2ND EDITION

ARTIFICIAL INTELLIGENCE

TECHNOLOGY APPLIED TO CONFLICT MANAGEMENT WITHIN THE BRAZILIAN JUDICIARY

COORDINATOR LUIS FELIPE SALOMÃO

FGV CONHECIMENTO
CENTRO DE INOVAÇÃO, ADMINISTRAÇÃO E PESQUISA DO JUDICIÁRIO

ISBN ~~000-00-00000-00-0~~

SUMMARY

ACKNOWLEDGEMENT

Fundação Getulio Vargas' Center of Innovation, Administration and Judiciary Research would like to thank the National Council of Justice and the courts that were the subject of the present research and that kindly made themselves available to provide the information contained in this report.

PREFACE

PREFACE

The Brazilian Judiciary has made a relevant commitment regarding the achievement of the Sustainable Development Goals (SDGs) of the United Nations 2030 Agenda, being recognized as the first Judiciary in the world to officially integrate the SDGs into its daily procedures.

The implementation of SDG 16, which refers to the promotion of peaceful and inclusive societies with universal access to justice as well as effective, responsible, and inclusive institutions at all levels, highlights the importance of incorporating technology in judicial routines, with the expectation of providing better and faster jurisdictional provision to users of the justice system.

In this context, in 2019 Fundação Getulio Vargas' Center of Innovation, Administration and Judiciary Research started an unprecedented survey named "Technology Applied to Conflicts Management in the Judiciary, with Emphasis in the Use of Artificial Intelligence", which sought to verify the state of the art of AI in Brazilian courts in order to identify the initiatives and delimit the results obtained from the following elements: current situation, problems that it seeks to solve, expected results and obtained results.

The first phase of the research mapped that 47 courts employed some form of AI in their activities, in addition to the Sinapse Platform from the National Council of Justice. The use of this technological tool is significantly important in the Brazilian context, especially considering the need to rationalize resources, and excessive judicialization, reflected in the more than 75 million cases in progress, according to the Justice in Numbers Report 2021.

In 2021, the survey was replicated in Brazilian courts, given the data dynamics that must be constantly updated. This 2nd edition expanded the survey of the technologies used, which now includes computational intelligence / artificial intelligence and analytics/business intelligence initiatives. Mapped information included: staff, technical aspects, data bases, evaluation and monitoring methodologies. The sample of the surveyed courts remained the same as it was in the 1st edition, and the methodology included a more comprehensive form, with a larger number of variables for investigation.

This study is highly relevant for both the judicial institutions and the public sector, since the innovation in this field has a great impact, from improving the services provided, to issues such as accessibility, information security, and privacy, for example.

The Center of Innovation, Administration, and Judiciary Research, through its inter-institutional research network, expects this report to promote multidisciplinary knowledge on the use of technology by Brazilian justice.

Minister Luis Felipe Salomão

Coordinator of the Center of Innovation, Administration and Judiciary Research

INTRODUCTION

01

INTRODUCTION

1.1. Mission of the Center of Innovation, Administration and Judiciary Research

The Center of Innovation, Administration, and Judiciary Research's mission is to contribute to improving the justice system, promoting the development of research, studies, discussion forums, events, and academic activities.

1.2. Interinstitutional Research Network

The present research was developed by an Interinstitutional Group of Researchers:

André Carlos Ponce de Leon Ferreira de Carvalho
Full Professor at the Institute of Mathematics and Computer Sciences and Research Productivity Scholar at CNPq. He was Visiting Researcher at the Alan Turing Institute, UK. He is a member of the Governance Committee in the Brazilian Strategy in Artificial Intelligence, member of the Strategy and Partnerships Board of the UKRI Center for Doctoral Training in Accountable, Responsible and Transparent AI, United Kingdom and member of the Directive Committee for Latin America and the Caribbean of the International Network for Government Science Advice. External Collaborating Researcher at the Center for Innovation, Administration and Judiciary Research at Fundação Getulio Vargas.

André Nascimento
Graduated in Computer Engineering from the University of Pernambuco, Master and Doctor in Computer Science from the Federal University of Pernambuco. Adjunct Professor at the Department of Computing at the Federal Rural University of Pernambuco and one of the coordinators of the Artificial Intelligence Laboratory (https://aiboxlab.org/).

Caroline Tauk
Federal Judge at the Federal Regional Court of the 2nd Region. Master in Public Law from the State University of Rio de Janeiro. Visiting Scholar at Columbia Law School. Academic Coordinator at the Center for Innovation, Administration and Judiciary Research at Fundação Getulio Vargas.

Fernanda Bragança

Doctor in Law from the Fluminense Federal University. Visiting researcher at the Université Paris 1 Pantheón Sorbonne. Researcher at the Center for Innovation, Administration and Judiciary Research at Fundação Getulio Vargas. Attorney. Mediator at FGV Mediation.

José Leovigildo Coelho

Coordinator of Technology Projects at FGV Knowledge. Researcher at the Center for Innovation, Administration and Research of the Judiciary. Partner at NESS Tecnologia, an IT Outsourcing company. Partner and Founder of Ionic Health, a technology company applied to diagnostic medicine. Partner at IARA Health, AI company in voice recognition for healthcare. Doctoral student in Technology Applied to Diagnostic Medicine at the Federal University of São Paulo. Executive Certification in Leadership & Management by MIT. Specialization in Exponential Technologies from Singularity University. Degree in Computer Science from the Federal University of Pernambuco. External Collaborating Researcher at the Center for Innovation, Administration and Judiciary Research at Fundação Getulio Vargas.

Juliana Loss

Academic Coordinator at the Center for Innovation, Administration and Research of the Judiciary. Technical Coordinator of Mediation at Fundação Getulio Vargas. Doctor in Law from Université Paris 1 Pantheón Sorbonne and Universidad Carlos III de Madrid.

Péricles Miranda

Adjunct Professor at the Federal Rural University of Pernambuco. Doctor in Computer Science from the Federal University of Pernambuco, with emphasis on Artificial Intelligence. Expertise in topics related to optimization, machine learning, meta-heuristic, and hyper-heuristic algorithms. Master in Computer Science from the Federal University of Pernambuco (2012) and Bachelor in Computer Engineering from the University of Pernambuco (2010).

Rafael Ferreira Mello

Professor at the Federal Rural University of Pernambuco. Coordinator of the Artificial Intelligence Laboratory (https://aiboxlab.org/). Postdoctoral fellow at the University of Edinburgh's School of Informatics in 2018. Doctor in Computer Science from the Federal University of Pernambuco with research interests encompassing natural language processing, learning analytics and educational technology. Participation in several multinational research projects, involving academic partners and companies in Europe, Australia, United States and Latin America.

Renata Braga

Adjunct Professor of the Law Course at UFF – Volta Redonda. Coordinator of the Study and Research Group on Consensual Methods of Conflict Resolution. Post-Doctor at UFRJ/University of Coimbra. Doctor in Law from the Federal University of Santa Catarina. Master in Civil Law from the State University of Rio de Janeiro. External Researcher Collaborating at the Center for Innovation, Administration and Research of the Judiciary at FGV.

Trícia Navarro
Judge at the Court of Justice of the State of Espírito Santo, currently in the position of Assistant Judge of the Presidency of the National Council of Justice. Professor of the Undergraduate and Master's programs at the Federal University of Espírito Santo. Post-Doctorate in Law from USP. Doctor in Procedural Law from the State University of Rio de Janeiro. Master in Procedural Law from the Federal University of Espírito Santo.

