

On the Brazilian Observatory for Artificial Intelligence

Rafael Meira Silva,
Cristina Godoy Oliveira
CIAAM, C4AI, Univ. of São Paulo
São Paulo, Brazil
rafael@meirasilva.com.br
cristinagodoy@usp.br

Luiz Costa, Alexandre Barbosa
CETIC, OBIA
São Paulo, Brazil
tuca@nic.br
alexandre@nic.br

Joao Paulo Candia Vieira
CIAAM, C4AI, Univ. of São Paulo
São Paulo, Brazil
candia@usp.br

Joao Pita Costa*
IRCAI, Quintelligence
Ljubljana, Slovenia
Joao.pitacosta@quintelligence.com

ABSTRACT

Artificial Artificial Intelligence (AI) is rapidly transforming industries and economies worldwide, with Brazil and South America emerging as significant players in this global shift. The fundamental need to monitor the impact of artificial intelligence (AI) in the verticals for sustainable development, government engagement, investment and society at large motivated the Brazilian Artificial Intelligence Observatory (OBIA). It is also an integral part of the Brazilian Artificial Intelligence Plan (PBIA), and a former objective of the Brazilian Strategy of AI aims to become the leading platform for monitoring the uses of AI in the country. OBIA is part of Axis 5 of the PBIA focused on supporting the regulatory and governance process of AI. This research paper explores the current state, challenges, and potential of AI development in the region, examining how technological advancements are influencing economic growth, societal change, and policy-making across South America, with a particular focus on Brazil as a leading hub of innovation. It is also investigating common aspects of the research agendas as with IRCAI's SDG Observatory, particularly in what regards machine learning workflows and approaches complementing traditional and crowdsourced heterogeneous data collection and analysis.

KEYWORDS

Artificial Intelligence, Observatory, Survey Data Analysis, Complex Data Visualization, Multidisciplinary Collaboration.

1 Introduction

AI is increasingly shaping the economic landscape and societal dynamics across Brazil and South America, positioning the region as a growing hub for technological innovation. Despite challenges such as uneven infrastructure and regulatory hurdles, Brazil is making significant strides in AI research and development, contributing to the regulation and better understanding of the impact of AI in South America. OBIA [5] is answering this need, serving as a platform to support the strategy and other government actions with data on the uses and impacts of AI (see Figure 1).

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
Information Society 2024, 7–11 October 2024, Ljubljana, Slovenia
© 2024 Copyright held by the owner/author(s).
<https://doi.org/10.70314/is.2024.sikdd.18>



Figure 1: Screenshot of OBIA showing some results on the preparedness of Brazilian industry to adopt AI workflows.

The objectives of OBIA include compiling, recording, and providing information related to Artificial Intelligence in Brazil, enabling analyses of its adoption and its main impacts on society. It also has the mission of consolidating and disseminating knowledge about the repercussions of this technology, providing support to guide policies, strategies, and actions in promoting development and responsible use of AI. The observatory gathers Brazilian data on the use and adoption of Artificial Intelligence by different sectors, such as education, business, government, health, and others (see Figure 2). The currently available indicators rely mostly on traditional data sources for analysis, such as surveys and data sets made available for the team. The first product of OBIA is the book "Artificial Intelligence in Healthcare - Potentialities, Risks and Perspectives", published in July 2024. In a second line of action, it functions as a repository of guiding documents in the area, originating from all parts of the world. In a third line, it acts as an "information exchange point" between AI centers operating in Brazil: the IAX. All indicators collected will be public and can be accessed on the OBIA portal [4].

The Center for Artificial Intelligence (C4AI) at the University of São Paulo, funded by FAPESP (the public agency for research funding in the State of São Paulo) and IBM, participates in the OBIA through its Humanities area. C4AI will contribute with qualitative research in the horizontal axes of "Legislation, Regulation, and Ethical Use" and "AI Governance," while also conducting studies across various vertical axes to be monitored. The research group dedicated to this effort comprises scholars from the fields of law, computer science, electrical engineering, sociology, and political science, allowing for an interdisciplinary analysis of the key topics monitored by

OBIA. This interdisciplinary approach will provide a comprehensive view of the current state of AI development and implementation in Brazil. Various reports, articles, and data will be provided to support OBIA in fulfilling its mission.

