adaptation, both language models (BERT Base and BERT Task-adapted) were fine-tuned on the labeled data of Portuguese administrative burdens to assess the impact of further pre-training. Considering the trade-off between performance and running time, the following hyperparameters were used: 3 epochs, a batch size of 16 (8 x 2 gradient accumulation steps), and a learning rate equal to 5e-5 with the AdamW optimizer. 20% of the supervised data was held out for testing, using a stratified split on the target feature. The following section presents the performance of the models in terms of F1-score, Precision, and Recall.

5- Results and Discussion

Table 2 presents the MLM Loss of both language models on the held-out set of the task-related corpus. The significant reduction in loss after task-adaptive pre-training suggests the initial BERT Base's distribution over its original web corpus considerably differs from the Portuguese regulation. That is to say, the high loss value indicates that BERT Base was often unsure of which words best replaced the masked tokens, consequently increasing the cross-entropy loss. BERT

* https://huggingface.co/neuralmind/bert-base-portuguese-cased

Task-adapted's small loss value suggests the opposite: it is able to assign high probabilities, thus indicating the model has learned patterns within the task-related corpus. As the regulatory documents present more formal writing with more complex sentence structures and domain-specific terms—along with the regional difference (BERT Base was trained on Brazilian Portuguese)—these may result in an increase in BERT Base's MLM Loss.

Table 2. MLM Loss on the held-out set of the task-related corpus

Model	Loss
BERT Base	9.327
BERT Task-adapted	0.315

The experimental results are aligned with the theory behind task adaptation in deep learning models like BERT. In particular, task adaptation helps the model specialize and improve its performance on the target task, thus allowing it to achieve better performance on the considered task. In more detail, models like BERT are pre-trained on large, diverse corpora using unsupervised learning objectives (e.g., masked language modeling or next-sentence prediction). Thus, these models can learn general language representations that capture a wide range of linguistic patterns and semantics. However, each specific task or domain is characterized by its own specific nuances, vocabulary, or contextual understandings that cannot be fully captured with a general pre-trained model. In this sense, task adaptation allows for leveraging the pre-trained model's general knowledge and adapting it to the specific features of the target task.

Concerning this study, a base BERT model, pre-trained using Portuguese vocabulary, was initially considered. However, as it emerged from the experimental results, this general-purpose model cannot deal with the specific terminology characterizing legislative documents. As can be observed from the experimental results, the task adaptation process adjusted the weights of the base BERT model, thus allowing it to better represent the specific features of the considered task. Thus, by adapting the base BERT model to the domain considered in this study (i.e., legislative documents and information obligations), it is possible to capture domain-specific terminology, jargon, or context, which are crucial for better performance in that domain.

Finally, the choice to rely on task adaptation is motivated by its ability to retain the knowledge gained during the pre-training phase while allowing the model to specialize in the new task. This is a fundamental property that prevents the loss of general information (in this case, concerning the semantics of Portuguese sentences and words) while the model is adjusting its weight to target the specific task under exam.

Table 3 shows a significant improvement in the downstream task after task-adaptive further pre-training, reinforcing the notion of new task-specific knowledge acquired. On the other hand, the small amount of labeled data for fine-tuning might not be enough to customize BERT Base on the downstream task to achieve satisfactory results.

Table 3. Fine-tuned models' results on unseen data (binary target). "F1", "P", and "R" specify F1, Precision, and Recall scores, respectively

Model	F1	P	R
BERT Base – fine-tuned	0.5	0.524	0.478
BERT Task-adapted – fine-tuned	0.794	0.833	0.758

From the perspective of impact assessment as a practice, these initial results suggest a promising path. With only 41 annotated documents, the best model was able to achieve an F1-score of 0.794. Yet, additional labeled data might corroborate significant performance improvements. This represents an opportunity to simplify the impact assessment process, alleviating experts from a time-consuming step while calculating administrative costs.

Moreover, there is room for improvement. For instance, it would be possible to tune the hyperparameters to eventually achieve better results. In the context of this research, a full exploration of the hyperparameter values was not performed due to limited computational resources. Additionally, the main objective of this study was to provide a proof-of-concept to assess the suitability of AI-based methods for the identification of administrative burdens.

To finish, the system is currently able to identify, on average, 75.8% of the administrative burdens present within a legal document (i.e., Recall score). On the other hand, the Precision score indicates that of all the passages returned by the system, 83.3% actually contain a burden. Although a good balance is observed between both scores, for a real scenario, one could argue that it is preferable to have a higher Recall, even if it means sacrificing some Precision. In fact, it is less time-consuming to review whether the identified passages have an administrative burden than to look for those missed by the model within the entire text. In future work, considering a weighted score, such as the F2-score, might be beneficial for better responding to experts' needs.