1.3. Objectives

The general objective of this research is to survey the use of data science and analytics, as well as technologies based on artificial intelligence (AI) and machine learning in the Brazilian Judiciary. The study sought to identify ongoing projects and initiatives, their respective functionalities, the impact produced using AI, and to verify the impact of AI on the celerity, efficiency, and productivity of the courts.

1.4. Methodology

The present research is exploratory, descriptive, and analytical to identify and describe the initiatives and national experiences of technologies with the use of computational and artificial intelligence in the courts, aimed at improving the justice system from the perspective of its functioning and structure.

The research sample covered the National Council of Justice, the Federal Supreme Court, the Superior Court of Justice, the Superior Labor Court, the Regional Labor Courts, the Federal Regional Courts, and the Courts of Justice.

A standard form, with open and closed questions, was sent to all courts participating in the survey. The consolidation of this study sought to maintain the text prepared by the courts to faithfully represent the data sent.

The quantitative data analysis was carried out from two main scenarios: the total set of initiatives and a cut of systems with informed status "in production".

The model of the survey form sent to the courts is presented below:

1. NAME OF THE ARTIFICIAL INTELLIGENCE INITIATIVE:

2. COURT:

3. YEAR OF IMPLEMENTATION:

4. CURRENT STATUS OF THE INITIATIVE:

( ) in ideation

( ) in development

( ) executing the pilot project

( ) in implementation

( ) in production

TEAM INFORMATION

5. PROJECT OWNER:

To answer this question, the product owner is understood as the team member who oversees defining and prioritizing project requirements and expediting the execution of priorities while maintaining the conceptual and technical integrity of features or components for the team.

6. CONTACT INFORMATION:

E-mail:

Phone:

7. TECHNICAL TEAM:

( ) internal (government employee, service providers, etc)

( ) external (Universities, Foundations, Companies, etc)

( ) internal and external

8. If there are external members, please inform their origin (private initiative, university):

9. CURATORY AND TRAINING TEAM:

( ) internal (government employee, service providers, etc)

( ) external (Universities, Foundations, Companies, etc)

( ) internal and external

10. If there are external members, please inform their origin (private initiative, university):

TECHNICAL ASPECTS

11. WHICH PROBLEM(s) OR MANAGEMENT OBJECTIVE(s) OF THE COURT IS ADDRESSED BY THE SOLUTION?

[Please indicate more than one, if applicable.]

12. IS THE INITIATIVE VINCULATED TO ANY EXISTING SYSTEM IN THE COURT?

(ex.: PJe, saj, eproc, etc)

13. TYPE OF TECHNOLOGY:

( ) Computational / Artificial Intelligence

( ) Analytics / Business Intelligence

14. What models and computational intelligence techniques are applied in the computational intelligence initiative?

[ ] Natural Language Processing (LNP)

[ ] Machine Learning

[ ] Deep learning

15. WHICH LEARNING PROCESS IS EMPLOYED?

( ) Supervised Learning

( ) Unsupervised Learning

( ) Semi-Supervised Learning

( ) Reinforcement Learning

( ) Other: _____

16. WHICH PROBLEM(S) IS MOSTLY ADDRESSED IN THE SOLUTION?

(CHECK MORE THAN ONE IF APPLICABLE)

[ ] Visualizing / Exploring data

[ ] Analytics / Business Intelligence

[ ] Jurimetry

[ ] Summarization

[ ] Extracting data

[ ] Signal processing (audio, video, etc.)

[ ] Categorization / Classification of documents

[ ] Clustering

[ ] Topics modeling

[ ] Anomaly detection

[ ] Information retrieval (ex.: Search engines, indexing, etc)

[ ] Text generation (ex.: support to the semi-automated writing of minutes, decisions, etc)

[ ] Optimization (ex.: allocation of resources, processes, personnel, routes, etc)

[ ] Recommendation system

[ ] Other: _____

17. BRIEFLY DESCRIBE THE SOLUTION FUNCTIONING (Architecture, training process, etc.):

18. WHAT TYPE OF TEXTUAL REPRESENTATION IS USED?:

[ ] Term Frequency (TF, TF-IDF,etc)

[ ] Word Embeddings (word2vec, glove, etc)

[ ] Contextual embeddings (FLAIR, BERT, ULMFit, etc)

[ ] Other: _____

19. WHICH PROGRAMMING LANGUAGE IS USED (CHECK MORE THAN ONE IF APPLICABLE):

[ ] Python

[ ] R

[ ] Java

[ ] C/C++

[ ] Javascript

[ ] OTHER: _____

20. WHICH IS THE MAIN ARCHITECTURE / PLATFORM OF THE SOLUTION:

[ ] Web

[ ] Desktop

[ ] Microservices

[ ] Mobile devices

[ ] Cloud providers (AWS SageMaker, Azure ML, Google AI, etc)

21. TECHNOLOGIC TOOLS USED (Ex.: Tensorflow, Keras, Scikit-Learn, Weka, Gensim, etc)

22. WHICH ALGORITHM PROVIDED BEST RESULT / WAS IMPLEMENTED? (Ex.: Random Forest, SVM, CNN, LSTM, CRF, XGBoost etc.

23. DATA PROCESSING INFRASTRUCTURE

24. INDICATE THE MAIN PLATFORM FOR EXECUTING THE PIPELINES OF DATA PROCESSING AND TRANSFORMATION, MODEL TRAINING, ETC.

( ) Local device (ex.: Desktop, etc.)

( ) Distributed computation (ex.: Hadoop, Kubernetes, etc)

25. DATA STORAGE INFRASTRUCTURE

( ) Local network (e.g.: on-premises, etc.)

( ) Cloud computing (ex.: HDFS, AWS S3, Google Cloud Storage, etc)

BASE DE DADOS

26. WAS ANY DATABASE USED FOR MODEL TRAINING?

( ) Yes

( ) No

If the previous answer was YES, please answer questions 27-31

27. BRIEFLY DESCRIBE THE DATABASE USED:

28. QUANTITY OF INSTANCES (ex.: Documents, Images, etc) IN THE DATABASE

29. WHICH DATA ANNOTATION PROCESS IS USED?

( ) Pre-annotated base (history data)

( ) Annotation done during the project

30. WAS THE DATABASE ENOUGH FOR TRAINING THE PROJECT MODELS? WAS THERE ANY CHALLENGES RELATED TO THE DATABASE (ex.: noise, low quality images, etc)?

31. POINT THE TYPES OF DATA USED FOR BUILDING THE SOLUTION

[ ] Texts (natural language)

[ ] Logs

[ ] Tabular data (e.g., transactional systems, etc)

[ ] Images

[ ] Audio

[ ] Video

[ ] Other: _____

EVALUATION AND MONITORING

32. HOW WERE THE COMPUTATIONAL INTELLIGENCE MODELS EVALUATED?

( ) Accuracy

( ) AUC / AUPR

( ) Precision

( ) Revocation

( ) F1-Score

( ) Other metrics: _____

33. THE SYSTEM DECISIONS/RESULTS ARE VALIDATED BY A HUMAN? (ex.: the predictions are used as suggestions/recommendations, which are subjected to validation by a human being?)