In addition to the participation of professionals from various NIC.br departments, the Observatory has a network of external partners, including the Center for Management and Strategic Studies (CGEE), the São Paulo State System Data Analysis Foundation (SEADE), C4AI, CIAAM (Center of Artificial Intelligence and Machine Learning) and others. The following will explore how C4AI contributes to OBIA through a complementary approach, focusing on the qualitative analysis of decisions by the São Paulo Court of Justice related to AI.

2. Data and Methodology

2.1. Legislation, Regulation, and Ethical Use: A Qualitative Analysis

The research presented in this paper is the base of an action contributing to implement the PBI strategy [7], responsible for monitoring AI regulation and legislation. It has divided its research into three main areas: the Executive, the Judiciary, and the Legislative branch, combining traditional and modern data collection methods. Regarding the Executive branch, monitoring is being conducted through data scraping of government transparency websites based on a curated and continuously updated list of AI-related terms developed by the group. This monitoring aims to understand what AI systems are being purchased or contracted by public authorities. For the Judiciary, we have been analyzing court decisions from the São Paulo Court of Appeal (TJSP) related to AI, to understand judicial interpretations and rulings in the absence of specific AI legislation [2]. As of the latest data scraping in August 2024, more than 13.000 relevant decisions have been identified. Lastly, in relation to the Legislative Branch, the group is closely following the progress of discussions on Bill 2338/2023, which focuses on AI regulation, by participating in public hearings and issuing technical notes to guide legislators. The goal is to expand this research to monitor AI-related legislation at the state and municipal levels, as many municipalities are legislating on the matter to prepare their cities to assume roles of "smart cities".

2.2. Monitoring and exploring the local data

To effectively monitor developments in AI, it is essential to establish a comprehensive list of AI-related terms that can guide data collection efforts. This list is derived from multiple sources, including scientific articles, standards like [3], and reports such as OECD's [1]. The monitoring process involves monthly web scraping of court rulings, based on the AI-related terms list, from TJSP (Judiciary Power) and data from the Brazilian Transparency Portal (Executive Power), which occurs on the 15th of each month. For the Judiciary Power, the scrapes and data treatment are performed with scripts developed in

Python and R programming languages, based on the TJSP API, by Jesus Filho (github.com/jjesusfilho/tjsp). For the Executive Power, a script was developed to scrape data from the Data Download section of the Brazilian Transparency Portal (portaldatransparencia.gov.br/download-de-dados).

Currently, we are developing an automation tool, based on NLP techniques, to enhance the qualitative analysis of these court rulings, allowing for more efficient identification and categorization of data relevant to AI research. The first approach for this automation tool is using a NER (Named Entity Recognition) model, to automate the identification of relevant entities, including litigants and court judgments. The next step would be to apply a classification model, yet to be chosen, to filter out noise data. The process of constructing the terms for web scraping is a critical step to ensure the relevance and accuracy of the data collected for AI research. This process begins with the development of a comprehensive list of AI-related terms, which is built using multiple authoritative sources. One primary source is the OECD's report "Identifying and Measuring Developments in Artificial Intelligence," which offers a foundation of 226 AI-related terms identified through extensive analysis of scientific articles, open-source systems, and patents. Another source is the ISO/IEC 22989:2022 standard [3], which provides a framework for AI concepts and terminologies. These terms are carefully selected, refined, and translated into Portuguese by experts working within the Brazilian Technical Standards Association (ABNT) to ensure that only those terms that are highly relevant and specific to AI are included. Terms that are too general or contextually irrelevant—such as "transparency," which could result in unrelated hits concerning Brazil's Access to Information Law—are excluded to avoid false positives in the scraping process. The final list of terms, consisting of 103 terms in both English and Portuguese, is used to guide the web scraping data collection processes, allowing a focused and efficient retrieval of information that aligns with specific research objectives.

