

**COREJUR e UE CEIA/UFG**

**PROJETO DE P&D DE ENCOMENDA TECNOLÓGICA DE MÓDULO DE  
INSTRUÇÃO ASSISTIDA POR INTELIGÊNCIA ARTIFICIAL**

ARLINDO RODRIGUES GALVÃO FILHO

DUÍLIO CALAÇA JUNIOR

MATEUS SOARES JORGE

GOIÂNIA  
2022

ARLINDO RODRIGUES GALVÃO FILHO

DUÍLIO CALAÇA JUNIOR

MATEUS SOARES JORGE

**PROJETO DE P&D DE ENCOMENDA TECNOLÓGICA DE MÓDULO DE INSTRUÇÃO  
ASSISTIDA POR INTELIGÊNCIA ARTIFICIAL**

Projeto de pesquisa e desenvolvimento apresentado em resposta ao chamamento público para encomenda tecnológica de instrução assistida por inteligência artificial de processo nº TC 036.798/2020-7 para seleção em processo de contratação de ETEC

GOIÂNIA  
2022

# PROJETO DE P&D DE ENCOMENDA TECNOLÓGICA DE MÓDULO DE INSTRUÇÃO ASSISTIDA POR INTELIGÊNCIA ARTIFICIAL

## ÍNDICE

<b>APRESENTAÇÃO</b>	<b>5</b>
Objetivo	5
Proponentes	5
Introdução	7
<b>DA SOLUÇÃO</b>	<b>8</b>
Síntese	8
Detecção de objetos para extração de parágrafos das peças processuais	10
Rota tecnológica	11
Etapas para desenvolvimento da solução	14
Cronograma	16
Classificação dos parágrafos, reconhecimento de entidades nomeadas, painel de jurimetria e redação das instruções e comunicações	16
Classificação quanto à admissibilidade	17
Classificação quanto aos requisitos de concessão de medida cautelar	20
Classificação quanto aos itens do formulário de petição inicial	22
Classificação quanto às comunicações a serem enviadas aos interessados	26
Classificação quanto aos encaminhamentos	27
Painel de jurimetria e redação das instruções e comunicações	28
Rota tecnológica	33
Etapas de desenvolvimento do projeto	40
Cronograma	42
Redação automática de instruções e sumarização de peças processuais – Geração automática de texto	45
Solução e aderência ao edital	45
Rota tecnológica	47
Etapas de desenvolvimento do projeto	55
Cronograma	56

	4
<b>CICLOS TEMÁTICOS</b>	<b>58</b>
Cronograma do segundo ciclo temático	59
<b>DO ESTUDO TÉCNICO PRELIMINAR - ETP</b>	<b>62</b>
<b>DA METODOLOGIA DE GESTÃO DO PROJETO</b>	<b>63</b>
<b>PROPOSTA</b>	<b>63</b>
<b>CRONOGRAMA</b>	<b>64</b>
<b>CONCLUSÃO</b>	<b>64</b>
<b>REFERÊNCIAS</b>	<b>66</b>

## 1) APRESENTAÇÃO

### Objetivo

Este projeto tem como objetivo descrever o processo de pesquisa e desenvolvimento de Módulo de Instrução Assistida por Inteligência Artificial de Representações de Aquisições Públicas para concorrência e seleção em processo de contratação direta pelo TCU por meio de encomenda tecnológica - ETEC, em resposta ao chamamento público previsto no edital de processo nº TC 036.798/2020-7.

### Proponentes

As proponentes deste projeto de encomenda tecnológica são:

**REDEEMPRESARIAL SERVIÇOS WEB LTDA**, inscrita no CNPJ sob o nº 05.907.755/0001-98, detentora da marca e plataforma de automação **COREJUR**, e

**CENTRO DE EXCELÊNCIA EM INTELIGÊNCIA ARTIFICIAL - UNIDADE EMBRAPII CEIA/UFG**, vinculado à **UNIVERSIDADE FEDERAL DE GOIÁS - UFG** – instituição federal de ensino e pesquisa de nível superior, inscrita no CNPJ sob o nº 01.567.601/0001-43, que em conjunto formam o consórcio **COREJUR.CEIA** para a propositura do projeto, conforme previsto no item 3 - Da participação na Encomenda Tecnológica do edital de referido chamamento público.

A **REDEEMPRESARIAL SERVICOS WEB LTDA (COREJUR)** e Centro de Excelência em Inteligência Artificial - Unidade Embrapii CEIA/UFG, assinaram em em 21/07/2021 com vigência de 18 meses o projeto "Classificação de Documentos e Extração Inteligente de Informações em Textos Jurídicos" sob o código Embrapii PEIA-2107.0014 com valor total de R\$ 1.076.986,20, que apresenta resultados promissores e serão aplicados diretamente nesta ETEC.

A classificação de parágrafos das petições iniciais desenvolvido nesta parceria já alcança uma acurácia global de 87% para identificação de teses jurídicas nos parágrafos analisados. Esta parceria entre as proponentes mostra o grau de alinhamento existente entre as equipes, que atualmente já produzem, em conjunto, soluções que buscam alcançar o estado-da-arte na área de Inteligência Artificial.

No mesmo sentido, esta parceria pré-existente demonstra a capacidade de apresentação de uma solução viável e adequada ao problema proposto, uma vez que já há experiência de ambas as proponentes com soluções praticamente idênticas. Os resultados obtidos serão apresentados no capítulo referente às soluções

O CEIA possui um amplo portfólio de projetos com atuação em diversos segmentos tais como, saúde, varejo, energia, eletrônicos, logística, dentre outros. Nas últimas três edições do Challenge de IA da RSNA (Radiological Society of North American) sempre houve um membro do CEIA no pódio, com destaque para a terceira posição geral em 2017 entre mais de 300 competidores. No campo de NLP (Natural Language Processing) destacamos alguns prêmios

internacionais recebidos:

- 1) Winner IberLEF: Iberian Languages Evaluation Forum, Iberian Languages Evaluation Forum.
- 2) Winner Assin2 - Evaluation of Semantic Textual Similarity and Textual Inference in Portuguese (Avaliação de Similaridade Semântica e Inferência Textual), STIL - Symposium in Information and Human Language Technology (Propor 2019)
- 3) Reconhecimento do Allen Institute for Artificial Intelligence (<https://allennlp.org/elmo>), mantido por Paul Allen fundador da Microsoft, que definiu a rede neural da **CEIA** como referência de resultado para problemas envolvendo a língua portuguesa.
- 4) Terceiro lugar geral no Pediatric Bone Age Challenge da Sociedade Americana de Radiologia - RSNA em 2017.
- 5) Duas vezes (2019 e 2021) projeto de destaque (spotlight talks) no principal evento de IA do mundo ( NeurIPS).
- 6) Em 2021, pesquisadores do CEIA conquistaram o primeiro lugar no COUGH Challenge do INTERSPEECH 2021, principal evento de processamento de voz do mundo.

Importante ressaltar também que a infraestrutura utilizada pelo CEIA para o treinamento de modelos de Inteligência Artificial é sem paralelo na América Latina, contando com capacidade de processamento de dados essencial para a execução deste projeto. Como exemplo, pode-se citar que o CEIA conta com um exemplar da **NVIDIA DGX A100, que conta com oito núcleos de GPUs NVIDIA A100, totalizando 640 GB de memória e apresenta os melhores resultados do mercado quanto ao treinamento de modelos de IA**, o que demonstra a infraestrutura de ponta e única no Brasil disponível para utilização neste projeto. Registre-se que sem uma estrutura similar, o presente projeto tem sérios problemas para ser viável, uma vez que o treinamento de modelos do tamanho proposto nesta ETEC exige a utilização de infraestrutura de ponta, em especial o modelo GPT-2, proposto para a solução de redação/sumarização automática de textos.

Já a **COREJUR** é uma plataforma líder de mercado em automação de processos jurídicos, com experiência consolidada nesta área por meio de serviços como o processamento automático de publicações, redação assistida de documentos jurídicos, RPA e desenho de fluxos de processos relativos à área jurídica. A **COREJUR** possui em seu portfólio a prestação de serviços de automação jurídica para diversos agentes de relevância nacional, como o Supremo Tribunal Federal, a OAB/GO, a Procuradoria Geral do Estado de Goiás, além de diversas empresas e escritórios de advocacia de grande porte. Desta maneira, a **COREJUR tem amplo domínio da área de negócios relevante à presente ETEC**, sendo essencial para o direcionamento das equipes técnicas e para as definições das diretrizes de negócio a serem observadas na execução deste projeto, bem como para o desenvolvimento dos eventuais modelos de documento a serem utilizados pelo TCU nesta ETEC.

## **Introdução**

Consciente da capacidade e maturidade tecnológica disponível a nível mundial, em que as possibilidades de aplicação da inteligência artificial (I.A.) surpreendem e oferecem soluções precisas em situações diversas antes inviáveis através de rotas tecnológicas tradicionais, o TCU percebeu a possibilidade de expressiva melhoria de processos em desempenho, produtividade e grande redução de custo com a aplicação de técnicas avançadas em I.A. no âmbito da análise de representações de aquisições públicas.

No entanto, ainda que a aplicação de I.A. tenha resolvido questões antes sem solução, muito do que é a capacidade da tecnologia encontra-se em processo de amadurecimento, conceituação ou experimentação, sem a possibilidade de aplicação a casos concretos. Soma-se a este entrave a realidade altamente customizável e específica de projetos de I.A., que precisam ser implementados e ajustados para atuações únicas e personalizadas.

Diante deste cenário e da realidade única da atuação do tribunal, o TCU lançou a concorrência para contratação direta por meio de encomenda tecnológica - ETEC, conforme previsto na Lei 10.973/04 para o desenvolvimento de uma solução de Instrução Assistida por Inteligência Artificial.

Para atingir os desafios inovadores propostos, o edital desta ETEC tem por objeto as seguintes soluções, a serem realizadas por módulos de inteligência artificial:

- 1) Detecção de significado nas peças processuais. Inclui identificação das alegações, exame de admissibilidade, cálculo da probabilidade de concessão de medidas cautelares.
- 2) Painel de jurimetria. Inclui priorização de processos e comparação com causas anteriores;
- 3) Redação de peças. Inclui geração de comunicações aos interessados e de instruções contendo sumarização de teses e predição da análise técnica e das propostas de encaminhamento.

Os desafios propostos no referido edital são complexos, ousados e de entrega incerta. Por um lado tais características justificam a contratação da ETEC apresentada, mas por outro mantém a expectativa de entrega em risco.

## 2) DA SOLUÇÃO

### Síntese

As proponentes apresentam sua solução à problemática delimitada no objeto deste edital, e, em especial, no Anexo I do Termo de Referência desta ETEC. A solução apresentada, conforme será detalhada a seguir, é composta de quatro componentes essenciais:

- 1) Um modelo de visão computacional para a extração de parágrafos das peças processuais;
- 2) Um conjunto de classificadores para a identificação de significado das peças, cálculo de probabilidades de procedência, concessão de cautelares, admissibilidade, comunicações, demais encaminhamentos e elaboração do painel de jurimetria;
- 3) Um conjunto de modelos para extração de entidades nomeadas também para identificação de significado das peças, cálculo de probabilidades de procedência, concessão de cautelares, admissibilidade, comunicações, demais encaminhamentos e elaboração do painel de jurimetria; e
- 4) Um modelo de geração automática de textos, com o fim de redigir automaticamente componentes das propostas de encaminhamento e comunicações a serem enviadas aos interessados.

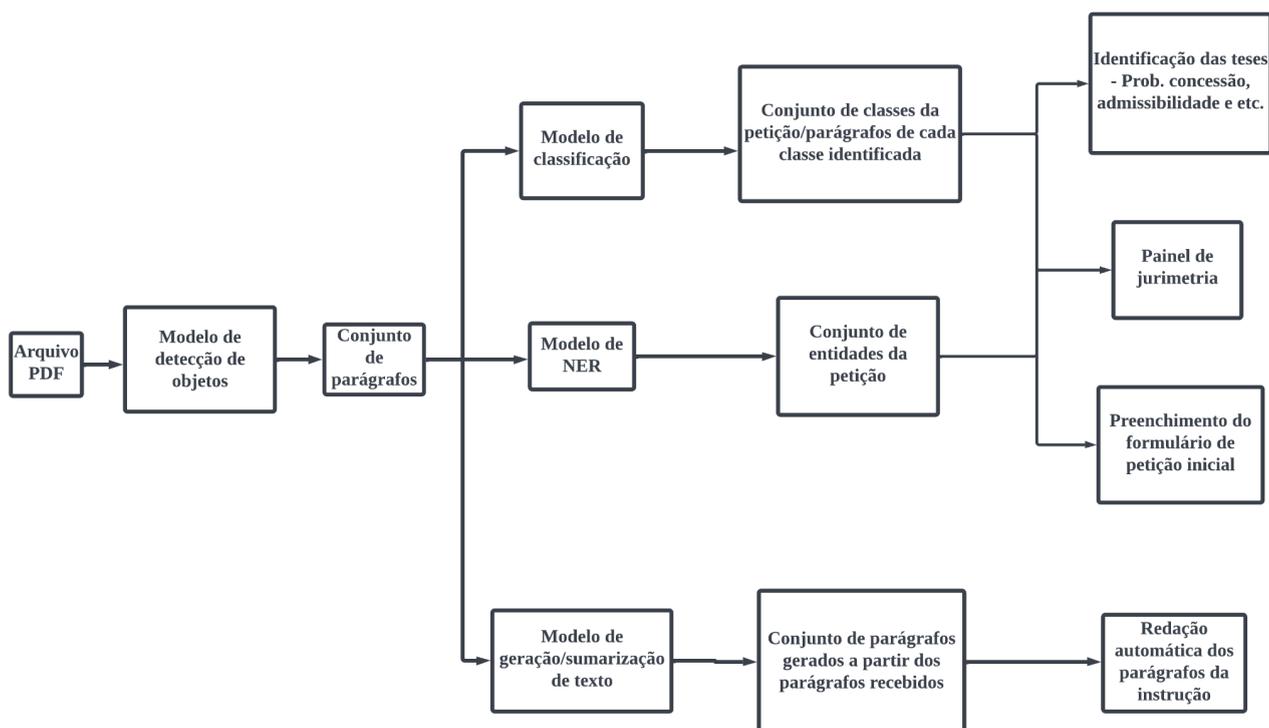
As quatro etapas da solução serão detalhadas em nível acadêmico nos capítulos a seguir, bem como suas interdependências, fundamentos teóricos, científicos e práticos, inclusive a sua adequação aos problemas definidos nesta ETEC. As proponentes se utilizam de evidências científicas colhidas em pesquisa abrangente da literatura científica mais avançada acerca dos temas debatidos, trazendo resultados e experiências que atingem o estado-da-arte em relação às tarefas propostas pelas proponentes.