( ) Yes

( ) No

34. FOR SYSTEMS IN PRODUCTION: HOW THE MODEL IS MONITORED? (ex.: concept drift detection, anomalies, etc.)?

35. THE OPERATIONAL COMPUTATIONAL INTELLIGENCE SYSTEMS HAVE ALREADY HAD ANY UPDATE OR IMPROVEMENT?

( ) Yes

( ) No

36. THE CALIBRATION / ADJUSTMENT IS MANUAL OR AUTOMATIC?

37. IS THE SYSTEM AUDITABLE (OR WAS AUDITED DURING ITS DEVELOPMENT) FROM THE POINT OF FAIRNESS, ACCOUNTABILITY, TRANSPARENCY (FAT)?

( ) Yes

( ) No

38. WHAT IS THE CURRENT STAGE OF THE RESULTS ALREADY OBTAINED FROM THE COMPUTATIONAL INTELLIGENCE MODELS APPLIED IN THESE INITIATIVES?

39. IS IT ALREADY POSSIBLE TO MAP THE RESULTS OF USING COMPUTATIONAL INTELLIGENCE IN THE COURT?

( ) Yes

( ) No

40. WERE THE SHORT, MEDIUM AND LONG TERM EXPECTATION MET?

( ) Yes

( ) No

( ) Not yet evaluated

41. A STUDY WAS DONE TO VALIDATE THE BENEFITS OR GAINS IN PRODUCTIVITY BY USING THE SOLUTION? (ex.: a person needed X hours and such time decreased to Y minutes by using the system, etc)

42. WHICH ARE THE MAIN NEEDS THAT COULD BE MET BY THE TECHNOLOGY?

43. WHAT ARE THE CURRENT LIMITATIONS OF THE EXPECTED RESULTS FROM THE COMPUTATIONAL INTELLIGENCE MODELS APPLIED IN THIS INITIATIVE?

44. DESCRIBE THE RISKS OF THE COMPUTATIONAL INTELLIGENCE MODELS APPLIED IN THIS INITIATIVE.

CONCLUSION

45. WHAT IS THE COURT'S INVESTMENT PLAN IN TECHNOLOGY FOR THE COMING YEARS?

46. THE COURT HAS A CLUSTER FOR DEVELOPING MACHINE LEARNING PROJECTS?

( ) No

( ) Yes (no GPU)

( ) Yes (with GPU)

( ) Other: ____

47. WHICH EXTERNAL DATABASE INTEGRATIONS AND EXTERNAL SYSTEMS WOULD BE IMPORTANT FOR THE IMPROVEMENT AND ENRICHMENT OF THIS COMPUTER INTELLIGENCE SYSTEM?

48. WHAT, IF ANY, WERE THE DISCOVERIES MADE AFTER PROCESSING THE DATA USING COMPUTER INTELLIGENCE TECHNIQUES FROM THESE INITIATIVES? INDICATE.

49. WHAT OTHER PROJECTS USING COMPUTATIONAL INTELLIGENCE COULD BE DEVELOPED TO IMPROVE JUSTICE IN THE COUNTRY?

C

CONSOLIDATIONS

02

DATA CONSOLIDATION AND ANALYSIS

This chapter aims to present the consolidated results of the survey carried out through the application of the research described in chapter 3, covering the STF, STJ, TST, TRTs, TRFs, and TJs. The questionnaire consisted of a set of open-ended and multiple-choice questions and covered several practical aspects of the use and development of AI initiatives in the context of each court.

2.1 Data standardization methodology

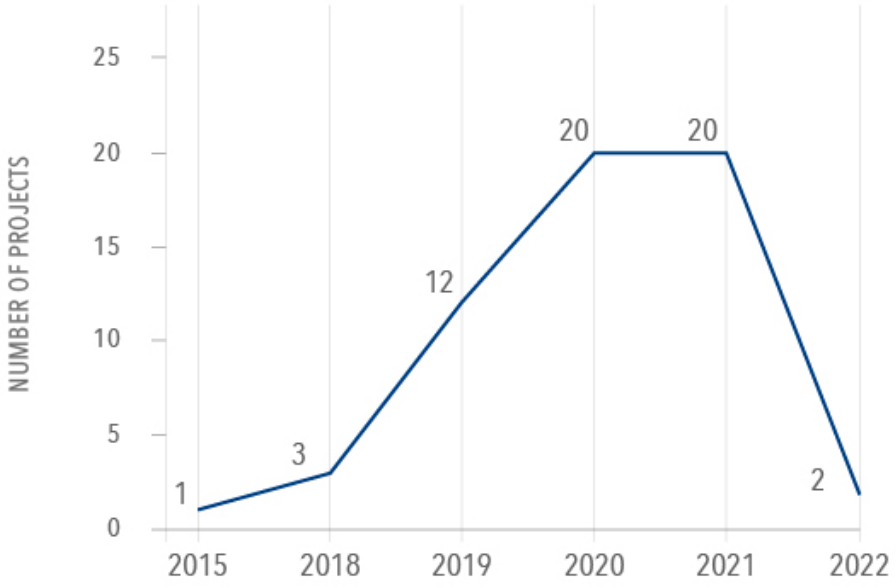
The quantitative data analysis was carried out considering two main scenarios: the total set of initiatives and a cohort of systems with the status informed "In Production". The fields presented here went through a normalization process according to the following criteria:

1. Capitalization standardization (all capital letters).
2. Replacement of names in full by the abbreviations of institutions/systems.
3. Separation of multivalued fields (separated by comma or semicolon) into distinct values.

2.2 Distribution of initiatives by region

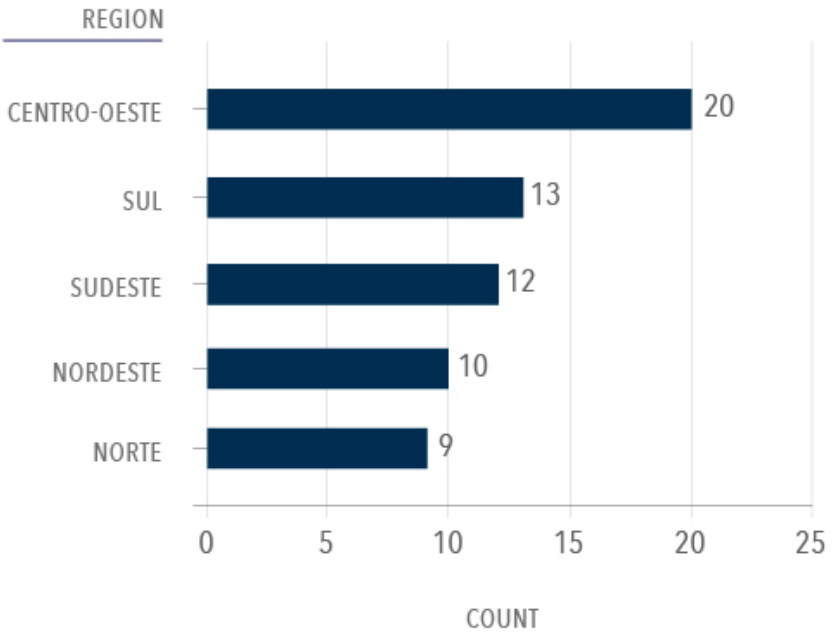
The survey pointed to a growth in the number of AI projects implemented in Brazilian courts. However, one can see the impact of the pandemic on the growth rate observed in previous years. It is worth mentioning that the graph already presents a forecast of the systems that will be implemented in 2022.