All in all, considering the results achieved in this study, it is possible to identify several benefits associated with the use of AI for supporting the legislative impact assessment process. First of all, efficiency emerges as one of the most important benefits: the proposed system was able to enhance government operations' efficiency by automating the time-consuming process of identifying information obligations within a legislative document. In this sense, the proposed system may open new perspectives and discussions concerning the use of AI in public governance.

Second, the proposed AI-based system may lead to economic benefits by minimizing processing times and reducing costs. Moreover, by integrating the AI system within the existing procedures performed in the legislative impact assessment process, AI can help reduce administrative burdens, automate routine actions, save hours worked, or reduce administrative expenses.

Third, the proposed system results in data and information processing benefits. In particular, impact assessment experts can rely on AI to analyze and extract useful insights from vast amounts of legislative documents in a limited time. Thus, considering that the proposed AI system analyzes a legislative document in a few seconds (instead of the hours required by a human expert), it is possible to perform the impact assessment process on a vast number of legislative documents, thus improving the quality of the legislation by reducing the administrative burdens for citizens.

Finally, the proposed AI system makes the legislative impact assessment process less subjective, thus increasing its transparency. In particular, AI can foster citizen trust by enabling individuals to witness the evolution of legislation and understand its implications. This transparency makes citizens more confident about the impartiality of the process that leads to new legislation, thus increasing trust between governance and society.

6- Conclusion

The primary objective of this study is to streamline some steps of the impact assessment process, supporting human experts in identifying administrative burdens embedded within legislative documents. The paper introduces an artificial intelligence-based system to address the complexities inherent in this task, shedding light on potential challenges that may arise. Notably, issues such as class imbalance and a scarcity of training data emerged due to the nature of administrative burdens within legislative documents. To overcome these issues, the study employed pre-processing techniques designed for legal documents and implemented transfer learning techniques specifically adapted for the task under exam. The use of these techniques mitigated the data imbalance issue and allowed the proposed AI-based system to produce satisfactory performance.

In particular, the proposed system has been tested on Portuguese legislative documents to assess its suitability for identifying paragraphs containing administrative burdens. The experimental results show that the proposed system achieves promising performance (considering precision, recall, and F1) and correctly identifies the majority of administrative burdens characterizing the considered legislative documents. From the analysis of the results, it emerged that task adaptation was fundamental to improving the performance of the base BERT model considered in this study.

All in all, the results demonstrated the suitability of AI to support the legislative impact assessment process; in particular, AI can reduce the burden faced by impact assessment experts in the identification and assessment of administrative costs.

Finally, it is possible to hypothesize that better results can be achieved when considering a larger corpus of legislative documents for fine-tuning the transformer-based models. In fact, the promising results reported in this study were achieved by only considering a limited number of training observations due to the difficulty in collecting labeled data. Thus, future research will explore methods involving weak supervision to increase the number of labeled training observations and automate the labeling process of paragraphs within legislative documents.

7- Declarations

7-1- Author Contributions

Conceptualization, V.C., P.C., and M.C.; methodology, V.C.; software, V.C.; validation, V.C., P.C., and M.C.; formal analysis, V.C., P.C., and M.C.; investigation, V.C.; resources, P.C. and M.C.; writing—original draft preparation, V.C.; writing—review and editing, V.C., P.C., and M.C.; visualization, V.C.; supervision, M.C.; project administration, P.C., and M.C.; funding acquisition, P.C. and M.C.; All authors have read and agreed to the published version of the manuscript.

7-2- Data Availability Statement

The data presented in this study are listed in Appendix I.

7-3- Funding

This work was supported by national funds through the FCT (Fundação para a Ciência e a Tecnologia) by the project UIDB/04152/2020-Centro de Investigação em Gestão de Informação – MagIC/NOVA IMS. This work was performed in the context of the the project "AI2A – Avaliação de Impacto e Inteligência Artificial" (POCI-05-5762-FSE-000226), funded by the program PORTUGAL 2020.

7-4- Institutional Review Board Statement

Not applicable.

7-5- Informed Consent Statement

Not applicable.

7-6- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

8- References

[1] Vanneschi, L., Horn, D. M., Castelli, M., & Popovič, A. (2018). An artificial intelligence system for predicting customer default in e-commerce. Expert Systems with Applications, 104, 1–21. doi:10.1016/j.eswa.2018.03.025.