Ao total serão realizadas quatorze macro entregas, uma vez que tem-se três ciclos temáticos, cada um com três tarefas associadas (classificação, reconhecimento de entidades nomeadas e geração/sumarização automática de texto). Para cada ciclo temático serão realizadas quatro entregas, sendo a primeira o dataset anotado, a segunda o modelo de classificação, a terceira o modelo de reconhecimento de entidades nomeadas e o quarto o modelo de geração/sumarização automática de textos. Adicionalmente, antes do primeiro ciclo temático, será entregue um modelo de detecção de objetos, e por fim, será entregue a plataforma de inferência para consumo do TCU:

## MACRO ENTREGAS

- 1 - Modelo de detecção de objetos;
- 2 - Dataset anotado - Ciclo temático 1
- 3 - Modelo de classificação - Ciclo temático 1
- 4 - Modelo de reconhecimento de entidades nomeadas - Ciclo temático 1
- 5 - Modelo de geração/sumarização automática de textos - Ciclo temático 1
- 6 - Dataset anotado - Ciclo temático 2
- 7 - Modelo de classificação - Ciclo temático 2
- 8 - Modelo de reconhecimento de entidades nomeadas - Ciclo temático 2
- 9 - Modelo de geração/sumarização automática de textos - Ciclo temático 2
- 10 - Dataset anotado - Ciclo temático 3
- 11 - Modelo de classificação - Ciclo temático 3
- 12 - Modelo de reconhecimento de entidades nomeadas - Ciclo temático 3
- 13 - Modelo de geração/sumarização automática de textos - Ciclo temático 3
- 14 - Plataforma de inferência para consumo pelo TCU

A solução final vislumbrada pela **COREJUR** e pelo **CEIA** se dá enquanto um *pipeline* que recebe a informação binária do arquivo em formato pdf sobre o qual será realizada a inferência. A partir do binário do pdf, serão extraídos os parágrafos por meio do modelo de detecção de objetos. Em sequência, cada um dos parágrafos será dado como entrada para as inferências dos três modelos propostos nesta solução, isto é: classificação, reconhecimento de entidades nomeadas e geração/sumarização automática de texto. Desta maneira, ao final das inferências, os modelos indicarão as classes existentes no documento, bem como os parágrafos que se enquadram em cada uma das classes, o conjunto de entidades nomeadas e a redação dos parágrafos da instrução:



A ideia de solução final, portanto, é que as petições recebidas pelo TCU em arquivos pdf já passem pelo *pipeline* apontado acima de maneira automática, gerando as informações necessárias para que o auditor elabore a instrução de maneira assistida, conforme é objeto deste edital, bem como sejam refletidas no painel de jurimetria. O desenvolvimento das aplicações para o consumo das inferências dos modelos referidos acima faz parte deste projeto, como descrito no cronograma.

Percebe-se, por conseguinte, que as proponentes apresentam soluções adequadas aos problemas levantados por esta ETEC. Nesse sentido, se utilizam da vasta experiência do **Centro de Excelência em Inteligência Artificial - UE CEIA/UFG** da Universidade Federal do Goiás, e da correlação temática da **COREJUR**, apresentando soluções já aplicadas por essas organizações com grande sucesso, como se passa a demonstrar em maior detalhamento.

## **Detecção de objetos para extração de parágrafos das peças processuais**

### **Solução e aderência ao edital**

A primeira etapa do projeto deve ser o treinamento de um modelo de visão computacional para que seja possível segmentar as peças processuais em parágrafos. Esta etapa do projeto é essencial para o andamento das demais, uma vez que se mostra uma estratégia viável para a segmentação de documentos extremamente longos em unidades de texto menores, porém, igualmente informativas.

Utilizar as peças processuais de maneira integral impede o uso de diversas técnicas de *machine learning*, como qualquer arquitetura baseada no modelo BERT, entre outros modelos que se utilizam de *transformers*, que possuem limitações expressivas quanto às quantidades de *tokens*.

Da mesma maneira, para textos muito longos a utilização de modelos sequenciais, como RNNs e LSTMs fica prejudicada, uma vez que as informações recursivamente passadas adiante pelo modelo acabam se perdendo para sequências muito longas. Portanto, justifica-se a necessidade de uma etapa de segmentação antes da utilização dos textos contidos nas peças processuais a serem utilizadas no treinamento.

Cabe ressaltar que os resultados obtidos pelo próprio TCU na classificação de acórdãos confirmam a necessidade de uma segmentação prévia do texto em parágrafos, na qual a acurácia do modelo se mostra muito inferior quando tenta classificar textos tão extensos.

Nesse sentido, percebe-se que a segmentação das peças processuais em parágrafos é uma condição necessária para a viabilidade de qualquer projeto que pretenda classificar as peças processuais do TCU sem que passem pelo mesmo obstáculo que impede a classificação de acórdãos completos, em razão do texto ser demasiadamente extenso.

Todas as soluções propostas neste projeto de P&D se utilizarão dos parágrafos extraídos por meio da Detectron 2, uma vez que permitem a utilização de modelos que alcançam o estado da arte, bem como permitem uma análise em que a informação extraída das petições é maximizada, tendo em vista a divisão da peça em várias unidades informativas, isto é, parágrafos.

### **Rota tecnológica**

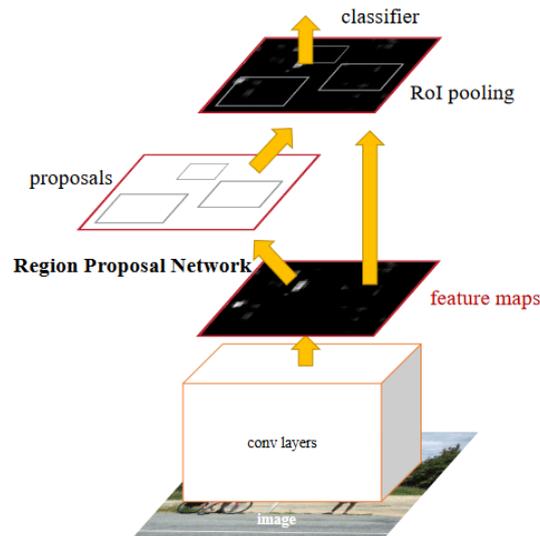
Quanto à rota tecnológica a ser utilizada, existem diversos modelos que desempenham a função de detectar objetos em imagens e poderiam ser treinados para o reconhecimento e extração adequadas dos parágrafos. Aponta-se, inicialmente o modelo chamado Detectron 2 (<https://github.com/facebookresearch/detectron2>), desenvolvido pelo Facebook, **o qual já é utilizado pela COREJUR para o mesmo objetivo**, porém com petições iniciais cíveis, com grande sucesso.

O Detectron2 se utiliza de uma arquitetura chamada Mask R-CNN<sup>1</sup> que é uma extensão da arquitetura Faster R-CNN<sup>2</sup>, sendo ambas baseadas na paralelização de convoluções nas redes neurais para reconhecimento de regiões, ou *bounding boxes*, e para a classificação da região proposta em uma classe que representa um objeto.

---

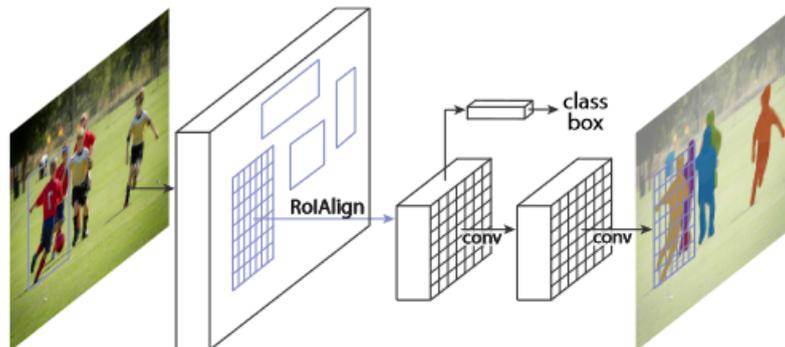
<sup>1</sup>HE, Kaiming et al. Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. 2017. p. 2961-2969.

<sup>2</sup> REN, Shaoqing et al. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, v. 28, 2015.



Arquitetura Faster R-CNN

A Mask R-CNN adiciona à Faster R-CNN dois elementos essenciais, um alinhamento das regiões de interesse da imagem de input e das convoluções realizadas, bem como um outro ramo na qual o modelo tenta prever uma máscara para o objeto a ser reconhecido. Essencialmente a Mask R-CNN se dá apenas pela adição de uma FCN<sup>3</sup> (*Fully Convolutional Neural Network*) à Faster R-CNN e a utilização do RoIAlign, no lugar da técnica de RoIPooling.



Arquitetura Mask R-CNN

A FCN tem como objetivo a possibilidade de classificação da imagem pixel a pixel, por meio de convoluções sucessivas até que as convoluções possam segmentar a imagem com um *kernel* 1x1, isto é, com apenas um pixel como janela de convolução:

<sup>3</sup> LONG, Jonathan; SHELHAMER, Evan; DARRELL, Trevor. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 3431-3440.

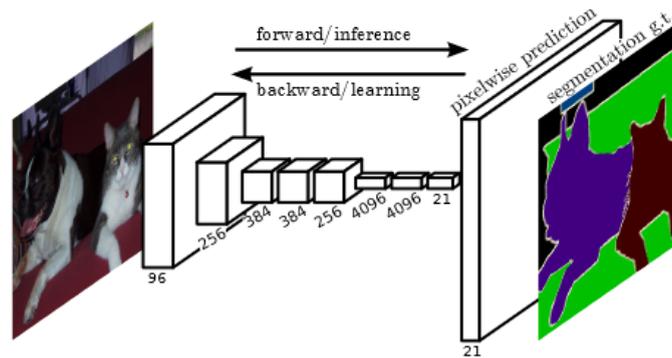


Figure 1. Fully convolutional networks can efficiently learn to make dense predictions for per-pixel tasks like semantic segmentation.

A Mask R-CNN atinge resultados compatíveis com o estado-da-arte na tarefa de detecção de objetos, o que demonstra a adequação do modelo ao problema proposto pelo TCU, tendo em vista que possibilita a classificação das peças processuais, conforme será apontado à frente. Os resultados de detecção de objetos da Mask R-CNN tem como base o dataset COCO (*Common objects in context*)<sup>4</sup>:

	backbone	AP <sup>bb</sup>	AP <sup>bb</sup> <sub>50</sub>	AP <sup>bb</sup> <sub>75</sub>	AP <sup>bb</sup> <sub>S</sub>	AP <sup>bb</sup> <sub>M</sub>	AP <sup>bb</sup> <sub>L</sub>
Faster R-CNN+++ [19]	ResNet-101-C4	34.9	55.7	37.4	15.6	38.7	50.9
Faster R-CNN w FPN [27]	ResNet-101-FPN	36.2	59.1	39.0	18.2	39.0	48.2
Faster R-CNN by G-RMI [21]	Inception-ResNet-v2 [41]	34.7	55.5	36.7	13.5	38.1	52.0
Faster R-CNN w TDM [39]	Inception-ResNet-v2-TDM	36.8	57.7	39.2	16.2	39.8	52.1
Faster R-CNN, RoIAlign	ResNet-101-FPN	37.3	59.6	40.3	19.8	40.2	48.8
Mask R-CNN	ResNet-101-FPN	38.2	60.3	41.7	20.1	41.1	50.2
Mask R-CNN	ResNeXt-101-FPN	39.8	62.3	43.4	22.1	43.2	51.2

Table 3. Object detection *single-model* results (bounding box AP), vs. state-of-the-art on *test-dev*. Mask R-CNN using ResNet-101-FPN outperforms the base variants of all previous state-of-the-art models (the mask output is ignored in these experiments). The gains of Mask R-CNN over [27] come from using RoIAlign (+1.1 AP<sup>bb</sup>), multitask training (+0.9 AP<sup>bb</sup>), and ResNeXt-101 (+1.6 AP<sup>bb</sup>).

Utilizando a arquitetura mencionada acima é possível atingir o estado-da-arte nas tarefas de segmentação semântica e detecção de objetos, de acordo com os artigos citados, o que permite compreender que a arquitetura do Detectron 2 é extremamente promissora para uma tarefa de relativamente baixa complexidade, como a detecção de parágrafos em um texto.

<sup>4</sup> IN, Tsung-Yi et al. Microsoft coco: Common objects in context. In: European conference on computer vision. Springer, Cham, 2014. p. 740-755.

Ademais, a divisão das peças processuais em parágrafos permite ampliação significativa do conjunto de dados, uma vez que, por exemplo, cada uma das representações feitas ao TCU tem dezenas de parágrafos, que podem ter seu conteúdo analisado separadamente.

Outro benefício da presente estratégia é evitar a utilização de classificações *multilabel* na qual cada um dos textos poderia ter mais de uma rotulação. No caso da divisão entre parágrafos, a grande maioria poderá ser classificada com apenas um rótulo, sem maiores problemas, enquanto o conjunto das classificações dos parágrafos de uma dada peça representa fielmente as classes às quais o referido documento integra.

No mesmo sentido se encontra a dificuldade apontada pelo TCU em extrair adequadamente o texto de petições digitalizadas, uma vez que contém outros objetos, que não textos, a serem excluídos da análise, tais como carimbos, rodapés, cabeçalhos etc. **Este problema já é solucionado pela COREJUR** na utilização do Detectron 2, uma vez que se pode ensinar ao modelo que os mencionados objetos pertencem a classes que não a do texto da peça propriamente dito.