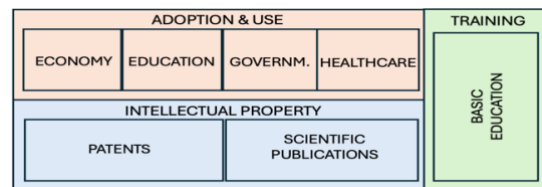


Figure 2: Current Dimensions of OBIA's monitoring topics

2.3. How to implement and classify repositories with reference documents and statistics?

As part of the data collection and structuring process for qualitative analysis, we are implementing and classifying repositories containing reference documents and statistics. These repositories will focus on key thematic areas, such as "Legislation, Regulation, and Ethical Use" and "AI Governance," and will be populated with data from sources like TJSP, the

Transparency Portal, and other relevant databases. By combining different methods, data retrieval becomes more efficient and targeted, ensuring the collection of relevant information. Web scraping supplements this process by capturing data unavailable through APIs, ensuring comprehensive coverage. The data is regularly updated, with documents classified by relevance to AI terms, creating a dynamic and organized repository (see Fig 3) described in [6].

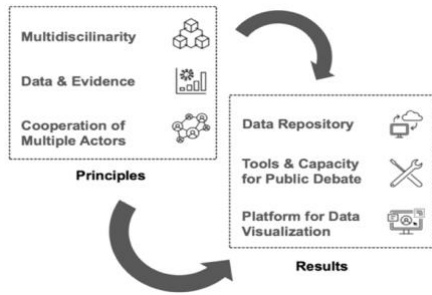


Figure 3: OBIA's guiding principles and expected results [6]

2.4. How to establish and maintain cooperation networks?

Establishing and maintaining cooperation networks requires fostering collaboration among interdisciplinary researchers from fields such as law, computer science, engineering, sociology, and political science. These networks are essential for sharing insights and methodologies related to AI monitoring. Using APIs and web scraping tools enables access to current data, supporting continuous knowledge exchange. Regular workshops, webinars, and joint research projects help keep participants engaged. Publishing reports, articles, and datasets strengthens the network and supports OBIA's mission to monitor AI developments comprehensively.

3 Discussion of initial results

As of June 28, 2024, a total of 13,064 decisions were scraped from the São Paulo State Court of Justice based on AI-related terms. Out of 103 terms searched, 45 returned at least one result. Graph 1 shows the monthly distribution of all results, while Figure 5 (logarithmic scale) displays the distribution of results by AI term. Both Portuguese and English terms were used for scraping. The top 15 terms with the most occurrences were analyzed over time, and Figure 6 presents the temporal evolution of these results by publication date. A qualitative review of 597 decisions from the São Paulo Court of Justice (TJSP) using a detailed list of AI-related terms, focused on terms like "Facial Recognition" and "Facial Biometrics," showing they are often used in various legal contexts, sometimes diverging from their technological meanings. Terms like "Facial Expression Recognition" and "Learning Agent" were often interpreted in psychological or social contexts rather than purely technological ones. The analysis used analytical, comparative, and monographic methods, with

the latter focusing on case-specific factors to draw broader generalizations. From TJSP's website, 597 rulings were reviewed: "Facial Recognition" (1), "Facial Expression Recognition" (1), "Machine Learning" (7), "Artificial Intelligence" (163), "Artificial Intelligence" in English (4), "Machine Learning" in Portuguese (3), "Learning Agent" (1), and "Facial Biometrics" (417).

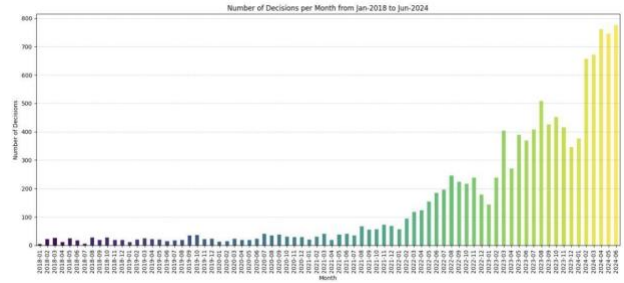


Figure 4: Nr. of Decisions per Month from Jan 2018 to Jun 2024.