Figure 1: Number of implemented initiatives along the latest years



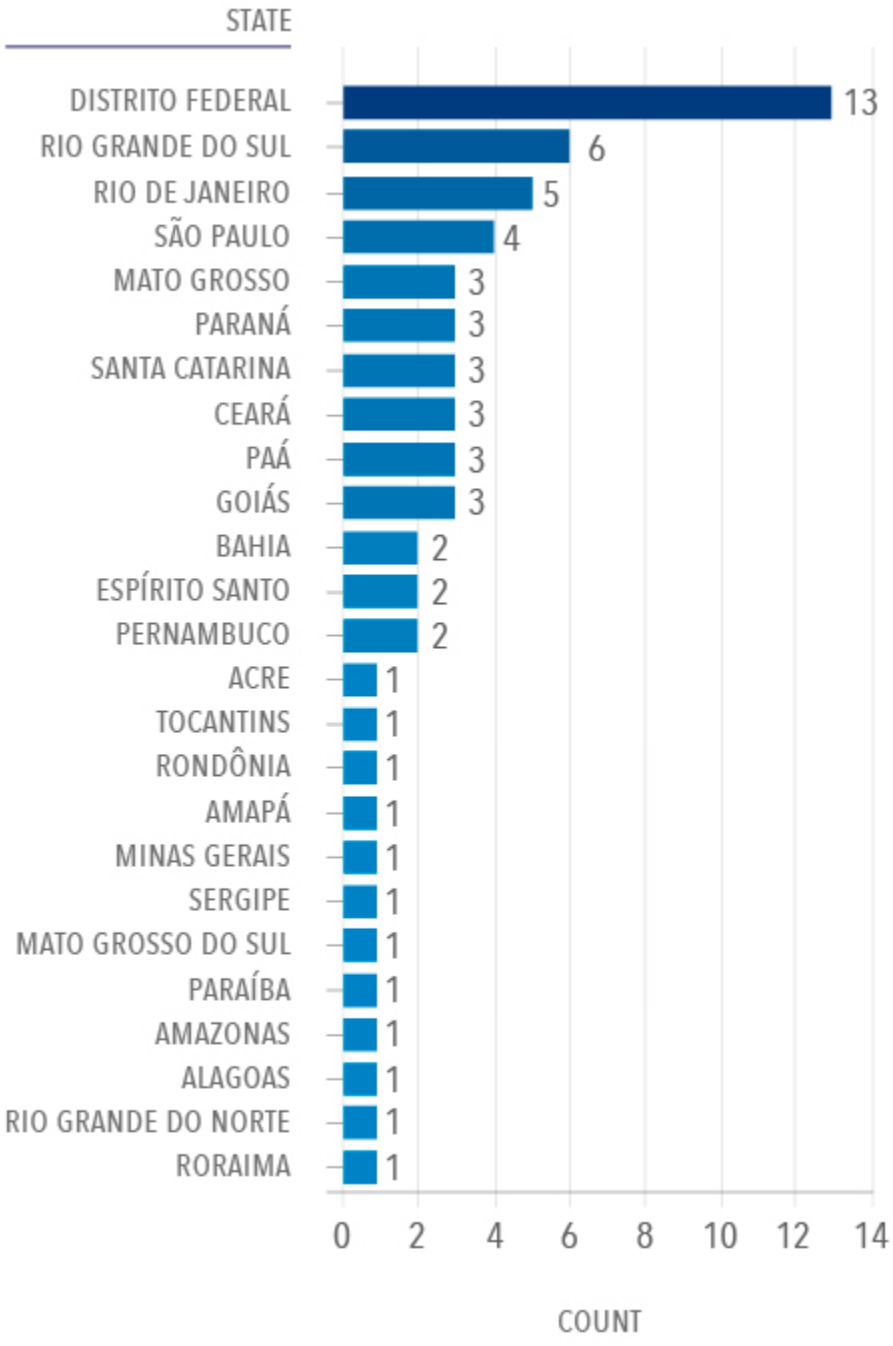
As shown in Figure 2, most of the initiatives are in the midwest region, mainly focused on courts located in Brasília. The other regions have a similar number of systems under implementation, with the variation from south to north being only four initiatives.

Figure 2: Number of initiatives by region



On the other hand, Figure 3 highlights the states with the highest number of initiatives to use artificial intelligence for the legal sector. The Federal District, Rio Grande do Sul, Rio de Janeiro and São Paulo states have more initiatives than the other states.

Figure 3: Number of initiatives by state



2.3 Current situation of the initiatives

Figure 4 shows the relationship between the number of initiatives per court and by the situation of development/adoption of the initiative. This figure shows that courts of law have had more experience in adopting artificial intelligence techniques. It is also possible to identify that the TST and STF have few initiatives, but all of them are already in production, which shows a higher level of maturity.

Figure 4: Quantity of initiatives by court and by situation

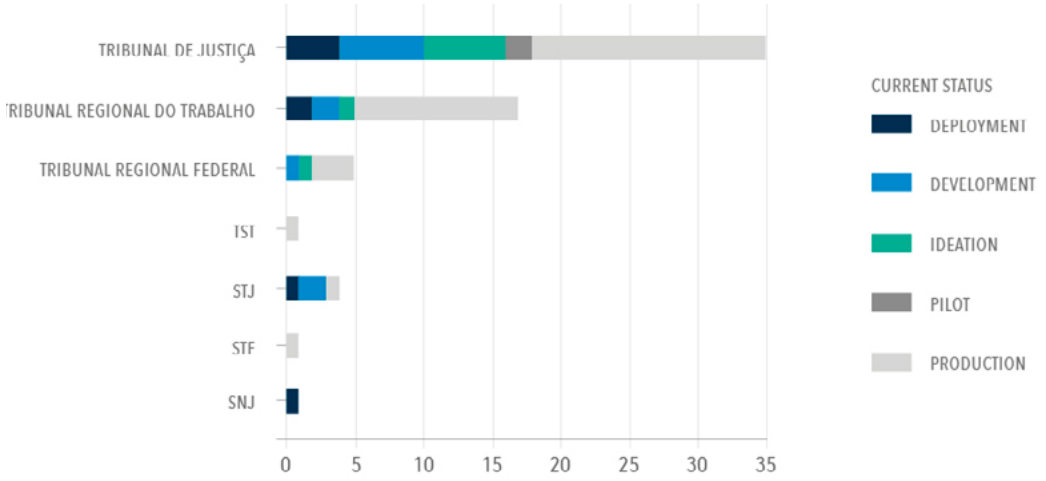
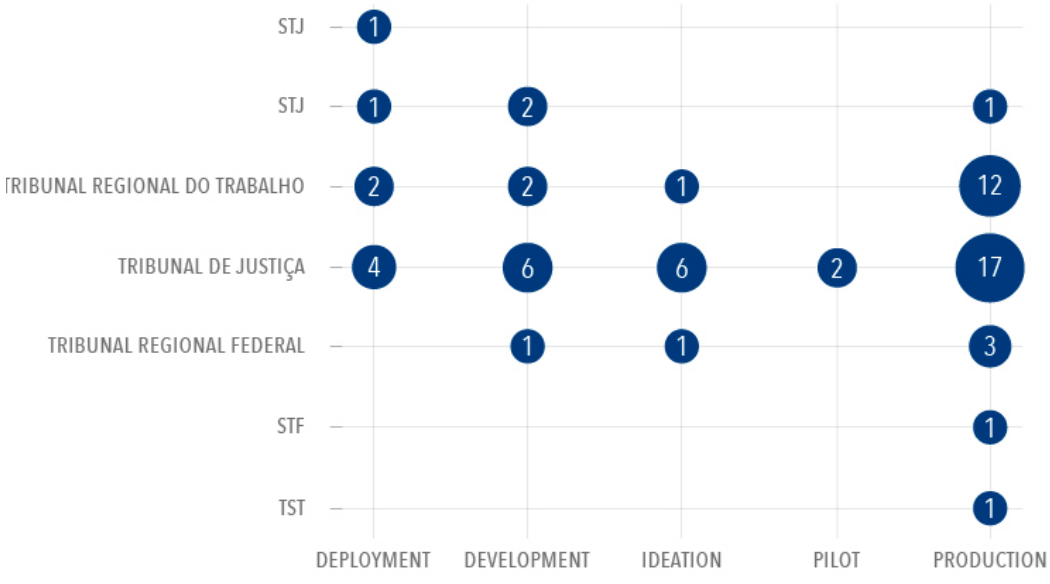