Por fim, quanto à maturidade tecnológica da solução, a **COREJUR**, em conjunto com a **UE CEIA/UFG** já realizou o treinamento de um modelo Detectron 2, para a mesma tarefa proposta neste projeto, isto é, a detecção e extração de parágrafos oriundos de peças processuais. O modelo já foi treinado com resultados satisfatórios em ambiente relevante, e atualmente se encontra em fase de produção, sendo utilizada para a extração de parágrafos, de maneira idêntica ao que será desenvolvido para o TCU, tornando relativamente simples atingir o nível de TRL 9.

### **Etapas para desenvolvimento da solução**

Etapa 1 – Recebimento e armazenamento das peças processuais em formato pdf, conversão dos pdfs em imagem e realização de upload na plataforma de anotação escolhida (atualmente, utiliza-se o LabelStudio - <https://labelstud.io/>), definição de diretrizes e orientação dos anotadores;

Etapa 2 - Anotação manual do texto das imagens geradas a partir das peças processuais, separando os objetos de interesse (texto da peça) dos demais (cabeçalhos, rodapés, carimbos etc.).

text 1 | title 2 | list 3 | table 4 | figure 5 | legend 6 | header 7 | footer 8

11

Fabio Francisco e Heitor Barros da Cruz  
Advogados Associados

O enunciado da Súmula 297 do Superior Tribunal de Justiça, diz que o Código de Defesa do Consumidor é aplicável às instituições financeiras. Ademais, determina o artigo 1º, inciso III, da Carta Magna, que o princípio da dignidade humana deverá estar presente nas relações de Direito Privado.

Outrossim, a Súmula 479, do C. Superior Tribunal de Justiça, foi editada para fixar a responsabilidade objetiva das instituições financeiras por fraudes e delitos praticados por terceiros, impedindo ilações a respeito de culpa exclusiva do consumidor para casos como o desta demanda, como bem descrito acima de como se deram os fatos.

Isso porque, para a teoria da responsabilidade integral, é risco do próprio empreendimento o que aqui se caracteriza como fortuito interno, e intrínseco à atividade econômica pela massificação dos contratos e impessoalidade nos tratos, de onde advém grande lucro às poderosas instituições, como os Bancos.

Tal responsabilidade acarreta no dever de indenizar o lesado pela ocorrência de fraude, pois é uma ação previsível e, normalmente, evitável. A responsabilidade objetiva, com base na teoria do risco-proveito adotada pelo Código de Defesa do Consumidor, tem como pontos fundamentais: a produção em massa, a vulnerabilidade do consumidor, a insuficiência da responsabilidade subjetiva, a existência de precedentes legislativos e principalmente o fato de o prestador ou o fonecedor de serviços responder pelos riscos que os seus produtos acarretam, já que lucra (muito) com seus negócios.

Av. Salgado Filho, 610 - Jd. Maia - Curitiba/PR - SP - CEP: 07115-000 - PABX (11) 2137-7077  
www.fbfbc.adv.br | rf@fbcadv.br

Este documento é cópia do original, assinado digitalmente por FABIO FRANCISCO e HEITOR BARROS DA CRUZ em 15/10/2021 às 10:04:59 e número 1015248020216290016. Para conferir o original, acesse o site <https://esaj.tjsp.jus.br/pesquisa/verificarConteudoDocumento.do>, informe o processo 1015248-09.2021.8.26.0016 e código BC5CAAA.

### Exemplo de anotação de petição cível com o LabelStudio

Etapa 3 – Treinamento do modelo Detectron 2 a partir das peças anotadas;

Etapa 4 – Utilização do modelo treinado para a extração de parágrafos das peças processuais;

### Cronograma

Período	Atividades	Entregáveis
Sem 1	Elaboração das diretrizes de anotação, orientação e treinamento dos anotadores - Início do período de anotação	Guidelines de anotação e 33% das imagens anotadas
Sem 2	Anotação das imagens extraídas dos arquivos pdf	66% das imagens anotadas
Sem 3	Anotação das imagens extraídas dos arquivos pdf	100% das imagens anotadas
Sem 4	Treinamento e teste do modelo de extração de parágrafos, transferência do conhecimento ao TCU	Entrega final do modelo de extração de parágrafos treinado - Extração dos parágrafos necessários à anotação das próximas fases do projeto - Entrega do dataset não rotulado <b>MACRO ENTREGA</b>

### Classificação dos parágrafos, reconhecimento de entidades nomeadas, painel de jurimetria e redação das instruções e comunicações

#### Soluções propostas e aderência ao edital

Em primeiro lugar, faz-se necessário apontar as soluções oferecidas aos problemas de classificação e reconhecimento de entidades nomeadas (NER) apontados no edital, uma a uma, de maneira detalhada, apontando sua estruturação enquanto projeto de aprendizado de máquina, bem como sua adequação às necessidades do TCU expostas no edital da presente ETEC.

De maneira geral, este projeto de P&D compreende que para a solução completa e adequada das referidas problemáticas do TCU, a solução proposta deve realizar classificações quanto a 6 grupos de classes, dos quais, 2 possuem entidades nomeadas a serem extraídas. Os referidos grupos serão analisados individualmente neste projeto, e podem ser listados como segue:

- i. Classificação quanto à admissibilidade;
- ii. Classificação quanto aos requisitos de concessão de medida cautelar;
- iii. Classificação quanto aos itens do formulário de petição inicial;

- iv. Classificação quanto aos tipos de irregularidade/Classificação quanto aos temas residuais (Segundo ciclo);
- v. Classificação quanto às comunicações a serem enviadas aos interessados;
- vi. Classificação quanto aos encaminhamentos.

As classificações acima delineadas serão utilizadas para todos os três ciclos temáticos, com exceção da classificação quanto aos tipos de irregularidade, que provavelmente será substituída por uma classe que permita a identificação de quais dos temas residuais se encontram na peça processual, para o segundo ciclo temático. Frise-se que este projeto prevê a definição dessas classes no período de saneamento de dados, conforme será detalhado no cronograma.

Já a classificação quanto aos itens do formulário da petição inicial será utilizada, porém faz parte deste projeto o acompanhamento do TCU para a elaboração de um formulário específico para o segundo ciclo temático. O terceiro ciclo se aproveitará do formulário já existente, com os ajustes necessários em razão da Lei 14.133/2021, que também serão previstos no cronograma, e também se utilizará da classificação quanto ao tipo de irregularidade.

Por fim, deve-se compreender que as classificações indicadas a seguir, bem como as classes elencadas, são suficientes para a solução dos problemas propostos nesta ETEC, com exceção da redação automática de texto e da sumarização, que serão propostas em outro capítulo. Entretanto, o projeto define uma etapa de validação e definição final das classes e entidades, momento no qual os especialistas da área fim do TCU serão convocados para a delimitação final das classificações, suas respectivas classes e as entidades nomeadas a serem extraídas dos parágrafos, com possibilidade de alteração e adição de entidades nomeadas e classificações, bem como suas classes, a depender do entendimento dos especialistas da área fim do TCU.

### **Classificação quanto à admissibilidade**

Nos termos do Anexo I do Termo de Referência desta ETEC, a primeira solução necessária ao TCU se dá pela classificação de existência dos requisitos de admissibilidade das representações e denúncias, nos termos dos arts. 234 a 237 do RITCU.

Nesse sentido, os parágrafos das petições iniciais devem ser classificados quanto à indicação de existência ou não de algum destes requisitos, que serão tratados como classes do problema. Da mesma forma, há duas entidades nomeadas que auxiliam na compreensão da existência ou não dos requisitos de admissibilidade da petição inicial.

Em adequação ao Anexo I do Termo de Referência da ETEC, inicialmente propõe-se a estruturação do projeto de classificação dos parágrafos quanto à existência dos requisitos de admissibilidade com as seguintes classes:

**i. Indício concernente à irregularidade**

Parágrafo no texto da petição inicial que aponta um indício de irregularidade, no entendimento estabelecido pelo TCU para anotação. Frise-se que não basta a afirmação na petição inicial para que a classe seja anotada, mas sim que o entendimento do TCU confirme que, dado o parágrafo em questão, entende-se haver indício de irregularidade, por meio de sua jurisprudência e seus critérios legais de avaliação.

**ii. Indício concernente à regularidade**

Parágrafo no texto da petição inicial que aponta um indício de ausência de irregularidade, no entendimento estabelecido pelo TCU para anotação. O autor jamais afirmará que a aquisição pública se deu de maneira regular, por definição, portanto esta classe deve ser anotada quando, pela interpretação do parágrafo, seja possível inferir o entendimento do TCU de que não haja irregularidade, por meio de sua jurisprudência e seus critérios legais de avaliação, mesmo que alegado pelo autor.

**iii. Competência do TCU**

Parágrafo no texto da petição que indique o pertencimento da matéria tratada na petição inicial à competência do TCU. Da mesma forma, não basta a afirmação na petição inicial para que a classe seja anotada, mas sim que o entendimento do TCU confirme que, dado o parágrafo em questão, entende-se haver indicação clara da matéria ser da competência deste TCU.

**iv. Incompetência do TCU**

Parágrafo no texto da petição que indique o não pertencimento da matéria tratada na petição inicial à competência do TCU. O autor jamais afirmará que a corte de contas é incompetente, por definição, portanto esta classe deve ser anotada quando, pela interpretação do parágrafo, seja possível inferir o entendimento do TCU de que esse não é competente para a análise, por meio de sua jurisprudência e seus critérios legais de avaliação, mesmo que alegado pelo autor.

**v. Presença de interesse público**

Parágrafo no texto da petição que indique a existência de interesse público nas alegações do autor. Da mesma forma, não basta a afirmação na petição inicial para que a classe seja anotada, mas sim que o entendimento do TCU confirme que, dado o

parágrafo em questão, entende-se haver indicação clara da existência de interesse público.

**vi. Ausência de interesse público**

Parágrafo no texto da petição que indique a inexistência de interesse público nas alegações do autor. O autor jamais afirmará que sua alegação carece de interesse público, por definição, portanto esta classe deve ser anotada quando, pela interpretação do parágrafo, seja possível inferir o entendimento do TCU de que não há interesse público, por meio de sua jurisprudência e seus critérios legais de avaliação, mesmo que alegado pelo autor.

**vii. Parágrafo de qualificação**

O parágrafo no qual consta a qualificação do autor, necessário para a admissibilidade da inicial, nos termos do art. 235 do RITCU.

Registre-se que as classes referentes aos indícios de irregularidades não irão compor as classes para a identificação e cálculo de probabilidade da tese de admissibilidade no segundo ciclo temático, fase na qual serão definidos substitutas ou apenas removidas as classes referidas.

Apresentam-se também três entidades nomeadas úteis à definição da presença ou não dos requisitos de admissibilidade da inicial. Frise-se que, enquanto as classes são associadas a parágrafos inteiros, as entidades nomeadas são trechos destes parágrafos a serem anotados no nível de *tokens* e não de sentenças:

**i. CNPJ/CPF do autor**

Deve-se anotar o trecho do parágrafo que contém o CNPJ/CPF do autor, para checagem de sua legitimidade para postular perante a corte de contas.

**ii. Endereço do autor**

Deve-se anotar trecho do parágrafo no qual consta o endereço do autor, necessário para a admissibilidade da inicial, nos termos do art. 235 do RITCU.

**iii. Código de Unidades de Administração de Serviços Gerais - UASG**

Deve-se anotar o Código de Unidades de Administração de Serviços Gerais – UASG, que identifica a unidade da administração pública envolvida no processo junto ao TCU.

### **Classificação quanto aos requisitos de concessão de medida cautelar**

Dando sequência às soluções solicitadas pelo edital desta ETEC, passa-se à classificação das petições iniciais quanto à existência dos requisitos autorizadores da concessão de medida cautelar, nos termos dos arts. 273 a 276 do RITCU.

Da mesma forma que a classificação anterior, as classes serão divididas entre aquelas que apontam a existência de um dado requisito e aquelas que apontam sua inexistência.

Porém, a classificação em relação aos requisitos de concessão de medidas cautelares se dá de forma objetiva, dados os requisitos elencados pelo TCU no Anexo I do Termo de Referência desta ETEC. Isto significa que apesar de possuir um caráter jurídico, sendo necessária a consulta do entendimento do TCU, como a corte já desdobrou os requisitos legais em fatos objetivos, basta a afirmação do autor para que seja possível identificar as teses relevantes.

Em adequação ao Anexo I do Termo de Referência da ETEC, inicialmente propõe-se a estruturação do projeto de classificação dos parágrafos quanto à existência dos requisitos de concessão de medida cautelar com as seguintes classes:

#### **i. Pedido de concessão de medida cautelar**

Parágrafo que contém pedido expresso de concessão de medida cautelar ao TCU, nos termos dos arts. 276 e seguintes do RITCU.

#### **ii. Existência de contrato/ata assinado(a)**

Parágrafo por meio do qual é possível inferir a existência de um contrato administrativo ou ata de licitação assinados.

#### **iii. Inexistência de contrato assinado**

Parágrafo por meio do qual é possível inferir a inexistência de um contrato administrativo ou ata de licitação assinados.

**iv. Impugnação/decisão pendente**

Parágrafo por meio do qual é possível determinar a existência de impugnação ou decisão pendente de julgamento conclusivo pelo órgão.

**v. Inexistência de impugnação/decisão pendente**

Parágrafo por meio do qual é possível determinar a inexistência de impugnação ou decisão pendente de julgamento conclusivo pelo órgão.

**vi. Existência de decisão judicial suspensiva**

Parágrafo por meio do qual é possível determinar a existência de decisão judicial que suspende a aquisição pública.

**vii. Inexistência de decisão judicial suspensiva**

Parágrafo por meio do qual é possível determinar a inexistência de decisão judicial que suspenda a aquisição pública.

**viii. Bem/serviço essencial ao órgão**

Parágrafo por meio do qual é possível identificar que o bem ou serviço é essencial ao órgão envolvido na aquisição pública.

**ix. Bem/serviço não essencial ao órgão**

Parágrafo por meio do qual é possível identificar que o bem ou serviço não é essencial ao órgão envolvido na aquisição pública.