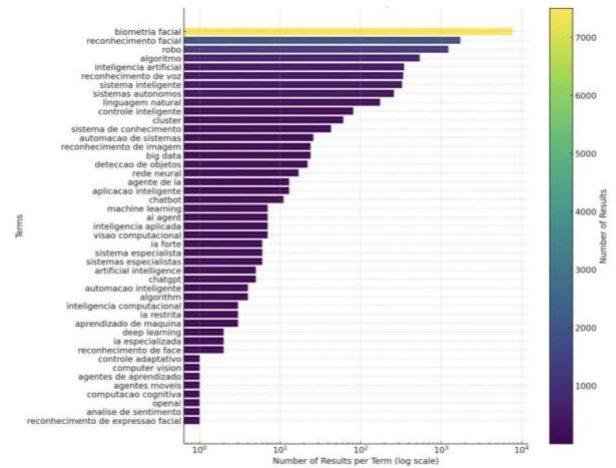


Figure 5: Number of results per AI term.

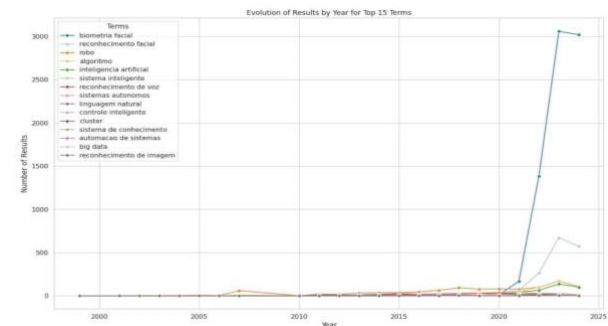


Figure 6: Evolution of results by year for top 15 terms.

The rulings followed a structured format, and the analysis included 14 categories, such as case number, appeal type, judge, district, and the context of term usage. Key findings highlighted the use of "Artificial Intelligence" and "Machine Learning" in commercial disputes and credit issues rather than solely technological matters. The rulings analyzed represent

decisions, rendered by collegiate bodies composed of multiple magistrates. Each ruling follows a structured format: Description and Qualification, covering aspects such as appeal, case number, judicial district, presiding judge, and parties involved; Summary of the ruling; Report, offering a brief description of the facts; Majority Opinion; and Dissenting Opinion (if applicable). The analysis was conducted with each of the 14 subcategories corresponding to columns in a single row: case number; type of appeal; reporting judge; district; judicial body; subject matter; judgment date; publication date; summary; parties; reasoning; final decision; context of term usage in the full text; and relevant jurisprudence. While the first nine categories were predefined based on the complete jurisprudence search, the remaining five were more subjective, created to enhance the understanding of the rulings' content and improve data visualization. Significant findings were noted in cases involving "Artificial Intelligence" and "Machine Learning," where the terms were often associated with commercial disputes, service contracts, or credit-related issues rather than purely technological applications. A recurrent theme in cases involving "Facial Biometrics" was the legality and validity of loan contracts signed through biometric recognition. The majority of decisions upheld the legality of such contracts, highlighting issues of consent and the technical reliability of biometric systems [1]. However, inconsistencies in judicial reasoning were identified, where similar cases had varying outcomes depending on the presiding judge. Overall, the analysis highlighted several gaps and challenges in the legal treatment of AI-related technologies, particularly concerning transparency, fairness, and consumer protection. The study underlined the need for more consistent legal standards and better understanding among judges of the technical nuances involved in AI applications to ensure fair and equitable rulings.

4 Conclusions and further work

The qualitative research findings from the analysis of court decisions related to AI reveal several key conclusions. AI-related terms such as "Facial Recognition," "Voice Recognition," and "Autonomous Systems" are frequently used in judicial contexts that extend beyond their traditional technological meanings, intersecting with areas like consumer protection, contract law, and fraud. The inconsistency in judicial reasoning and varying outcomes in similar cases highlight the need for clearer legal frameworks and a deeper understanding of AI's technological implications among judges. Moving forward, the incorporation of NLP techniques into the analysis will help extract key arguments from judicial decisions, providing deeper insights into the legal discourse on AI. This will enhance the robustness of future research on AI regulation and its implications for public policy.