Figure 5 provides more details on the status of court initiatives. For example, it is possible to identify that the courts of justice have a good distribution of initiatives in the different stages of adoption and development. On the other hand, some courts already have solutions in production but few at the beginning of development.

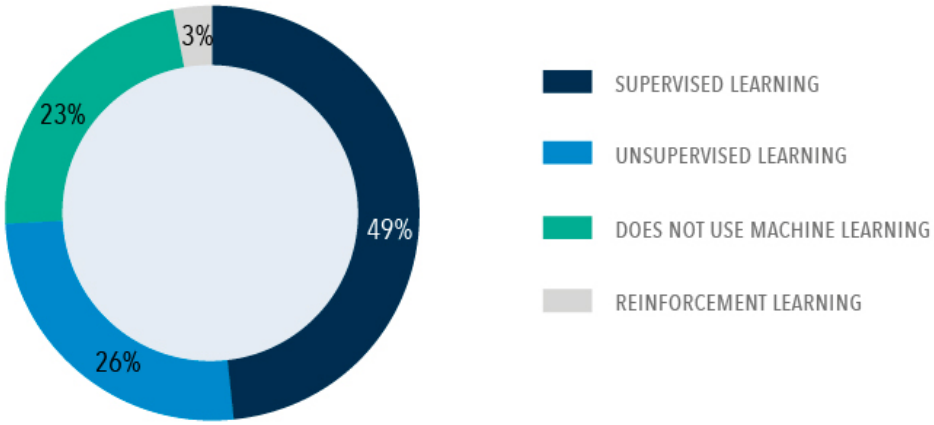
Figure 5: Situation of initiatives by court



2.4 Systems in production

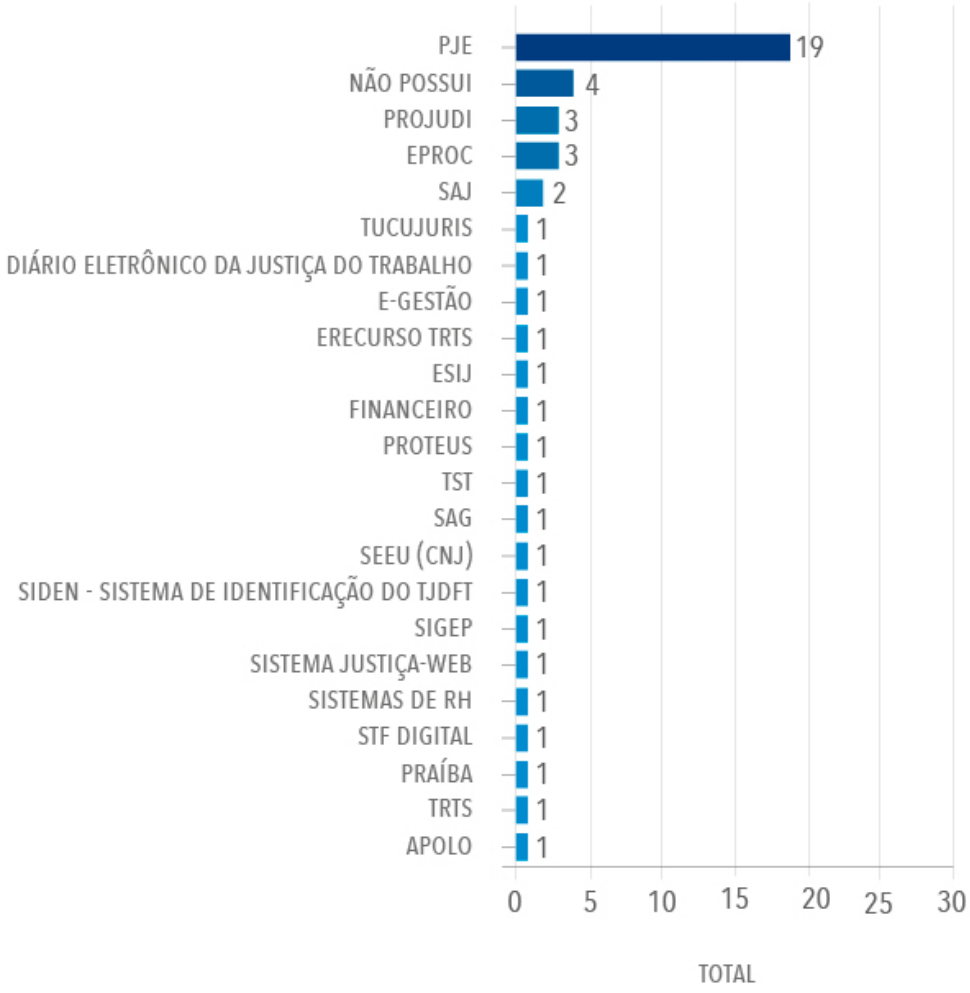
This section presents the results focused on the systems that are in production in the courts. Figure 6 shows that the approaches most employed in the initiatives are traditional supervised and unsupervised learning. However, it is worth mentioning that 23% of the initiatives do not use any data training approach, focusing more on automation-related techniques.

Figure 6: Approach employed



Another important piece of information that shows the maturity of the solutions in production is that only four do not have integration with other court systems, as shown in Figure 7. In addition, the PJE has 19 initiatives to use intelligent services in combination with existing systems.

Figure 7: Integration with existing systems



2.5 Team

Figures 8 and 9 show how the solutions development and data curation teams are divided into internal, external, and mixed (internal and external) teams. They indicate that the solutions development teams are mostly formed by people from the court itself.