[2] Castelli, M., Vanneschi, L., & Silva, S. (2013). Prediction of high performance concrete strength using Genetic Programming with geometric semantic genetic operators. Expert Systems with Applications, 40(17), 6856–6862. doi:10.1016/j.eswa.2013.06.037.

[3] Castelli, M., Vanneschi, L., & Silva, S. (2014). Prediction of the Unified Parkinson's Disease Rating Scale assessment using a genetic programming system with geometric semantic genetic operators. Expert Systems with Applications, 41(10), 4608–4616. doi:10.1016/j.eswa.2014.01.018.

[4] Besozzi, D., Manzoni, L., Nobile, M. S., Spolaor, S., Castelli, M., Vanneschi, L., Cazzaniga, P., Ruberto, S., Rundo, L., & Tangherloni, A. (2020). Computational Intelligence for Life Sciences. Fundamenta Informaticae, 171(1–4), 57–80. doi:10.3233/FI-2020-1872.

[5] Popovič, A., Hackney, R., Tassabehji, R., & Castelli, M. (2018). The impact of big data analytics on firms' high value business performance. Information Systems Frontiers, 20(2), 209–222. doi:10.1007/s10796-016-9720-4.

[6] Zuiderwijk, A., Chen, Y. C., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. Government Information Quarterly, 38(3), 101577. doi:10.1016/j.giq.2021.101577.

[7] Sousa, W. G. de, Melo, E. R. P. de, Bermejo, P. H. D. S., Farias, R. A. S., & Gomes, A. O. (2019). How and where is artificial intelligence in the public sector going? A literature review and research agenda. Government Information Quarterly, 36(4), 101392. doi:10.1016/j.giq.2019.07.004.

[8] Ojo, A., Mellouli, S., & Ahmadi Zeleti, F. (2019). A Realist Perspective on AI-era Public Management. Proceedings of the 20th Annual International Conference on Digital Government Research, 159-170. doi:10.1145/3325112.3325261.

[9] Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., … Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, 101994. doi:10.1016/j.ijinfomgt.2019.08.002.

[10] Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. Government Information Quarterly, 36(2), 368–383. doi:10.1016/j.giq.2018.09.008.

[11] Aoki, N. (2020). An experimental study of public trust in AI chatbots in the public sector. Government Information Quarterly, 37(4), 101490. doi:10.1016/j.giq.2020.101490.

[12] Wirtz, B. W., Weyerer, J. C., & Sturm, B. J. (2020). The Dark Sides of Artificial Intelligence: An Integrated AI Governance Framework for Public Administration. International Journal of Public Administration, 43(9), 818–829. doi:10.1080/01900692.2020.1749851.

[13] Wang, W., & Siau, K. (2018). Artificial intelligence: a study on governance, policies, and regulations. Proceedings of the Thirteenth Midwest Association for Information Systems Conference, 17-18 May, 2018, Saint Louis, United States.

[14] Margetts, H., & Dorobantu, C. (2019). Rethink government with AI. Nature, 568(7751), 163–165. doi:10.1038/d41586-019-01099-5.

[15] Coglianese, C. (2012). Measuring Regulatory Performance - Evaluating the Impact of Regulation and Regulatory Policy. Organisation for Economic Co-operation and Development (OECD), Paris, France.

[16] Carey, M. P. (2016). Methods of estimating the total cost of Federal regulations. Congressional Research Service, Washington, United States.

[17] Coffey, B., McLaughlin, P. A., & Peretto, P. (2020). The cumulative cost of regulations. Review of Economic Dynamics, 38. doi:10.1016/j.red.2020.03.004.

[18] Coletti, P. (2016). Standard cost model. In Handbook of Regulatory Impact Assessment. Edward Elgar Publishing, Cheltenham, United Kingdom. doi:10.4337/9781782549567.00019.

[19] George, B., Pandey, S. K., Steijn, B., Decramer, A., & Audenaert, M. (2021). Red Tape, Organizational Performance, and Employee Outcomes: Meta-analysis, Meta-regression, and Research Agenda. Public Administration Review, 81(4), 638–651. doi:10.1111/puar.13327.

[20] United States Office of Management. (2015). Budget of the United States Government. US Government Printing Office, Washington, United States.

[21] Meuwese, A., Scheltema, M., & van der Velden, L. (2015). The OECD framework for regulatory policy evaluation: An initial assessment. European Journal of Risk Regulation, 6(1), 101–110. doi:10.1017/S1867299X00004323S1867299X00004323.