**x. Ausência de contrato similar**

Parágrafo por meio do qual é possível determinar a inexistência de contrato de bem/serviço alegadamente essencial similar àquele que é objeto da inicial.

**xi. Existência de contrato similar**

Parágrafo por meio do qual é possível determinar a existência de contrato de bem/serviço alegadamente essencial similar àquele que é objeto da inicial.

**xii. Risco de lesão, inexecução ou execução insatisfatória**

Parágrafo por meio do qual é possível determinar a existência de risco de lesão, inexecução ou execução insatisfatória.

**xiii. Inexistência de risco de lesão, inexecução ou execução insatisfatória**

Parágrafo por meio do qual é possível determinar a inexistência de risco de lesão, inexecução ou execução insatisfatória.

As classes que apenas podem ser usadas para o primeiro ciclo temático deverão ser substituídas ou removidas por completo para o segundo ciclo temático, caso não façam parte da análise da concessão de medida cautelar para o referido ciclo. Nesse sentido, serão levantadas as classes necessárias para o segundo ciclo temático, conforme previsto no cronograma.

Para a identificação da presença dos requisitos autorizadores da concessão de medida cautelar, deve-se adicionar a necessidade de extração dos normativos alegadamente violados pelo órgão requerido pela petição inicial. Deve-se pontuar que neste caso, basta a alegação da parte autora para que a anotação do normativo seja útil. Assim, a entidade nomeada a ser anotada, a nível de *token* é a seguinte:

**i. Normativos violados**

Normativo indicado pela parte autora como violado pelos atos ou omissões do órgão requerido na petição inicial. Apenas devem ser anotados quando a parte autora indicar expressamente sua violação.

**Classificação quanto aos itens do formulário de petição inicial**

A próxima funcionalidade da solução proposta se dá pela classificação dos parágrafos das petições iniciais com o objetivo de realizar o preenchimento automático do formulário de estruturação da petição inicial, nos termos do edital. Deve-se frisar que as classes de preenchimento do referido formulário que aparecem nas classificações anteriores foram excluídas desta tarefa, uma vez que já serão identificadas pelos outros grupos de classes.

Ademais, o formulário de petição inicial para o segundo e terceiro ciclos temáticos deverá ser elaborado junto ao TCU, o que será previsto no cronograma deste projeto. O mais provável é que ocorra uma adaptação do formulário existente (primeiro ciclo) para os demais, principalmente em relação ao terceiro ciclo, que possui temática bastante semelhante ao primeiro, porém sob a égide de nova legislação. Para o segundo a adaptação também deverá ser temática, sem maiores problemas.

Nessa toada, segundo os termos do Anexo I do Termo de Referência desta ETEC, as classes identificadas para solucionar a necessidade de preenchimento automático dos formulários de estruturação da petição inicial de aquisições públicas são as seguintes:

**i. Registro de preço**

Parágrafo que questiona fato ou questão de direito relacionada ao registro de preços.

**ii. Origem dos recursos**

Parágrafo que aponta a origem dos recursos que são objeto da pretensão delineada na petição inicial. No Anexo I do Termo de Referência desta ETEC apenas é possível observar uma das origens possíveis, isto é, “Aplicação direta de recursos federais”, porém, cada possibilidade de origem presente no formulário será uma classe separada.

**iii. Recursos oriundos de convênio**

Parágrafo que permita concluir que a origem dos recursos que são objeto da petição inicial tem origem em convênio;

**iv. COVID**

Parágrafo que permita concluir que a petição versa acerca de recursos relacionados ao combate à COVID-19;

**v. Atividade meio**

Parágrafo que permita concluir que a aquisição pública diz respeito à atividade meio do órgão adquirente;

**vi. Atividade fim**

Parágrafo que permita concluir que a aquisição pública diz respeito à atividade finalística do órgão adquirente;

**vii. Questionamento contrato de gestão**

Parágrafo que permita identificar a existência de questionamento de um contrato de gestão;

**viii. Questionamento aspectos técnicos**

Parágrafo que permita identificar a existência de questionamento dos aspectos técnicos do objeto da aquisição pública

**ix. Questionamento aspectos procedimentais;**

Parágrafo que permita identificar a existência de questionamento dos aspectos procedimentais da seleção do fornecedor ou da condução do contrato.

Quanto às entidades nomeadas necessárias para o preenchimento do formulário de estruturação da inicial, deve-se registrar que o nome do órgão será extraído a partir de seu CNPJ ou CPF, conforme apontado acima neste projeto. No mais, resta necessário identificar o valor estimado da causa ou dos pedidos do autor:

**i. Valor estimado**

Valor atribuído pelo autor à causa ou aos seus pedidos, a depender da indicação do TCU.

**Classificação quanto aos tipos de irregularidade/Classificação quanto ao tema residual**

Outra funcionalidade da solução a ser desenvolvida para o TCU é a classificação dos parágrafos da petição inicial quanto aos tipos de irregularidades da aquisição pública apontados pelo autor em sua petição inicial.

Como já mencionado, para o segundo ciclo temático esta classificação será substituída por uma classificação quanto aos temas residuais cuja identificação da existência em peças processuais sejam de interesse do TCU.

Conforme mencionado no edital desta ETEC, bem como em seu Anexo I, para esta tarefa de classificação existem aproximadamente 100 classes, que devem ser atribuídas aos parágrafos quando a irregularidade indicada se enquadrar em uma dessas 100 classes. Deve-se registrar que, conforme mencionado pelo TCU, o desbalanceamento destas classes pode levar à necessidade de exclusão daquelas que tenham números de exemplos inferiores ao necessário para o treinamento adequado do módulo de inteligência artificial.

Importante indicar que para este grupo de classes basta a existência da alegação de irregularidade na petição, independentemente do entendimento ou jurisprudência deste TCU, uma vez que a simples constatação da existência do parágrafo associado a uma determinada classe já é o suficiente para o preenchimento do formulário.

Assim, será necessário um levantamento exaustivo destas classes, que aparecem de maneira incompleta no Anexo I do Termo de Referência desta ETEC. A título de exemplo, listam-se as classes indicadas no mencionado edital:

**i. Mesmo atestado de capacidade técnica**

Parágrafo que contém a alegação de que a empresa contratante e a subcontratada participaram do certame com o mesmo atestado de capacidade técnica

**ii. Ausência de diligências**

Parágrafo que indique a ausência de diligências para a apresentação de esclarecimentos

**iii. Decisões contrárias ao edital**

Parágrafo que indique a existência de decisões do pregoeiro em descumprimento ao edital da aquisição pública

**iv. Descumprimento do edital pela empresa vencedora**

Parágrafo que indique o descumprimento das exigências do edital pela vencedora do certame

**v. Habilitação indevida**

Parágrafo em que conste a alegação de habilitação indevida de algum dos participantes da aquisição pública;

**vi. Pedido de impugnação não respondido**

Parágrafo que alegue a existência de pedido de impugnação do edital junto ao órgão licitante, sobre o qual a autora não foi informada da decisão.

**Classificação quanto às comunicações a serem enviadas aos interessados**

Dando seguimento às funcionalidades necessárias para a efetiva solução do objeto desta ETEC, passa-se à classificação dos parágrafos das petições iniciais quanto à necessidade de emissão de comunicações aos interessados.

Da mesma forma que as soluções anteriores, os parágrafos da petição inicial serão anotados um a um quanto à possibilidade de identificar a partir desses textos a necessidade de alguma das comunicações pré-estabelecidas pelo TCU, nos termos do art. 157 e seguintes do RITCU.

Segundo as orientações do Anexo I do Termo de Referência desta ETEC, podem ser identificadas quatro classes de comunicações, quais sejam:

**i. Diligência**

Parágrafo no qual é possível perceber, de maneira explícita ou implícita, que será necessária a realização de diligências para dirimir o caso concreto

**ii. Oitiva**

Parágrafo no qual é possível perceber, de maneira explícita ou implícita, que será necessária a realização de oitiva de testemunhas para dirimir o caso concreto

**iii. Inspeção**

Parágrafo no qual é possível perceber, de maneira explícita ou implícita, que será necessária a realização de inspeção da aquisição pública para dirimir o caso concreto

#### **iv. Audiência**

Parágrafo no qual é possível perceber, de maneira explícita ou implícita, que será necessária a realização de audiência dos responsáveis para dirimir o caso concreto

#### **Classificação quanto aos encaminhamentos**

Por fim, quanto à classificação dos parágrafos quanto ao encaminhamento a ser utilizado na instrução, novamente, será necessária a definição exaustiva dos encaminhamentos possíveis para uma efetiva determinação das classes existentes para esta tarefa. Esta tarefa já foi realizada para o ciclo temático de aquisições públicas que seguem a Lei 8.666/93, porém, terá que ser realizada antes da anotação dos dados para os demais temas.

Dessa forma, encaminhamentos muito raros, isto é, que tenham quantidade de ocorrência inexpressiva, podem ser excluídos para maior qualidade da solução.

Deverão ser anotados os parágrafos que contenham alegações de irregularidade que se vinculam à necessidade de um encaminhamento específico.

O objetivo, portanto, é treinar o modelo para que identifique os parágrafos que indicam a necessidade de um encaminhamento específico, e dada essa classificação, aponte automaticamente o conjunto de encaminhamentos necessários que foram identificados na peça analisada, bem como o grau de confiança do modelo na necessidade do encaminhamento.

Importante frisar que as classes que poderiam ser adicionadas ao presente grupo, isto é, aquele referente às propostas de encaminhamento, mas já são identificadas em outro grupo de classes mais específico, como a concessão de medida cautelar, podem ser excluídas do presente grupo, dada a redundância.

Como exemplo, a partir dos encaminhamentos apontados no Anexo I do Termo de Referência desta ETEC algumas das classes a serem utilizadas para o treinamento do modelo são as seguintes:

#### **i. Procedência**

Parágrafos da petição inicial que levaram à procedência, no mérito, dos pedidos constantes da mencionada peça processual.

**ii. Improcedência**

Parágrafos da petição inicial que levaram à procedência, no mérito, dos pedidos constantes da mencionada peça processual.

**iii. Indeferimento ingresso como parte interessada**

Parágrafos da petição inicial que levaram ao indeferimento do pedido de ingresso do autor como parte interessada.

**iv. Deferimento ingresso como parte interessada**

Parágrafos da petição inicial que levaram ao deferimento do pedido de ingresso do autor como parte interessada.

**Painel de jurimetria e redação das instruções e comunicações**

Neste ponto do projeto, detalha-se o desenvolvimento de uma solução de previsão de resultados de decisões judiciais por meio de documentos que a precedem, no caso, quanto à admissibilidade, concessão de medidas cautelares, comunicações aos interessados e procedência (encaminhamentos) e, a partir dessas previsões, redija um documento adequado.

Essa é uma atividade que se encontra na fronteira da pesquisa atual, tendo em vista que grande parte da pesquisa realizada quanto a este tema tem como metodologia a classificação das decisões pelo seu resultado utilizando-se do próprio texto da decisão, o que não soluciona o problema proposto pelo TCU. A solução solicitada nesta ETEC é denominada de *outcome forecasting*, que representa a predição do resultado de uma decisão que ainda não foi proferida<sup>5</sup>.

Neste ponto, em relação à presente ETEC, deve-se frisar que as classificações a serem feitas para o *outcome forecasting* são aquelas em que não basta a afirmação da do autor acerca da existência de uma classe, sendo necessária a adequação do parágrafo à jurisprudência do TCU para a sua classificação correta.

Importante ressaltar que, a concessão de medidas cautelares, apesar de ser uma tarefa que pretende a previsão do resultado do julgamento do TCU, acaba por não se enquadrar como *outcome forecasting*, uma vez que os requisitos para a sua concessão são delimitados objetivamente pelo TCU, trazendo objetividade para a classificação a ser realizada.

Isto é, basta que o autor afirme a existência de um contrato já assinado para que essa classe seja reconhecida, enquanto para a competência do TCU, por exemplo, a análise do texto produzido pelo autor não é suficiente, sendo necessária a interpretação de acordo com a jurisprudência da corte

---

<sup>5</sup>MEDVEDEVA, Masha; WIELING, Martijn; VOLS, Michel. Rethinking the field of automatic prediction of court decisions. *Artificial Intelligence and Law*, p. 1-18, 2022.

de contas. O mesmo fato se dá com os tipos de irregularidade, que podem ser compreendidos a partir da mera citação pelo autor.