Figure 8: Technical team

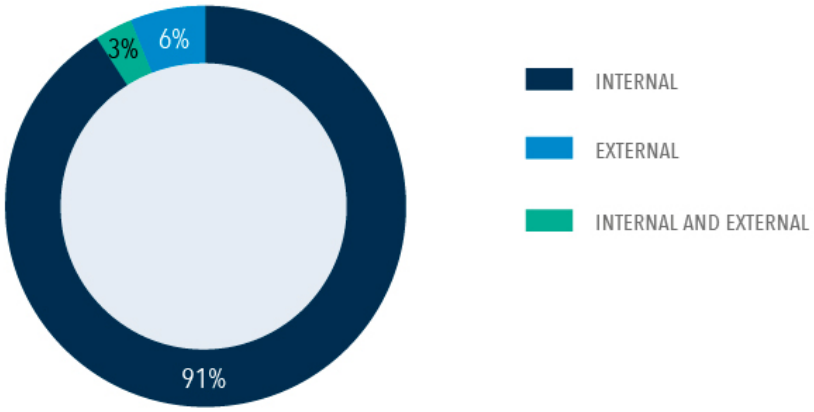
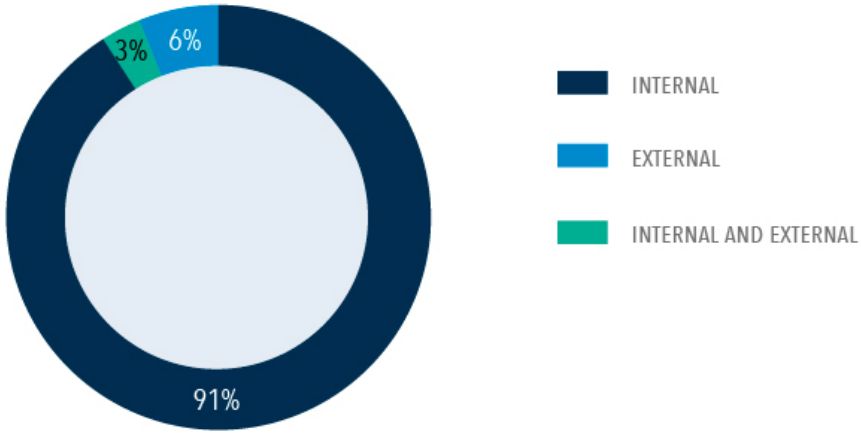


Figure 9: Training and curation teams



2.6 Results validation

In just over half of the initiatives (54%), the results were mapped after the project was implemented (Figure 10). The survey also showed that most projects currently in production met initial expectations (60%). However this leaves a high percentage of initiatives in which the result did not meet expectations (40%), as seen in Figure11.

Figura 10 – Status de projetos cujos resultados foram mapeados

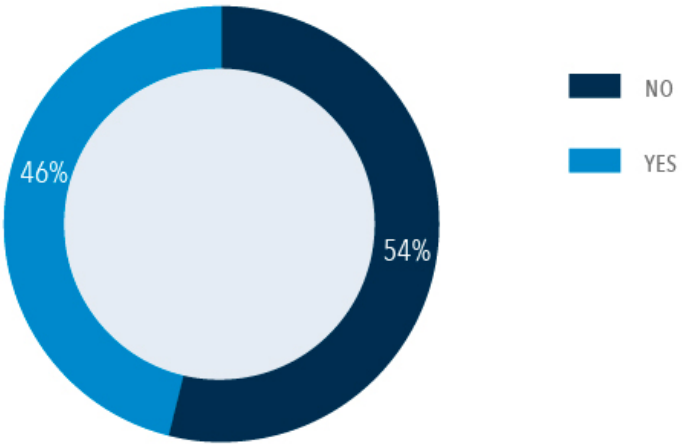
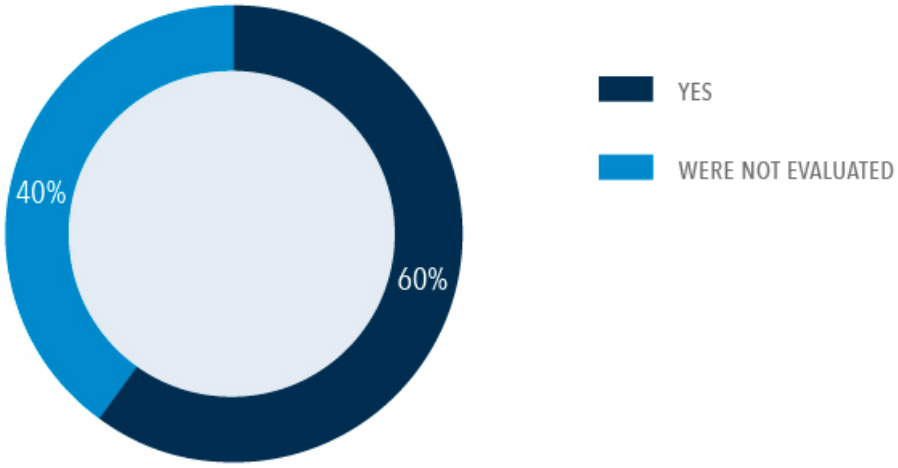
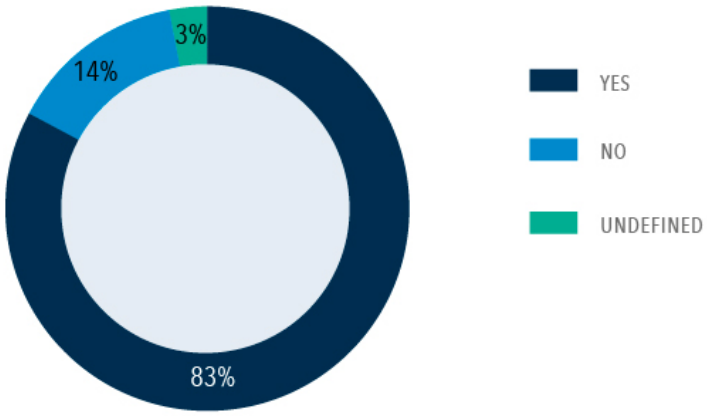


Figure 11. Distribution of projects in terms of meeting initial expectations



On the other hand, 83% of the projects indicated that the results obtained by the implemented solution undergo some type of human validation (Figure 12).

Figure 12. Validation of the obtained results.



2.7 Infrastructure

Regarding the infrastructure used for the development of the projects, most of the projects used the local infrastructure of the court itself to store the data used (or analyzed), as can be seen in Figure 13. In terms of processing, Figures 14 and 15, a prevalence of use of local infrastructure (computers, servers, etc.) was also observed, instead of using distributed or cloud computing platforms. However, only a small percentage (9%) of the respondents claimed to have a computing infrastructure equipped with graphic processing units (GPUs), which are extremely relevant for developing deep learning models.