[22] Otter, D. W., Medina, J. R., & Kalita, J. K. (2021). A Survey of the Usages of Deep Learning for Natural Language Processing. IEEE Transactions on Neural Networks and Learning Systems, 32(2), 604–624. doi:10.1109/TNNLS.2020.2979670.

[23] Qiu, X. P., Sun, T. X., Xu, Y. G., Shao, Y. F., Dai, N., & Huang, X. J. (2020). Pre-trained models for natural language processing: A survey. Science China Technological Sciences, 63(10), 1872–1897. doi:10.1007/s11431-020-1647-3.

[24] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30, 4-9 December, 2017, Long Beach, United States.

[25] Hu, D. (2020). An Introductory Survey on Attention Mechanisms in NLP Problems. Intelligent Systems and Applications, IntelliSys 2019, Advances in Intelligent Systems and Computing, 1038. Springer, Cham, Switzerland. doi:10.1007/978-3-030-29513-4_31.

[26] Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. Telecommunications Policy, 44(6), 101976. doi:10.1016/j.telpol.2020.101976.

[27] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. doi:10.1038/nature14539.

[28] Singh, P. (2023). Systematic review of data-centric approaches in artificial intelligence and machine learning. Data Science and Management, 6(3), 144–157. doi:10.1016/j.dsm.2023.06.001.

[29] Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115. doi:10.1016/j.inffus.2019.12.012.

[30] Castelli, M., Sormani, R., Trujillo, L., & Popovič, A. (2017). Predicting per capita violent crimes in urban areas: an artificial intelligence approach. Journal of Ambient Intelligence and Humanized Computing, 8(1), 29–36. doi:10.1007/s12652-015-0334-3.

[31] Redmond, M., & Baveja, A. (2002). A data-driven software tool for enabling cooperative information sharing among police departments. European Journal of Operational Research, 141(3), 660–678. doi:10.1016/S0377-2217(01)00264-8.

[32] Doulaverakis, C., Konstantinou, N., Knape, T., Kompatsiaris, I., & Soldatos, J. (2011). An Approach to Intelligent Information Fusion in Sensor Saturated Urban Environments. 2011 European Intelligence and Security Informatics Conference, Athens, Greece. doi:10.1109/eisic.2011.24.

[33] Cho, H., Seo, Y. W., Kumar, B. V. K. V., & Rajkumar, R. R. (2014). A multi-sensor fusion system for moving object detection and tracking in urban driving environments. Proceedings - IEEE International Conference on Robotics and Automation, 1836–1843. doi:10.1109/ICRA.2014.6907100.

[34] Mikhaylov, S. J., Esteve, M., & Campion, A. (2018). Artificial intelligence for the public sector: opportunities and challenges of cross-sector collaboration. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2128), 20170357. doi:10.1098/rsta.2017.0357.

[35] Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. Government Information Quarterly, 37(3), 101493. doi:10.1016/j.giq.2020.101493.

[36] Liu, H. W., Lin, C. F., & Chen, Y. J. (2019). Beyond state v loomis: Artificial intelligence, government algorithmization and accountability. International Journal of Law and Information Technology, 27(2), 122–141. doi:10.1093/ijlit/eaz001.

[37] Chen, T., Ran, L., & Gao, X. (2019). AI innovation for advancing public service. Proceedings of the 20th Annual International Conference on Digital Government Research, 100-108. doi:10.1145/3325112.3325243.

[38] Kouziokas, G. N. (2017). The application of artificial intelligence in public administration for forecasting high crime risk transportation areas in urban environment. Transportation Research Procedia, 24, 467–473. doi:10.1016/j.trpro.2017.05.083.

[39] Romic, D., & Halak, Z. V. (2014). Regulatory Impact Assessment in The Republic of Croatia-Situation and Perspective. Challenges of the Knowledge Society, Nicolae Titulescu University Editorial House, Bucharest, Romania.

[40] Zeitz, D. (2016). Better regulation in Germany as quality assurance system: Recent development and current challenges. Netherlands Administrative Law Library (Nall). doi:10.5553/nall/.000025.

[41] Karpen, U. (2017). Regulatory impact assessment: current situation and prospects in the German Parliament. Amicus Curiae, 2015(101), 14. doi:10.14296/ac.v2015i101.2342.