Furthermore, a preliminary analysis of news using the NLP capabilities of the *Eventregistry.org* system (see Figure 7) show how this source can provide complementary results to the

study when, e.g., capturing the attention of media on the terms "criminal law" and "AI" in "Brazil" in the past 12 months, where 1.4% exhibits discussions on Human Rights, and terms like "democracy" and "discrimination" are within the top 30. When performing sentiment analysis over these results we can see large variations after the summer of 2022 with a predominantly negative sentiment regarding this search topic.

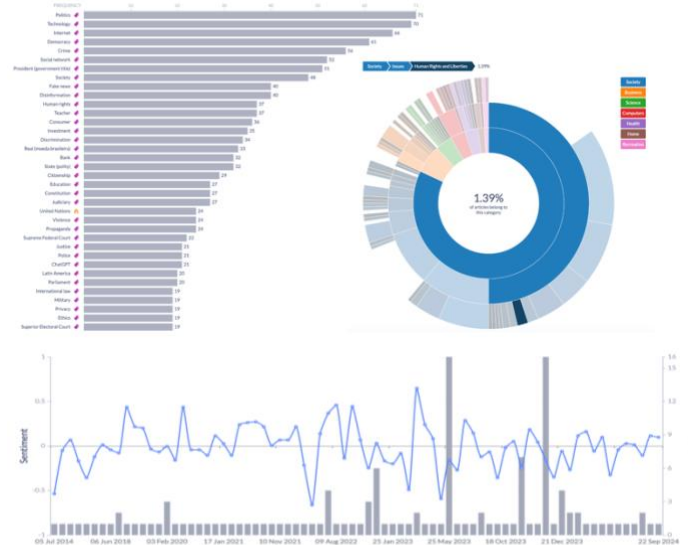


Figure 7: Significance of criminal law and AI in the news.

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to the Center for Artificial Intelligence (C4AI) at the University of São Paulo (USP), supported by FAPESP and IBM, for their invaluable support to the AI Observatory team. We thank to the CIAAM for their continued collaboration and contributions to this research. We thank the support of the European Commission project ELIAS - Lighthouse of AI for Sustainability (10080425).

REFERENCES

- [1] Baruffaldi, Stefano, et al. (2020) Identifying and measuring developments in artificial intelligence: Making the impossible possible. OECD.
- [2] Cristina Godoy B. de Oliveira, Otávio de Paula Albuquerque, Emily Liene Belotti, Isabella Ferreira Lopes, Rodrigo Brandão de A. Silva, Glauco Arbix. Intelligent Systems: 12th Brazilian Conference, BRACIS 2023, Belo Horizonte, Brazil, September 25–29, 2023, Proceedings, Part I, pp 18 – 32.
- [3] ISO (2022) Information technology — Artificial intelligence — Artificial intelligence concepts and terminology. ISO/IEC 22989:2022. [Online]. Available: <https://www.iso.org/standard/74296.html/> [27 8 2024]
- [4] Luiz Costa et al. (2024) The Brazilian Artificial Intelligence Observatory (OBIA). [Online]. Available: <https://www.obia.nic.br/> [27 8 2024]
- [5] MCTI (2021). Brazilian Strategy of Artificial Intelligence. [Online]. Available: [www.gov.br](https://www.gov.br/ebia-documento_referencia_4-979_2021.pdf) [07 9 2024]
- [6] MCTI (2023). OBIA: Observatório Brasileiro de Inteligência Artificial,. Available: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/1_ebia-reuniao-ro_7_24_05_2023_anexo_2_eixo2-pdf.pdf [27 8 2024]
- [7] PBI (2024). Brazilian Artificial Intelligence Plan . [Online]. Available: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/noticias/2024/07/plano-brasilero-de-ia-tera-supercomputador-e-investimento-de-r-23-bilhoes-em-quatro-anos/ia_para_o_bem_de_todos.pdf/view [07 09 2024]