Assim sendo, pode-se dividir as classificações realizadas em dois grupos, isto é, as classificações simples e as tarefas de classificação do tipo *outcome forecasting*:

### ***Outcome forecasting***

- i. Classificação quanto à admissibilidade;
- ii. Classificação quanto aos encaminhamentos.
- iii. Classificação quanto às comunicações a serem enviadas aos interessados;

### **Classificação simples**

- iv. Classificação quanto aos requisitos de concessão de medida cautelar;
- v. Classificação quanto aos itens do formulário de petição inicial;
- vi. Classificação quanto aos tipos de irregularidade;

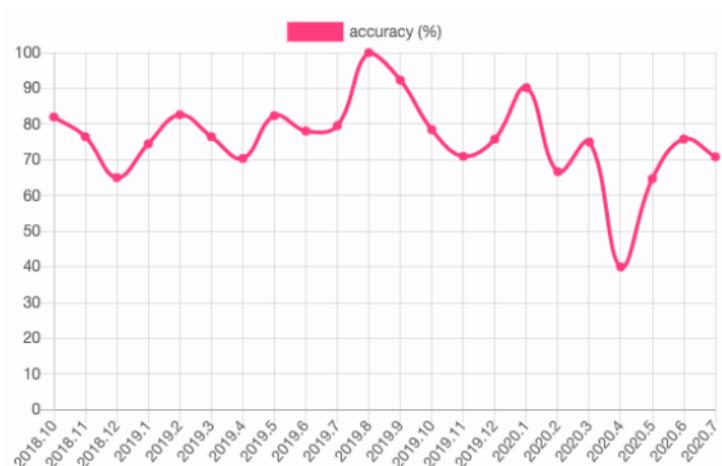
Atualmente, o melhor resultado para *outcome forecasting* encontrado na literatura entre 2015 e 2021 mostra uma acurácia de 75% (setenta e cinco por cento) na previsão de resultados da Corte Europeia de Direitos Humanos, conforme extrai-se da tabela abaixo, retirada do artigo inframencionado:

**Table 3** Research that falls under the category of outcome forecasting, including relevant court, the data used for forecasting, the best performance

Paper	Court	Data	Max. performance
Sharma et al. (2015)	SCOTUS	Court of Appeal info	70%
Katz et al. (2017)	SCOTUS	Court of Appeal info	70%
Waltl et al. (2017)	German Court of Appeal (Tax Law)	Decision of the lower (fiscal) courts	57% (F1)
Medvedeva et al. (2020b)	ECtHR	Facts as communicated to the parties	75%
Medvedeva et al. (2021)	ECtHR	Facts as communicated to the parties	66% (F1)

When instead of accuracy, the F1-score (the average between precision and recall) is used as a performance indicator, this is indicated

Deve-se apontar, ainda, que o referido modelo, denominado JURI SAYS se encontra em produção<sup>6</sup>, de acordo com as especificações encontradas no artigo de sua definição<sup>7</sup>, apresentando uma acurácia que varia de maneira significativa, passando de um método de previsão relativamente sólido à produção de previsões com pouca precisão, a depender dos documentos analisados a cada mês, conforme apontam os autores do artigo mencionado:

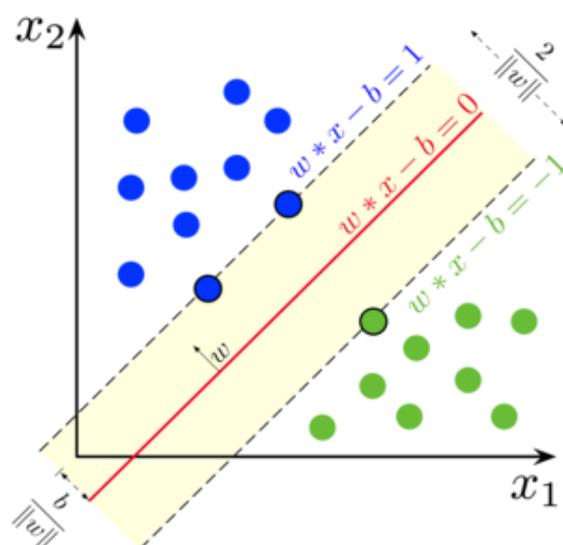


**Figure 1.** Accuracy of JURI SAYS over the past two years predicting future judgements.

A técnica utilizada no referido artigo é a SVM, técnica clássica que busca dividir as classes por meio da utilização de hiperplanos que se distanciem o máximo possível dos exemplos rotulados em cada uma das classes:

<sup>6</sup> <https://www.jurisays.com/>

<sup>7</sup> SAYS, JURI. Prediction System for the European Court of Human Rights. In: Legal Knowledge and Information Systems: JURIX 2020: The Thirty-third Annual Conference, Brno, Czech Republic, December 9-11, 2020. IOS Press, 2020. p. 277.



No que diz respeito à extração de características (*feature extraction*) o artigo se utiliza da técnica de *tf-idf*, isto é, *term frequency-inverse document frequency*, que procura determinar a relevância de uma determinada palavra em um documento, que será proporcional à sua frequência neste documento, e inversamente proporcional à frequência de aparição da palavra analisada nos documentos como um todo, isto é, no *corpus*.

Relevante apontar, portanto, que apesar de trazer resultados interessantes, as técnicas utilizadas na produção do artigo anteriormente citado não estão, necessariamente, entre aquelas que apresentam melhores resultados, como as técnicas mais modernas e robustas apresentadas no capítulo de classificação dos parágrafos. É certo que não se pode garantir que as técnicas mais novas e robustas para a classificação de textos alcançarão resultados superiores, porém, com certeza há chances relevantes a justificar a pesquisa e desenvolvimento de modelos que se utilizem desses métodos mais robustos para a tarefa de previsão dos resultados de julgamentos do TCU.

Baseada na classificação dos parágrafos referentes à admissibilidade, concessão da medida cautelar e encaminhamentos (procedência, diligências, entre outros). Os campos do painel de jurimetria podem ser construídos a partir das seguintes classificações, conforme delimitadas no capítulo anterior deste projeto e nos termos do Anexo I do Termo de Referência desta ETEC:

- i. Classificação quanto à admissibilidade – *outcome forecasting*;
- ii. Classificação quanto aos requisitos de concessão de medida cautelar – classificação simples;
- iii. Classificação quanto aos encaminhamentos – *outcome forecasting*.

Nesse sentido, a partir dos parágrafos identificados como relevantes para as classificações acima, podem ser estimadas as probabilidades de conhecimento, concessão de cautelares e procedência para uma dada petição inicial, ou peça processual qualquer. Isso se dá pela estimação

da probabilidade de um dado parágrafo fazer parte de uma peça processual a qual foi admitida, ou teve cautelar concedida, ou foi procedente, a depender destes resultados nos casos anteriores utilizados para o treinamento do modelo.

Acerca do apontamento automático dos processos relacionados, a similaridade entre peças processuais pode ser determinada a partir dos *embeddings* gerados pelo modelo para cada parágrafo. A representação vetorial dos parágrafos pode ser comparada, por diversas métricas, como a distância de cosseno e etc, a qual pode identificar os processos que contém os parágrafos mais semelhantes a aqueles contidos em uma peça processual nunca analisada pelo modelo.

Assim sendo, pode-se indicar os resultados destes processos relacionados, além da probabilidade prevista pelo modelo de classificação, como outro indicador que possa informar acerca da probabilidade de conhecimento, concessão de medidas cautelares, propostas de encaminhamento e etc.

Novamente, a segmentação das peças processuais em parágrafos se mostra essencial, uma vez que a análise de semelhança poderá ser analisada em relação a todos os tipos de classificação realizadas, no caso, para irregularidades, cautelares, admissibilidade e as demais medidas de similaridade a serem definidas pelo TCU, conforme expresso no Anexo I do Termo de Referência desta ETEC.

Quanto à redação das instruções, no que diz respeito à análise técnica e às propostas de encaminhamento, da mesma maneira, podem ser utilizadas as seguintes classificações já mencionadas:

- i. Classificação quanto aos encaminhamentos – *outcome forecasting*.
- ii. Classificação quanto aos tipos de irregularidade – Classificação simples;
- iii. Classificação quanto às comunicações a serem enviadas aos interessados – *outcome forecasting*.

Deve-se utilizar da base de encaminhamentos e análises técnicas elaborada previamente para relacionar os parágrafos classificados quanto à irregularidade e quanto aos encaminhamentos aos parágrafos a serem redigidos na instrução que o auditor deseja elaborar.

Assim, a partir da classificação dos parágrafos dentre as classes definidas para encaminhamentos e irregularidades, o modelo poderá sugerir as análises técnicas e os encaminhamentos vinculados às classes mencionadas. Cumpre registrar que, ao ser realizada a inferência do modelo para cada parágrafo, as análises e encaminhamentos também serão sugeridos de acordo com cada um dos parágrafos, permitindo uma análise mais detalhada e a realização de múltiplas sugestões de análise técnica e de encaminhamentos para cada peça processual entregue ao modelo, o que seria impossível sem a devida segmentação da peça processual em parágrafos.

Da mesma maneira, a partir da classificação dos parágrafos da peça processual quanto às comunicações a serem enviadas às partes, pode-se sugerir ao auditor quais comunicações processuais devem ser realizadas, utilizando-se da mesma técnica. Porém, quanto a este ponto, não é necessária a estruturação prévia de uma base de correspondências entre classes e sugestões, tendo em vista que a própria classificação já poderá identificar qual comunicação deve ser realizada, a partir das anotações de peças processuais.

Importante ressaltar que para a elaboração do painel de jurimetria e da redação das instruções para os demais ciclos temáticos, será necessário o levantamento inicial das correspondências entre as classes identificadas nas peças processuais e os encaminhamentos e análises técnicas relativos aos referidos ciclos, uma vez que apenas as aquisições públicas possuem esta correspondência já elaboradas. Essas etapas adicionais serão elencadas no cronograma do presente projeto.

### **Rota tecnológica**

Especificamente quanto à trilha tecnológica, deve-se frisar que o projeto de pesquisa e desenvolvimento inclui o teste de diferentes arquiteturas e técnicas para verificação empírica do melhor resultado. Nesse sentido, apresentam-se duas alternativas viáveis para o treinamento do modelo de linguagem e dos classificadores.

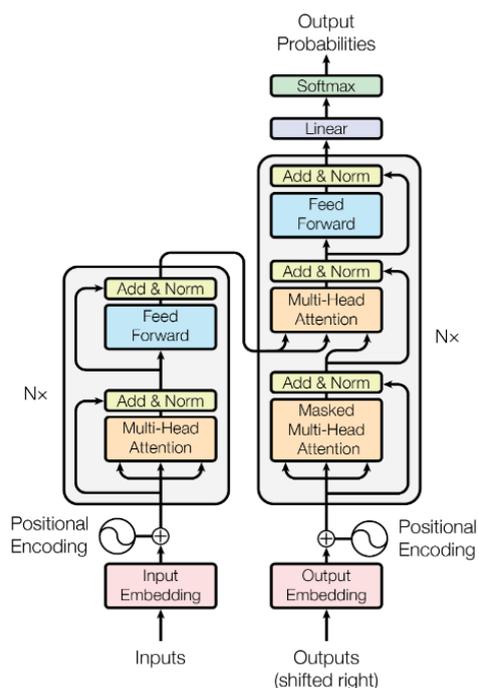
Inicialmente pode-se apontar o BERT<sup>8</sup> (*Bidirectional Encoding Representations from Transformers*) que, como sugerido pelo nome, faz o uso de *transformers*<sup>9</sup> para a modelagem de sequências sem a necessidade de recorrência, como é o caso das RNNs e LSTMs.

O transformer utiliza-se de *positional encoding*, que leva em consideração a posição do *token* para a geração dos *embeddings* que são passados como entrada para o modelo, o que permite a representação da ordem da sequência. Adicionalmente, o modelo se utiliza de um mecanismo de atenção para computar as relações existentes entre palavras de uma sequência, gerando um peso relativo a cada par de elementos da sequência, o que substitui o mecanismo de recorrência tradicionalmente utilizado na modelagem de sequências, tendo como *input* uma sequência completa, e não seus componentes, um por vez:

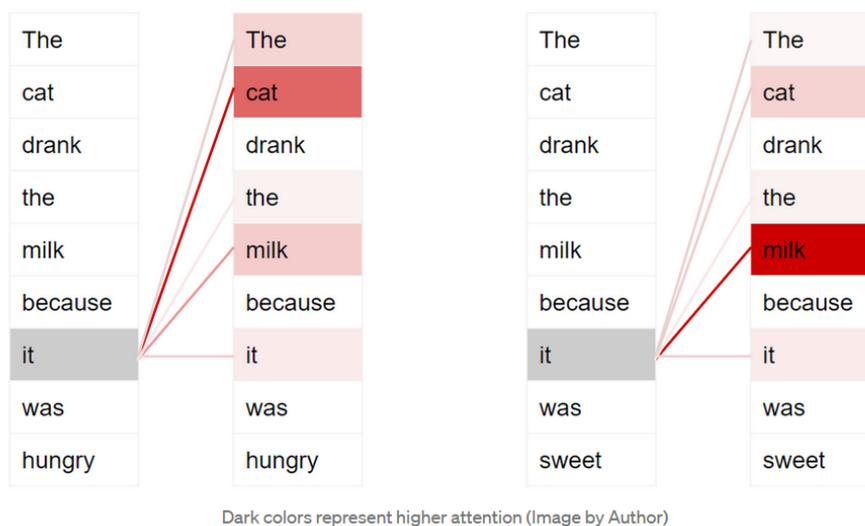
---

<sup>8</sup> DEVLIN, Jacob et al. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

<sup>9</sup> VASWANI, Ashish et al. Attention is all you need. Advances in neural information processing systems, v. 30, 2017.



Arquitetura do modelo *transformer*



Representação visual da atenção

O BERT<sub>BASE</sub> é composto de 12 camadas de *transformers*, 12 núcleos de atenção e contém 110 milhões de parâmetros, enquanto o BERT<sub>LARGE</sub> é composto 24 camadas de *transformers*, 16 núcleos de atenção e 340 milhões de parâmetros.

Usualmente o treinamento de um modelo BERT ocorre em duas etapas: inicialmente com o pré-treinamento do modelo utilizando-se das técnicas de *Masked Language Model* - MLM e *Next Sentence Prediction* - NSP, que consistem em treinar o modelo com dados não rotulados, para a

recomposição de uma sentença da qual se mascaram alguns *tokens*, e para a correta previsão da próxima sentença, dado uma sentença inicial. Por meio dessa técnica, o modelo é capaz de gerar *embeddings* de altíssima qualidade, utilizados posteriormente para a classificação das sentenças.

No caso do TCU, dada a quantidade relativamente reduzida de dados disponíveis, mesmo após a segmentação das peças processuais em parágrafos, talvez não seja possível treinar um modelo do zero a partir das peças disponibilizadas. O que pode ser feito, entretanto, é a utilização de um modelo BERT já treinado para o português para o treinamento com a linguagem específica do domínio jurídico com as peças do TCU. Deve-se registrar, ainda, que existem modelos disponíveis que já são pré-treinados com textos jurídicos e podem ser testados.

É necessário compreender que a utilização desta arquitetura só é possível graças à segmentação das peças processuais, uma vez que uma peça processual, especialmente as petições iniciais, tendem a ter bem mais que 512 *tokens*, limite estabelecido nesta arquitetura, tendo em vista a complexidade quadrática do mecanismo de atenção.

Após o pré-treinamento do modelo, por meio do MLM e NSP, passa-se ao chamado *fine-tuning*, no qual o modelo pré-treinado é utilizado para outro treinamento, desta vez da tarefa final, no caso, a classificação dos parágrafos e o reconhecimento de entidades nomeadas. Para isso, o modelo se utiliza dos *embeddings* gerados, que funcionam como as *features* do modelo de classificação.