Figure 13. Storage Infrastructure

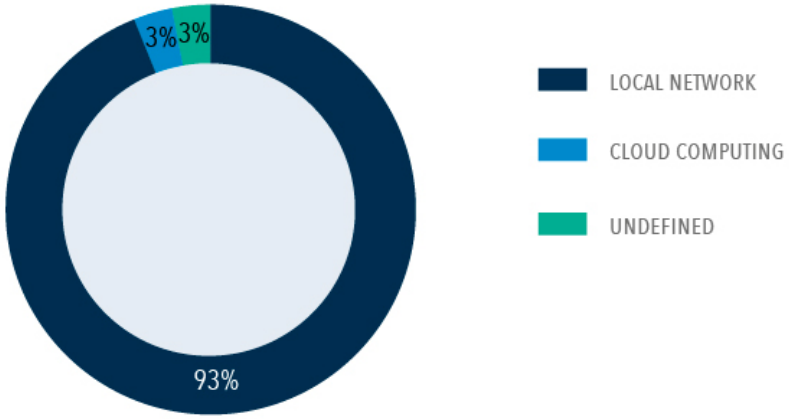


Figure 14. Processing infrastructure

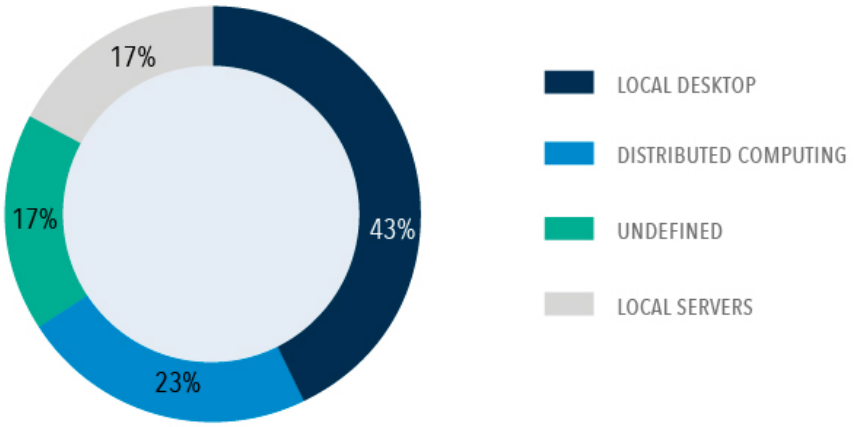
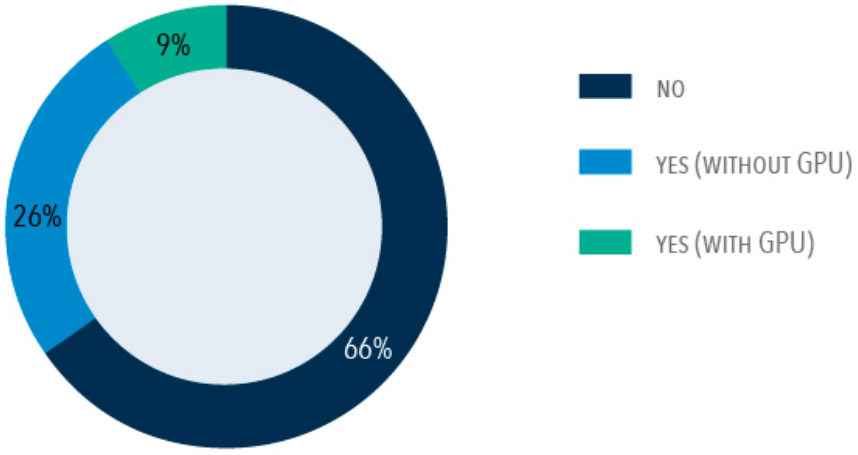


Figure 15. Availability of cluster for model training



2.8 Databases

Regarding the Databases used, 74% of the initiatives used this type of resource (Figure 16), with a prevalence of databases composed of text documents (pure text or PDFs), as can be seen in Figure 17. This justifies the great use of natural language processing techniques in the collected answers. The most used textual representation was the term (11), followed by dense vector representations (i.e., word embeddings), with six cases. However, we believe that the use of contextual embeddings is a trend that has already been observed in two of the mapped initiatives (Figure 18).

Figure 16. Use of databases

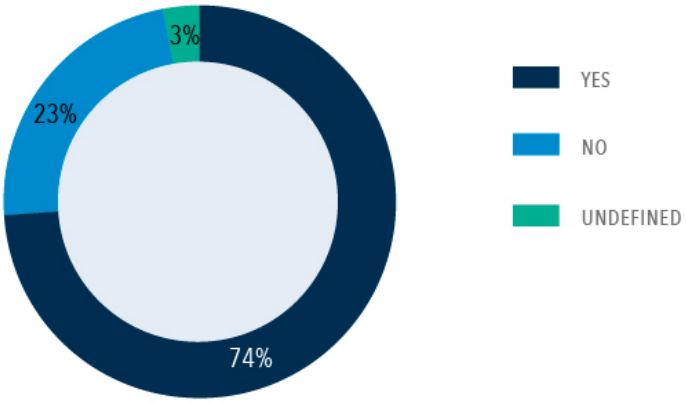


Figure 17. Availability of cluster for model training

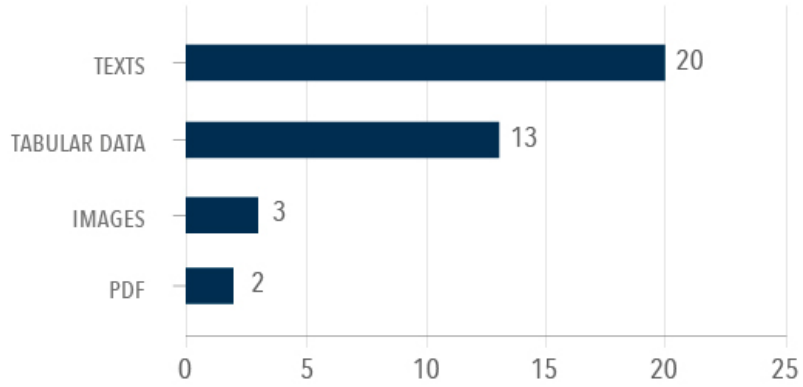
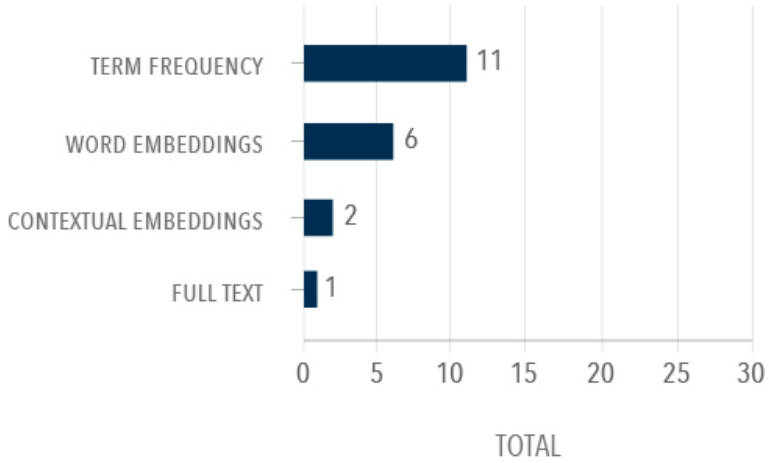


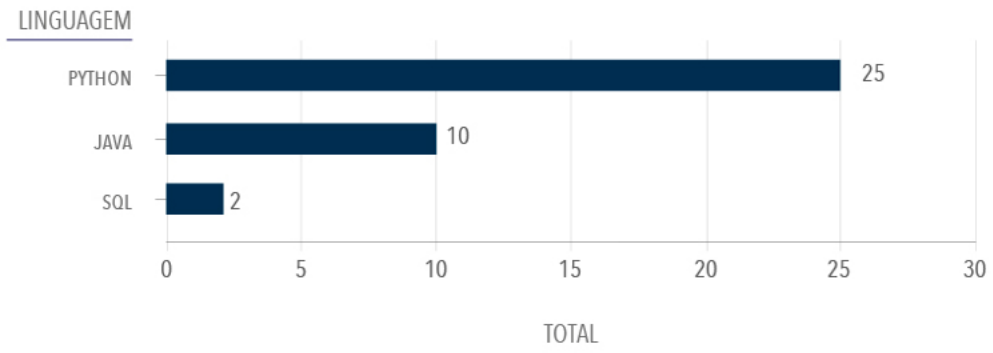
Figure 18. Textual representation



2.9 Methods and languages adopted

The most used language in the analyzed responses was Python, followed by Java and SQL (Figure 19). This was expected since this is one of the most popular languages worldwide for the most frequent tasks in the analyzed courts, also reflected in Figure 20.