[42] Kosach, I., Shaposhnykov, K., Chub, A., Yakushko, I., Kotelevets, D., & Lozychenko, O. (2022). Regulatory policy in the context of effective public governance: evidence of Eastern European Countries. Cuestiones Políticas, 40(72), 456–473. doi:10.46398/cuestpol.4072.26.

[43] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. doi:10.48550/arXiv.1810.04805.

[44] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized BERT pretraining approach. arXiv preprint, arXiv:1907.11692. doi:10.48550/arxiv.1907.11692.

[45] Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. Advances in neural information processing systems, 32, 8-14 December, 2019, Vancouver, Canada.

[46] Clark, K., Luong, M. T., Le, Q. V., & Manning, C. D. (2020). Electra: Pre-training text encoders as discriminators rather than generators. arXiv preprint arXiv:2003.10555. doi:10.48550/arXiv.2003.10555.

[47] Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345–1359. doi:10.1109/tkde.2009.191.

[48] Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., & Smith, N. A. (2020). Don't Stop Pretraining: Adapt Language Models to Domains and Tasks. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. doi:10.18653/v1/2020.acl-main.740.

[49] Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). LEGAL-BERT: The Muppets straight out of law school. arXiv preprint arXiv:2010.02559. doi:10.48550/arXiv.2010.02559.

[50] Sinha, K., Jia, R., Hupkes, D., Pineau, J., Williams, A., & Kiela, D. (2021). Masked Language Modeling and the Distributional Hypothesis: Order Word Matters Pre-training for Little. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. doi:10.18653/v1/2021.emnlp-main.230.

[51] Ratner, A. J., De Sa, C. M., Wu, S., Selsam, D., & Ré, C. (2016). Data Programming: Creating Large Training Sets, Quickly. Advances in Neural Information Processing Systems, 29, 5-10 December, 2016, Barcelona, Spain.

[52] Settles, B. (2012). Active learning (synthesis lectures on artificial intelligence and machine learning). Morgan Claypool Publishers, San Rafael, United States.

[53] Ren, P., Xiao, Y., Chang, X., Huang, P. Y., Li, Z., Gupta, B. B., Chen, X., & Wang, X. (2022). A Survey of Deep Active Learning. ACM Computing Surveys, 54(9), 1–40. doi:10.1145/3472291.

[54] Varma, P., & Re´, C. (2018). Snuba: Automating weak supervision to label training data. Proceedings of the VLDB Endowment, 12(3), 223–236. doi:10.14778/3291264.3291268.

[55] Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. doi:10.18653/v1/p18-1031.

[56] Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. NAACL HLT 2018-2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies-Proceedings of the Conference, 1-6 June, 2018, New Orleans, United States.

[57] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. Preprint, 1-12. Available online: https://www.mikecaptain.com/resources/pdf/GPT-1.pdf (accessed on May 2023).

[58] Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. arXiv preprint, arXiv:1711.05101. doi:10.48550/arXiv.1711.05101.

Appendix I: List of Annotated Documents

Decreto-Lei n.º 10/2019

Decreto-Lei n.º 102-D/2020

Decreto-Lei n.º 108/2009

Decreto-Lei n.º 128/2014

Decreto-Lei n.º 136/2019

Decreto-Lei n.º 149/2014

Decreto-Lei n.º 156/2005

Decreto-Lei n.º 159/2014

Decreto-Lei n.º 169/2012

Decreto-Lei n.º 17/2018

Decreto-Lei n.º 178/2006

Decreto-Lei n.º 225/2006

Decreto-Lei n.º 29/2008

Decreto-Lei n.º 43/2018

Decreto-Lei n.º 50/2013

Decreto-Lei n.º 52/2018

Decreto-Lei n.º 53/2019

Decreto-Lei n.º 555/99

Decreto-Lei n.º 67/2019

Decreto-Lei n.º 7/2019

Decreto-Lei n.º 73/2007

Decreto-Lei n.º 73/2020

Decreto-Lei n.º 78/2018

Decreto-Lei n.º 8/2007

Decreto-Lei n.º 80/2017

Decreto-Lei n.º 80/2018

Decreto-Lei n.º 80/2019

Decreto-Lei n.º 82/2019

Decreto-Lei n.º 83/2019

Decreto-Lei n.º 87/2018

Decreto-Lei n.º 91/2018

Lei n.º 32/2019

Lei n.º 98/2019

Portaria n.º 1069/97

Portaria n.º 1320/2008

Portaria n.º 201-A/2017

Portaria n.º 281/2015

Portaria n.º 307/2015

Portaria n.º 358/2009

Portaria n.º 651/2009

Portaria n.º 937/2008