Desta maneira o BERT alcança resultados equiparáveis ao estado-da-arte em classificação de textos nos seguintes *datasets* de classificação: CoLA<sup>10</sup> e SST-2<sup>11</sup>. Tal fato demonstra a adequação da rota tecnológica à problemática proposta pelo TCU, apontando à possibilidade de que sejam alcançados resultados semelhantes ao estado-da-arte:

System	MNLI-(m/mm) 392k	QQP 363k	QNLI 108k	SST-2 67k	CoLA 8.5k	STS-B 5.7k	MRPC 3.5k	RTE 2.5k	Average
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	93.2	35.0	81.0	86.0	61.7	74.0
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.8	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	87.4	91.3	45.4	80.0	82.3	56.0	75.1
BERT <sub>BASE</sub>	84.6/83.4	71.2	90.5	93.5	52.1	85.8	88.9	66.4	79.6
BERT <sub>LARGE</sub>	<b>86.7/85.9</b>	<b>72.1</b>	<b>92.7</b>	<b>94.9</b>	<b>60.5</b>	<b>86.5</b>	<b>89.3</b>	<b>70.1</b>	<b>82.1</b>

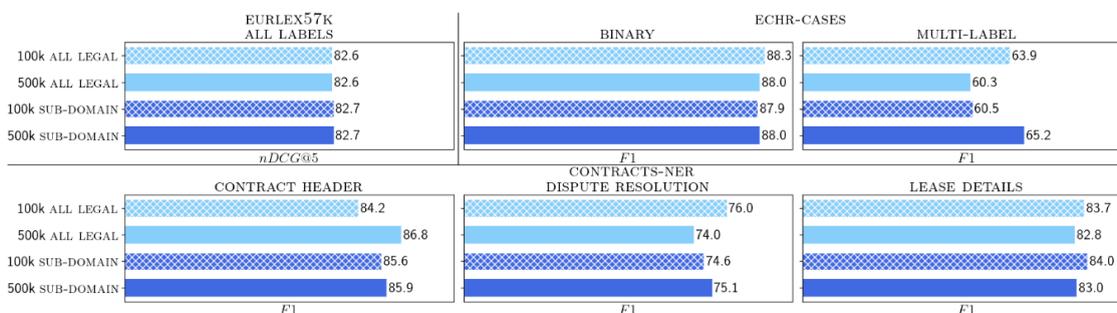
Table 1: GLUE Test results, scored by the evaluation server (<https://gluebenchmark.com/leaderboard>). The number below each task denotes the number of training examples. The “Average” column is slightly different than the official GLUE score, since we exclude the problematic WNLI set.<sup>8</sup> BERT and OpenAI GPT are single-model, single task. F1 scores are reported for QQP and MRPC, Spearman correlations are reported for STS-B, and accuracy scores are reported for the other tasks. We exclude entries that use BERT as one of their components.

Especificamente quanto ao domínio jurídico, o BERT já foi utilizado na classificação de textos e reconhecimento de entidades nomeadas, obtendo resultados compatíveis com o estado da

<sup>10</sup>WARSTADT, Alex; SINGH, Amanpreet; BOWMAN, Samuel R. Neural network acceptability judgments. Transactions of the Association for Computational Linguistics, v. 7, p. 625-641, 2019.

<sup>11</sup>SOCHER, Richard et al. Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the 2013 conference on empirical methods in natural language processing. 2013. p. 1631-1642

arte, por meio do denominado LEGAL-BERT<sup>12</sup>. Os datasets utilizados para classificação foram o EURLEX57K<sup>13</sup> que consiste em um conjunto de normas da União Europeia para classificação *multi label* e o ECHR-CASES<sup>14</sup> constituído de casos oriundos da Corte Europeia de Direitos Humanos. Já em relação ao reconhecimento de entidades nomeadas, o dataset utilizado foi o CONTRACTS-NER<sup>15</sup>, composto por contratos norte-americanos:



**Figure 3:** End-task results on development data across all datasets for LEGAL-BERT-FP variants.

Como segunda alternativa, é possível utilizar o modelo MultiFiT: *Efficient Multi-lingual Language Model Fine-tuning*<sup>16</sup>, arquitetura que já foi utilizada pelo TCU<sup>17</sup> para a classificação de acórdãos. Esta rota tecnológica pode ser usada novamente, uma vez que a segmentação das peças processuais em parágrafos elimina a dificuldade anteriormente sofrida na utilização deste modelo, isto é a classificação de textos demasiadamente longos.

Em síntese, o MultiFiT se utiliza de QRNNs – *Quasi Recurrent Neural Networks*<sup>18</sup> em sua arquitetura, com o objetivo de combinar a linearidade das RNNs e as convoluções das CNNs, com o objetivo de aplicá-las sem a sua característica inicial de serem invariáveis no tempo. Ademais, o uso de QRNNs demonstra-se eficaz pela possibilidade de paralelização das convoluções realizadas, enquanto modelos puramente recorrentes sofrem com a impossibilidade de computação paralela ao longo dos intervalos de tempo:

<sup>12</sup> CHALKIDIS, Ilias et al. LEGAL-BERT: The muppets straight out of law school. arXiv preprint arXiv:2010.02559, 2020.

<sup>13</sup> CHALKIDIS, Ilias et al. Extreme multi-label legal text classification: A case study in EU legislation. arXiv preprint arXiv:1905.10892, 2019.

<sup>14</sup> CHALKIDIS, Ilias; ANDROUTSOPOULOS, Ion; ALETRAS, Nikolaos. Neural legal judgment prediction in English. arXiv preprint arXiv:1906.02059, 2019.

<sup>15</sup> CHALKIDIS, Ilias; ANDROUTSOPOULOS, Ion; MICHOS, Achilles. Extracting contract elements. In: Proceedings of the 16th edition of the International Conference on Artificial Intelligence and Law. 2017. p. 19-28.

<sup>16</sup> EISENSCHLOS, Julian Martin et al. MultiFiT: Efficient multi-lingual language model fine-tuning. arXiv preprint arXiv:1909.04761, 2019.

<sup>17</sup> <https://github.com/piegu/language-models/blob/master/lm3-portuguese-classifier-TCU-jurisprudencia.ipynb>

<sup>18</sup> BRADBURY, James et al. Quasi-recurrent neural networks. arXiv preprint arXiv:1611.01576, 2016.

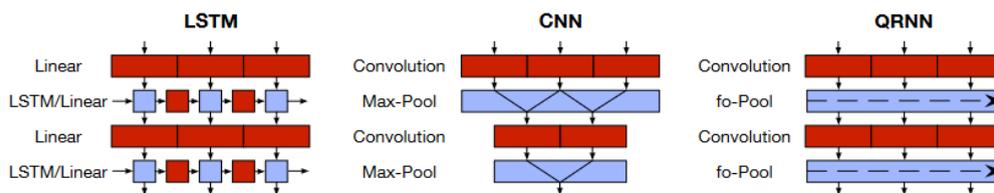
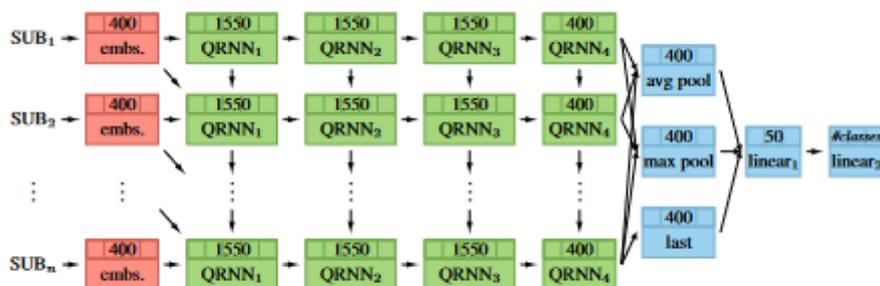


Figure 1: Block diagrams showing the computation structure of the QRNN compared with typical LSTM and CNN architectures. Red signifies convolutions or matrix multiplications; a continuous block means that those computations can proceed in parallel. Blue signifies parameterless functions that operate in parallel along the channel/feature dimension. LSTMs can be factored into (red) linear blocks and (blue) elementwise blocks, but computation at each timestep still depends on the results from the previous timestep.

A arquitetura do MultiFiT para classificação de sequências é composta por uma cama de *embedding*, quatro camadas de QRNNs, uma camada de agregação e duas camadas lineares:



Assim, inicialmente é treinado um modelo de linguagem genérico do idioma a ser utilizado, posteriormente, o modelo de linguagem é treinado com exemplos não rotulados pertencentes a um domínio de conhecimento específico, no caso o jurídico. Por fim, o modelo de linguagem é treinado com exemplos rotulados para uma tarefa de aprendizado supervisionado. Utilizando a técnica descrita, o artigo indica resultados que atingem o estado-da-arte para tarefas de classificação, tanto supervisionada quanto *zero-shot* utilizando o dataset CLS<sup>19</sup> em alemão, francês e japonês:

<sup>19</sup>PRETTENHOFER, Peter; STEIN, Benno. Cross-language text classification using structural correspondence learning. In: Proceedings of the 48th annual meeting of the association for computational linguistics. 2010. p. 1118-1127

		DE			FR			JA		
		Books	DVD	Music	Books	DVD	Music	Books	DVD	Music
<i>Zero-shot</i>	LASER, code	84.15	78.00	79.15	83.90	83.40	80.75	74.99	74.55	76.30
	MultiBERT	72.15	70.05	73.80	75.50	74.70	76.05	65.41	64.90	70.33
	MultiFiT, pseudo	<b>89.60</b>	<b>81.80</b>	<b>84.40</b>	<b>87.84</b>	<b>83.50</b>	<b>85.60</b>	<b>80.45</b>	<b>77.65</b>	<b>81.50</b>
<i>Translat.</i>	MT-BOW	79.68	77.92	77.22	80.76	78.83	75.78	70.22	71.30	72.02
	CL-SCL	79.50	76.92	77.79	78.49	78.80	77.92	73.09	71.07	75.11
	BiDRL	84.14	84.05	84.67	84.39	83.60	82.52	73.15	76.78	78.77
<i>Super.</i>	MultiBERT	86.05	84.90	82.00	86.15	86.90	86.65	80.87	82.83	79.95
	MultiFiT	<b>93.19</b>	<b>90.54</b>	<b>93.00</b>	<b>91.25</b>	<b>89.55</b>	<b>93.40</b>	<b>86.29</b>	<b>85.75</b>	<b>86.59</b>

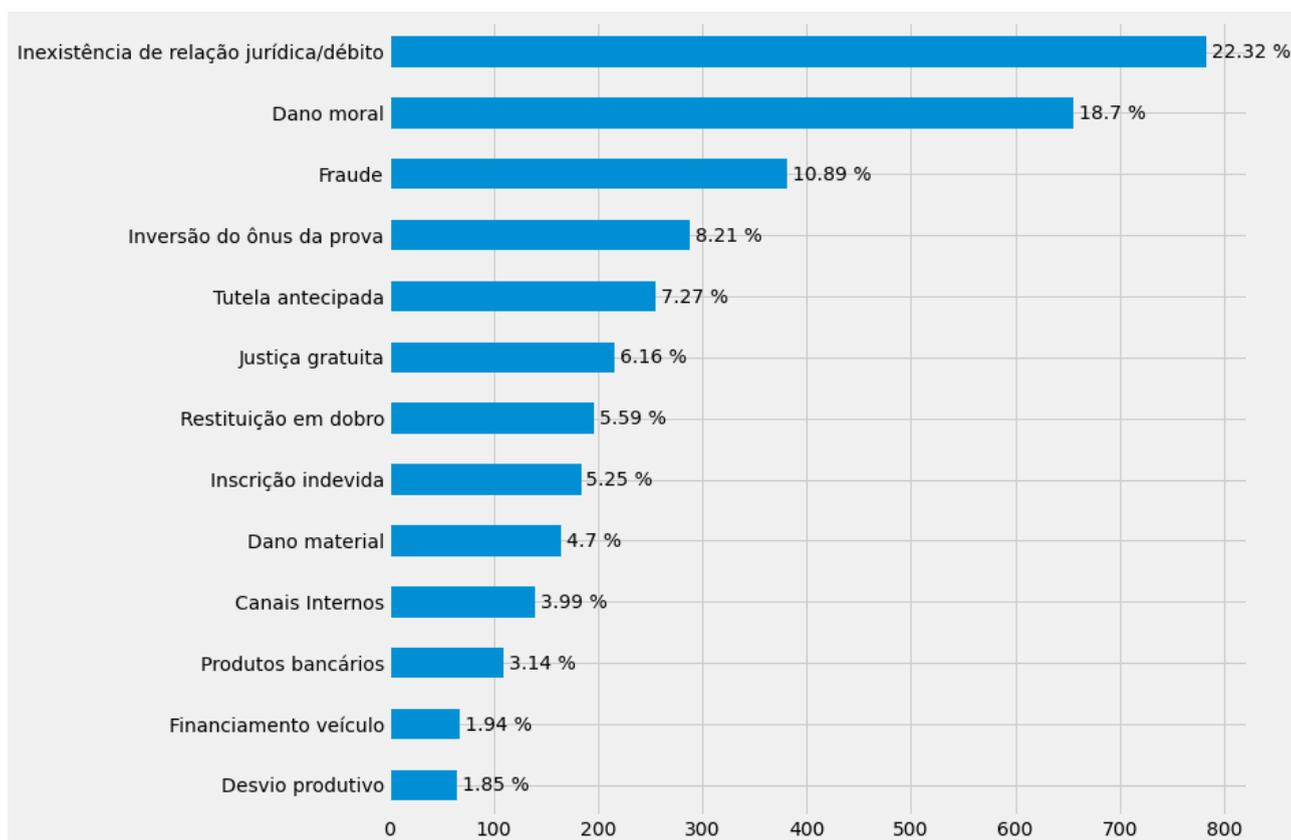
Table 3: Comparison of zero-shot, translation-based and supervised methods (with 2k training examples) on all domains of CLS. MT-BOW and CL-SCL results are from (Zhou et al., 2016).