Figure 19: Programming languages adopted in the solutions



Among the most used frameworks, scikit learn (Python) stands out in projects that involve document classification and grouping (Figure 21). Power BI and QlikSense stood out as BI platforms and elastic search and Solr in information retrieval projects.

Regarding the algorithms and models used in the implemented solution, tree-based algorithms (Random Forest and XGBoost) to classify documents stood out, while the k-Means algorithm was the most used in clustering problems (Figure 22).

Figure 20: Problems addressed x Programming languages

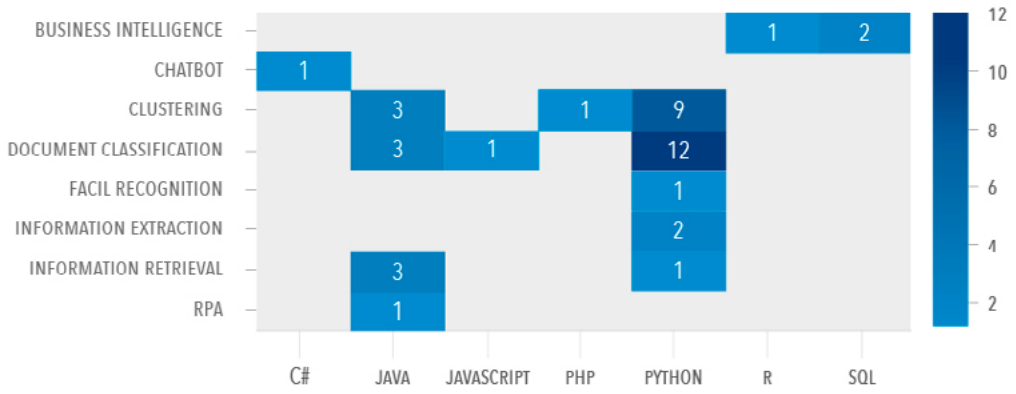


Figure 21: Frameworks x Problems addressed

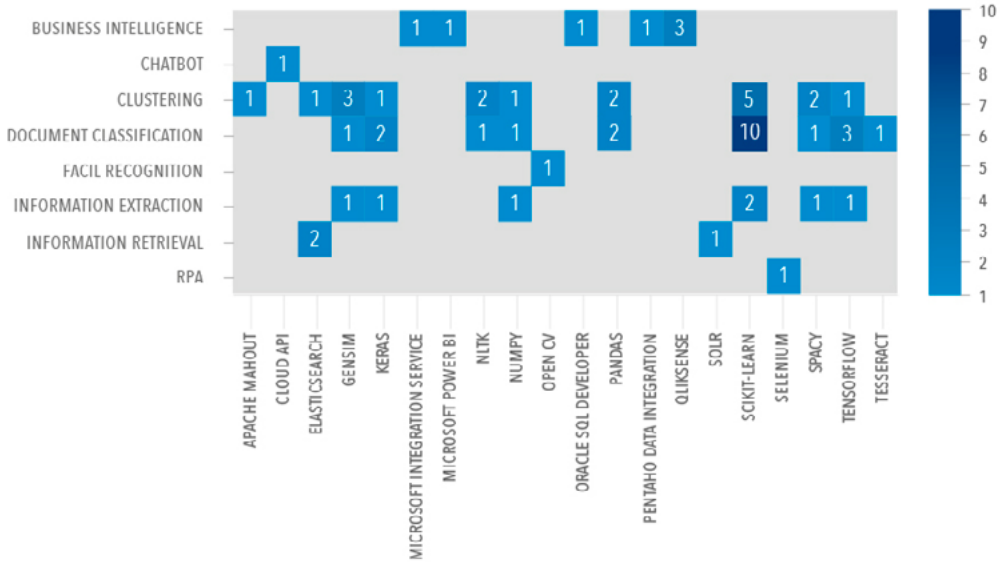
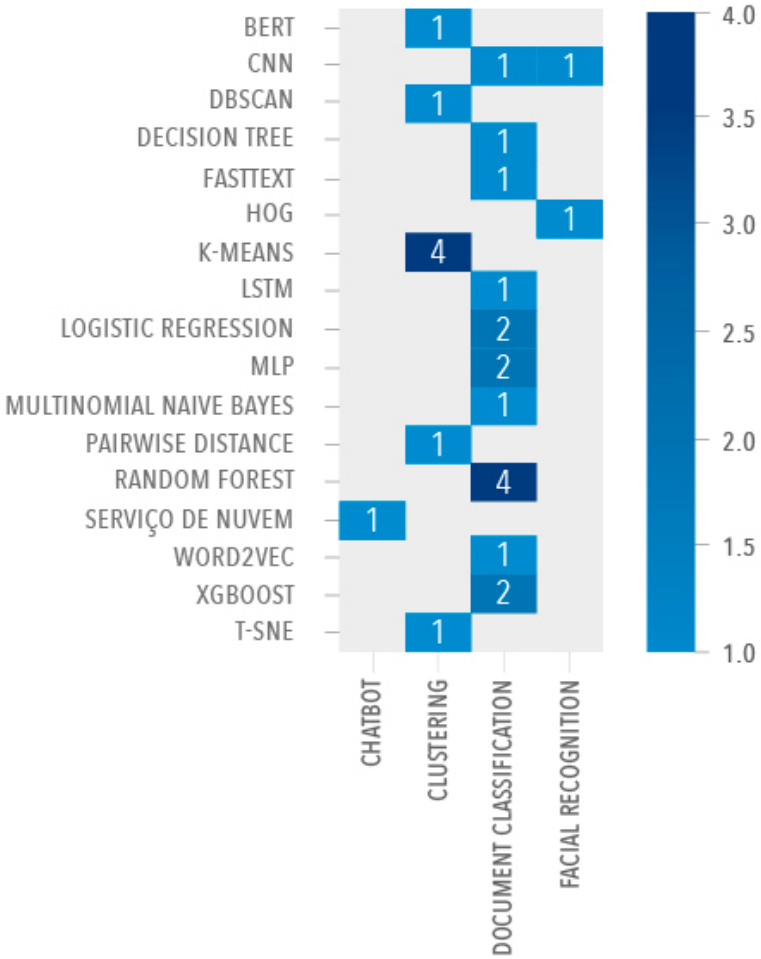


Figure 22: Methods x Problems addressed



FINAL CONSIDERATIONS

The universe of AI-based and data-intensive approaches and methods currently used in the Brazilian Judiciary is vast and diverse in terms of the technologies and problems tackled.

One can observe the prevalence of methodologies and algorithms widely used outside the legal context and especially common in text mining. This fact reinforces the need and characteristic of structuring and automating workflows involving analyzing and sorting a massive amount of text documents.

We believe that, as such initiatives gain more strength, the greater the need for investments in high-performance processing infrastructure within the Judiciary.

FGV CONHECIMENTO
CENTRO DE INOVAÇÃO,
ADMINISTRAÇÃO E PESQUISA
DO JUDICIÁRIO