Pode-se perceber que o MultiFiT alcança resultados adequados ao estado-da-arte para classificação supervisionada, que é o caso do presente projeto, em diversas línguas que não o inglês, apontando um grande benefício para um cenário de baixa disponibilidade de dados, como pode ser o caso do TCU, mesmo com a extração de parágrafos das peças processuais.

O próprio TCU já fez uso desta rota tecnológica, atingindo resultados relevantes na classificação dos excertos de acórdãos, cenário que passa a ser semelhante ao da presente ETEC, em razão da segmentação das peças processuais em parágrafos.

Em relação à maturidade tecnológica, a **COREJUR em conjunto com a UE CEIA/UFG desenvolveu uma solução de classificação de textos idêntica à proposta neste projeto de P&D**, apenas se diferenciando no tema das peças processuais, que no caso é o tema bancário. Porém, a **COREJUR** e a **UE CEIA/UFG**, em conjunto, já desenvolveram modelos de classificação para parágrafos de peças processuais, alcançando resultados sólidos.

A **COREJUR** e a **UE CEIA/UFG**, em conjunto, produziram um *dataset* anotado contendo 3.508 parágrafos de petições iniciais cíveis bancárias. O *dataset* é dividido em 13 classes desbalanceadas. O *dataset* de treino foi composto por 2.806 exemplos, enquanto os *datasets* de validação e teste possuem 351 exemplos. A distribuição do *dataset* é a seguinte:



Os resultados preliminares de modelos treinados a partir do BERT, já com *fine-tuning* com petições para aprendizado do domínio específico, demonstram uma acurácia global de 87% para o dataset de teste descrito acima:

	precision	recall	f1-score	support
Canais Internos	0.91	0.77	0.83	13
Dano material	0.81	0.81	0.81	16
Dano moral	0.93	0.95	0.94	65
Desvio produtivo	1.00	0.88	0.93	8
Financiamento veículo	0.86	1.00	0.92	6
Fraude	0.86	0.83	0.84	46
Inexistência de relação jurídica/débito	0.78	0.86	0.82	74
Inscrição indevida	0.79	0.95	0.86	20
Inversão do ônus da prova	1.00	0.90	0.95	29
Justiça gratuita	0.95	1.00	0.98	21
Produtos bancários	0.62	0.36	0.45	14
Restituição em dobro	0.93	0.87	0.90	15
Tutela antecipada	0.96	0.92	0.94	24
accuracy			0.87	351
macro avg	0.88	0.85	0.86	351
weighted avg	0.87	0.87	0.87	351

Importante frisar que o experimento apontado acima ocorre em cenário idêntico à solução proposta nesta ETEC, isto é, mediante a segmentação da peça processual em parágrafos e sua posterior classificação nas classes apontadas acima, referentes ao processo civil, na área consumerista.

No mais, apesar da diferença existente entre a classificação simples e o *outcome forecasting* delineada acima, a técnica a ser utilizada é a mesma, apenas com maior incerteza acerca do resultado final, o que não impede, porém, a sua efetivação com as mesmas técnicas já utilizadas e validadas pela **COREJUR** e pela **UE CEIA/UFG** no projeto citado acima.

A solução atualmente se encontra em utilização pela **COREJUR**. Desta maneira, percebe-se que os resultados já obtidos pelas proponentes se mostram extremamente satisfatórios para a solução do problema proposto. Assim, a **COREJUR** e o **CEIA** demonstram experiência específica com o tipo de tarefa proposta neste projeto, indicando a possibilidade de alcançar o TLR 9 com certa facilidade, tendo em vista que já realizam projeto com as mesmas especificações, apenas em outro tema.

Dessa maneira, amplamente demonstrado que ambas as soluções alternativas propostas têm plena capacidade de oferecer resultados atinentes ao estado-da-arte, o que comprova a adequação do presente projeto de P&D ao edital desta ETEC.

### **Etapas de desenvolvimento do projeto**

Etapa 1: **Desenho final da solução** – Definição de diretrizes: Definição do conjunto completo de classes e entidades nomeadas, definição do conjunto de respostas a serem utilizadas como encaminhamentos e análises técnicas, alinhamento e treinamento dos anotadores, definição dos entendimentos do TCU acerca de parágrafos que indiquem a existência das classes;

Etapa 2: **Corpus de dados** – Anotação manual dos parágrafos quanto às classes e entidades nomeadas, seleção final e treinamento dos modelos de linguagem de português jurídico a partir das peças processuais não rotuladas do TCU, com o teste de diversas arquiteturas de modelagem de língua;

Etapa 3: **Pesquisa e desenvolvimento e transferência do conhecimento** – Seleção final e treinamento do(s) modelo(s) de classificação e de reconhecimento de entidades nomeadas a partir dos parágrafos das peças processuais anotados. Testagem do modelo de língua para a tarefa de determinação de similaridade entre peças processuais;

Etapa 4: **Construção do *pipeline* de inferência** – Testagem final do pipeline completo – extração de parágrafos, classificação e NER – Teste final com petições reais recebidas pelo TCU para observação da acurácia em produção;

Etapa 5: **Integração com os sistemas do TCU** – *Deployment* dos modelos treinados, a partir das orientações do TCU quanto à melhor forma de integração dos modelos ao sistema já utilizado pelo tribunal de contas.

**Cronograma**

<b>Período</b>	<b>Atividade</b>	<b>Entregável</b>
<b>Etapa 1 - Desenho final da solução</b>		
Sem 1	Versão inicial das classes e entidades – Guidelines de anotação – Validação com especialistas do TCU	Documento de definição das guidelines de anotação
Sem 2	Definição dos encaminhamentos e análises técnicas a serem utilizados como classificação, conforme detalhado neste projeto	Relatório com a lista final de encaminhamentos
Sem 3	Treinamento dos anotadores – Anotações em fase de teste – Medida de concordância entre os anotadores	Relatório da anotação teste - Medida de concordância (kappa)
Sem 4	Treinamento dos anotadores - Segunda rodada de anotação em fase de teste – Medida de concordância entre os anotadores	Relatório da anotação teste 2 - Medida de concordância 2 (kappa)
<b>Etapa 2 - Corpus de dados</b>		
Sem 5	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem, classificador e NER	8,33% dos parágrafos anotados - Modelos v1 (Relatório de performance)
Sem 6	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	16,66% dos parágrafos anotados - Relatório da curadoria - Modelos v2 (Relatório de performance)
Sem 7	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	24,99% dos parágrafos anotados - Relatório da curadoria - Modelos v3 (Relatório de performance)
Sem 8	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	33,32% dos parágrafos anotados - Relatório da curadoria - Modelos v4 (Relatório de performance)

Sem 9	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	41,65% dos parágrafos anotados - Relatório da curadoria - Modelos v5 (Relatório de performance)
Sem 10	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	49,98% dos parágrafos anotados - Relatório da curadoria - Modelos v6 (Relatório de performance)
Sem 11	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	58,31% dos parágrafos anotados - Relatório da curadoria - Modelos v7 (Relatório de performance)
Sem 12	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	66,64% dos parágrafos anotados - Relatório da curadoria - Modelos v8 (Relatório de performance)
Sem 13	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	74,97% dos parágrafos anotados - Relatório da curadoria - Modelos v9 (Relatório de performance)
Sem 14	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	83,3% dos parágrafos anotados - Relatório da curadoria - Modelos v10 (Relatório de performance)
Sem 15	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	91,63% dos parágrafos anotados - Relatório da curadoria - Modelos v11 (Relatório de performance)
Sem 16	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior –	100% dos parágrafos anotados - Relatório da curadoria - Modelos v12 (Relatório de performance) - Dataset

	Treinamento do modelo de linguagem, classificador e NER - FIM DA FASE DE SANEAMENTO	completo <b>MACRO ENTREGA</b>
<b>Etapa 3 - Pesquisa e desenvolvimento</b>		
Sem 17	Pesquisa final, a partir dos resultados obtidos nas semanas anteriores, utilizando o dataset completo.	Relatório de pesquisa referente aos testes feitos com o dataset completo
Sem 18	Seleção final das técnicas de <i>machine learning</i> que produziram os melhores resultados na pesquisa final e nos testes anteriores, transferência de conhecimento	Entrega final dos modelos de classificação e NER selecionados <b>MACRO ENTREGA</b>
<b>Etapa 4 - Construção do <i>pipeline</i> de inferência</b>		
Sem 19	Testagem final do pipeline completo – extração de parágrafos, classificação e NER – Teste final com petições reais recebidas pelo TCU para observação da acurácia em produção	Relatório detalhando o funcionamento do <i>pipeline</i> , incluindo eventuais falhas, dificuldades e as respectivas soluções
<b>Etapa 5 - Integração com os sistemas do TCU e transferência do conhecimento</b>		
Sem 20	Definição dos métodos de integração para a utilização do modelo treinado pelo TCU – Testes para consumo do modelo com os métodos de integração definidos pelo TCU	Relatório de definição dos métodos de integração a serem utilizados e contendo o resultado dos testes preliminares
Sem 21	Desenvolvimento da plataforma inferência a ser consumida pelo TCU	Relatório de desenvolvimento a partir das necessidades levantadas na semana anterior
Sem 22	Finalização da plataforma de inferência com o modelo treinado para utilização em produção pelo TCU, minicurso para transferência do conhecimento ao TCU	Entrega final da solução para ser utilizada em produção pelo TCU <b>MACRO ENTREGA</b>

## Redação automática de instruções e sumarização de peças processuais – Geração automática de texto

### Solução e aderência ao edital

Seguindo o objeto estabelecido pelo Anexo I do Termo de Referência desta ETEC, passa-se a apontar a solução indicada pelas proponentes para a redação automática de elementos das instruções preliminares e de mérito, que terá método de construção e rota tecnológica similar à tarefa de sumarização dos textos.

De maneira preliminar, importante frisar que em relação à redação automática de parágrafos das peças processuais, a tarefa de classificação pode ser suficiente para a solução do problema, porém, necessitará da construção de respostas estruturadas para a possibilidade de que se passe à redação do parágrafo propriamente dito a partir da categorização dada pelo modelo.

Nesse caso, o modelo não gera o parágrafo, mas apenas aponta a classe a qual o parágrafo pertence e, a partir desta classificação, seleciona o parágrafo correspondente, que já deve estar estruturado anteriormente. Já esta solução, como se passará a apresentar, pretende que o modelo gere o texto automaticamente, sem necessidade da existência de uma resposta estruturada previamente.

Inicialmente, quanto à tarefa de geração automática de textos, a solução por meio da geração de texto pode ser definida como um modelo de linguagem capaz de prever a próxima palavra, dadas as palavras anteriores, tornando possível a construção de parágrafos das instruções a partir dos parágrafos das peças processuais anteriores.

Formalmente, a tarefa a ser realizada pode ser denominada de *Recognizing Textual Entailment* – (RTE)<sup>20</sup>, na qual o modelo é exposto a pares de sentenças, entre as quais há uma relação de consequência, ou inferência. Deve-se frisar que a técnica a ser utilizada é um RTE generativo, ou seja, não apenas classifica os pares de sentenças em “consequência”, “contradição” e “neutro”, mas sim uma técnica de *sequence-to-sequence* no qual o texto da sentença resultante deve ser gerado pelo modelo<sup>21</sup>.

O processo de treinamento, portanto, se dará mediante a ligação dos parágrafos relevantes da peça antecedente com os parágrafos relevantes da peça consequente, um a um. Importante frisar que alguns parágrafos podem não estar relacionados a nenhum parágrafo da peça consequente, bem como vários parágrafos da peça antecedente podem ser ligados a um mesmo parágrafo da peça consequente.

Nesse sentido, a anotação será realizada mediante o upload conjunto dos pares de peças processuais, por exemplo, petição inicial e instrução preliminar correspondente. Na ferramenta de

---

<sup>20</sup>POLIAK, Adam. A survey on recognizing textual entailment as an NLP evaluation. arXiv preprint arXiv:2010.03061, 2020.

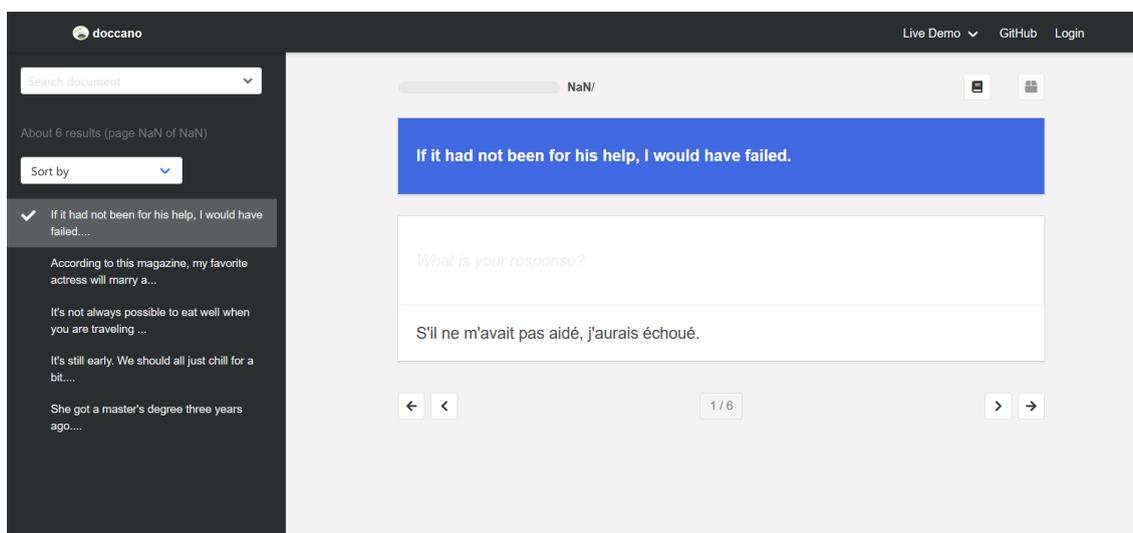
<sup>21</sup>KOLESNYK, Vladyslav; ROCKTÄSCHEL, Tim; RIEDEL, Sebastian. Generating natural language inference chains. arXiv preprint arXiv:1606.01404, 2016.

anotação escolhida, será feita a ligação dos parágrafos da peça antecedente aos parágrafos inseridos na petição consequente, observando uma relação de consequência, isto é, os parágrafos da peça consequente devem ser ligados aos parágrafos da peça antecedente que justificam a existência do referido parágrafo na peça processual consequente.

Assim, o *dataset* a ser gerado pelo processo de anotação manual é um conjunto de pares de parágrafos, um a um, que serão utilizados para o treinamento do modelo. Essencialmente, o modelo será treinado para maximizar a chance de que o output gerado automaticamente seja igual ao anotado no dataset. Desta maneira é possível treinar o modelo para que gere parágrafos consistentes e semelhantes aos gerados pelos especialistas do TCU:

Peça antecedente - Input	Peça consequente – Label
Parágrafo A	Parágrafo B
Parágrafo C	Parágrafo D

Quanto à forma de anotação, deverá ser utilizada uma ferramenta adequada à tarefa de *sequence-to-sequence*, por meio da qual os anotadores indicarão, em pares, os parágrafos das peças antecedentes e sua consequência da peça processual consequente. A seguir, um exemplo de ferramenta que pode ser utilizada para a anotação proposta nesta solução:



Exemplos similares de *dataset* são tipicamente utilizados em tarefas de *sequence-to-sequence* sendo a forma mais comum desta tarefa a de *machine translation*, como apontado acima, que também trabalha com *datasets* em que se tem uma sequência de tokens como entrada e como saída do modelo:

German	English
Ich befürworte die Feststellung des Berichterstatters, dass der Weltraum nicht b	I support the rapporteur's stipulation that space should not become weaponised.
Ogleich sie wohlwollend gemeint sind, werden die vorgeschlagenen Änderungen auf	Although meant well, the proposed amendments will, in my view, in fact, hamper t
Ich komme zum Schluß, Herr Präsident. Ich möchte sagen, daß das Recht, daß das G	To conclude, Mr President, I want to say that the law represents or should repre
Ich denke jedoch wirklich, dass wir in den nächsten paar Jahren erleben, dass si	I do think, however, that in the next few years, we will see them abandoning the
Deshalb sehe ich keinen Zusammenhang zwischen der Stilllegung von Ignalina und d	I therefore see no link between the closure of Ignalina and security of supply i

A sumarização, por sua vez, se dá como um caso especial neste tipo de anotação, que se dá quando vários parágrafos da peça antecedente geram como consequência apenas um parágrafo da peça consequente:

Peça antecedente - Input	Peça consequente – Label
Parágrafo A	Parágrafo D
Parágrafo B	Parágrafo D
Parágrafo C	Parágrafo D

Desta forma, esta solução é adequada para ambas as tarefas, isto é, para a geração automática de texto e para a sumarização de peças processuais. Caso a solução de geração automática de textos não seja necessária, em razão de eventual sucesso da solução de classificação somada à construção de respostas previamente estruturadas, esta solução ainda serve como proposta para a tarefa de sumarização, porém, apenas devem ser anotados os casos específicos de vários parágrafos da peça antecedente tendo como consequência o mesmo parágrafo da peça consequente, conforme a tabela exemplificativa acima.

Frise-se que a utilização de vários outputs para um mesmo input pode acarretar a geração repetida de parágrafos, ou na geração de parágrafos muito similares, o que pode ser corrigido por uma filtragem baseada nos *embeddings* dos parágrafos gerados, o que tornaria possível a remoção de parágrafos idênticos ou similares a outros já gerados para uma dada peça processual.

Desta feita, demonstrada a técnica de anotação e geração do *dataset* a ser utilizado no treinamento do modelo de linguagem, e a sua adequação ao problema proposto pelo TCU nesta ETEC, passa-se à rota tecnológica referente à tarefa de geração automática de textos.

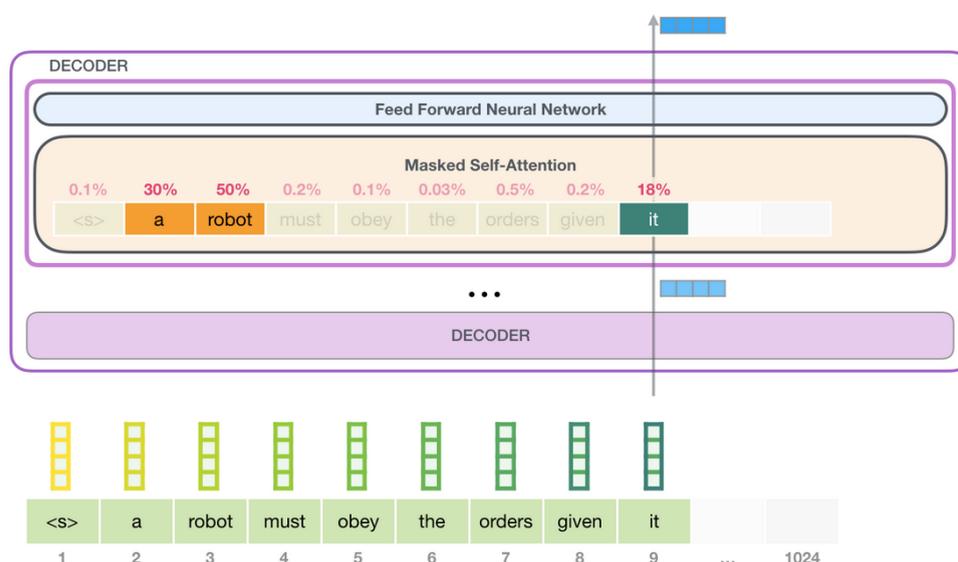
### Rota tecnológica

No que diz respeito à trilha tecnológica, deve-se, novamente, observar o estado-da-arte quanto às tarefas de sumarização e geração automática de textos. Da mesma maneira que a tarefa

anterior, a experimentação e análise empírica da performance dos modelos utilizados nas soluções propostas ao TCU fazem parte deste projeto de P&D. Nesta seção, apresenta-se uma rota viável para a realização da tarefa de geração automática de textos e sumarização generativa.

Aponta-se, portanto, o *Generative Pre-Training Transformer*<sup>22</sup>, em sua segunda versão GPT-2<sup>23</sup> enquanto arquitetura candidata à solução do problema proposto nesta ETEC, que também se utiliza da arquitetura *transformer*, porém de maneira diversa em relação ao BERT, já apresentado.

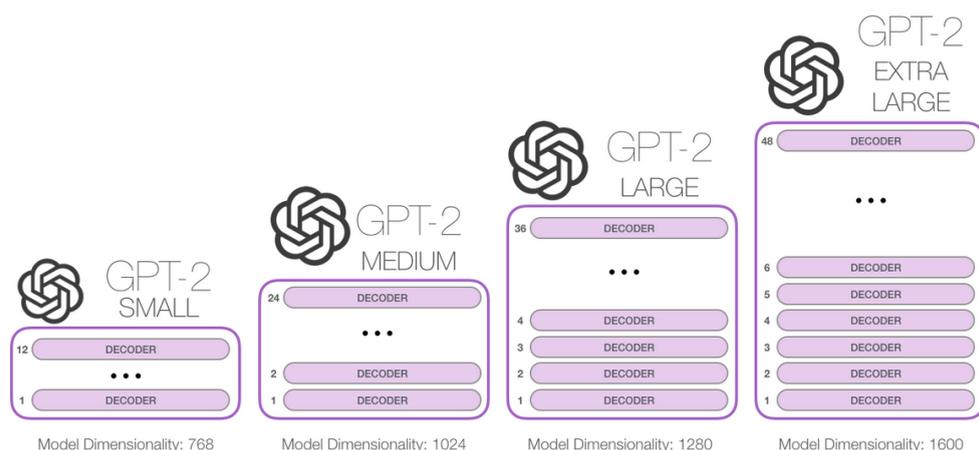
O GPT-2 faz o uso de camadas de *decoder* dos *transformers*, que, de maneira essencial, realizam operações de *multi-headed-attention* seguidas da aplicação de uma rede neural *feedforward* para gerar uma distribuição de probabilidades para cada um dos *tokens* do vocabulário, permitindo a previsão da próxima palavra:



Desta maneira, o GPT-2 é composto de um conjunto de camadas de *decoders* baseados na arquitetura *transformers*, sendo realizada a implementação com 12, 24, 36 e 48 camadas para os modelos *small*, *medium*, *large* e *extra-large*, respectivamente. Importante ressaltar que apenas se encontram versões *small* para o português, que, caso não seja suficiente, levará à necessidade de um treinamento do zero ou da utilização do modelo base em língua diversa, já o que o modelo é agnóstico quanto ao idioma:

<sup>22</sup> RADFORD, Alec et al. Improving language understanding by generative pre-training. 2018.

<sup>23</sup> RADFORD, Alec et al. Language models are unsupervised multitask learners. OpenAI blog, v. 1, n. 8, p. 9, 2019.



Utilizando a arquitetura apontada acima, o GPT-2 atinge o estado-da-arte em diversas tarefas, mesmo sem a utilização de qualquer treino específico, sendo apenas treinado enquanto modelo de linguagem<sup>24</sup>. Especificamente, a performance do GPT-2 foi medida em relação à habilidade para modelar dependências distantes em um texto (LAMBADA<sup>25</sup>), *part-of-speech tagging* (CBT-CN<sup>26</sup>) entidades nomeadas (CBT-NE<sup>27</sup>), modelagem de língua (WikiText<sup>28</sup>), entre outros, superando o estado da arte em sete das oito tarefas, sem qualquer treinamento específico para tanto, tendo apenas o treinamento com cerca de 40 GB de textos não rotulados:

	LAMBADA (PPL)	LAMBADA (ACC)	CBT-CN (ACC)	CBT-NE (ACC)	WikiText2 (PPL)	PTB (PPL)	enwik8 (BPB)	text8 (BPC)	WikiText103 (PPL)	1BW (PPL)
SOTA	99.8	56.25	85.7	82.3	39.14	46.54	0.99	1.08	18.3	<b>21.8</b>
117M	<b>35.13</b>	45.99	<b>87.65</b>	<b>83.4</b>	<b>29.41</b>	65.85	1.16	1.17	37.50	75.20
345M	<b>15.60</b>	55.48	<b>92.35</b>	<b>87.1</b>	<b>22.76</b>	47.33	1.01	<b>1.06</b>	26.37	55.72
762M	<b>10.87</b>	<b>60.12</b>	<b>93.45</b>	<b>88.0</b>	<b>19.93</b>	<b>40.31</b>	<b>0.97</b>	<b>1.02</b>	22.05	44.575
1542M	<b>8.63</b>	<b>63.24</b>	<b>93.30</b>	<b>89.05</b>	<b>18.34</b>	<b>35.76</b>	<b>0.93</b>	<b>0.98</b>	<b>17.48</b>	42.16

Por conseguinte, a utilização da arquitetura GPT-2 para a geração automática de parágrafos de peças processuais se mostra extremamente promissora, sendo possível buscar resultados em conformidade com o estado-da-arte na tarefa de modelagem de língua. Importante frisar, ainda, a relevância da segmentação da peça processual em parágrafos, sem a qual seria impossível a utilização do modelo GPT-2, tendo em vista a sua limitação da quantidade de tokens.

Importante mencionar o modelo mais recente da OPENAI, denominado GPT-3<sup>29</sup> que ainda não se encontra disponível de forma *open-source*, mas pode ser utilizado, mediante pagamento calculado a partir da quantidade de *tokens* necessários para o treinamento do modelo. O GPT-3 é

<sup>24</sup> PPL – *Perplexity*, ACC – *Accuracy*

<sup>25</sup> PAPERNO, Denis et al. The LAMBADA dataset: Word prediction requiring a broad discourse context. arXiv preprint arXiv:1606.06031, 2016.

<sup>26</sup> HILL, Felix et al. The goldilocks principle: Reading children's books with explicit memory representations. arXiv preprint arXiv:1511.02301, 2015. s

<sup>27</sup> Ibidem

<sup>28</sup> MERITY, Stephen et al. Pointer sentinel mixture models. arXiv preprint arXiv:1609.07843, 2016.

<sup>29</sup> BROWN, Tom et al. Language models are few-shot learners. Advances in neural information processing systems, v. 33, p. 1877-1901, 2020

constituído da mesma arquitetura do GPT-2, com pequenas modificações, porém, apresenta um modelo cerca de 100 vezes mais parâmetros, passando de cerca de 1.5 bilhão, para 175 bilhões de parâmetros:

Model Name	$n_{\text{params}}$	$n_{\text{layers}}$	$d_{\text{model}}$	$n_{\text{heads}}$	$d_{\text{head}}$	Batch Size	Learning Rate
GPT-3 Small	125M	12	768	12	64	0.5M	$6.0 \times 10^{-4}$
GPT-3 Medium	350M	24	1024	16	64	0.5M	$3.0 \times 10^{-4}$
GPT-3 Large	760M	24	1536	16	96	0.5M	$2.5 \times 10^{-4}$
GPT-3 XL	1.3B	24	2048	24	128	1M	$2.0 \times 10^{-4}$
GPT-3 2.7B	2.7B	32	2560	32	80	1M	$1.6 \times 10^{-4}$
GPT-3 6.7B	6.7B	32	4096	32	128	2M	$1.2 \times 10^{-4}$
GPT-3 13B	13.0B	40	5140	40	128	2M	$1.0 \times 10^{-4}$
GPT-3 175B or “GPT-3”	175.0B	96	12288	96	128	3.2M	$0.6 \times 10^{-4}$

**Table 2.1:** Sizes, architectures, and learning hyper-parameters (batch size in tokens and learning rate) of the models which we trained. All models were trained for a total of 300 billion tokens.

Caso seja de interesse do TCU, a mesma tarefa apontada neste capítulo pode se utilizar do GPT-3, sem maiores alterações, sendo apenas necessário o pagamento para a utilização do modelo. Observando as instruções constantes da API da OPENAI pode-se perceber que o treinamento se daria de maneira idêntica ao proposto neste projeto, a partir de pares de sentenças (parágrafo da peça antecedente, parágrafo da peça consequente):

```

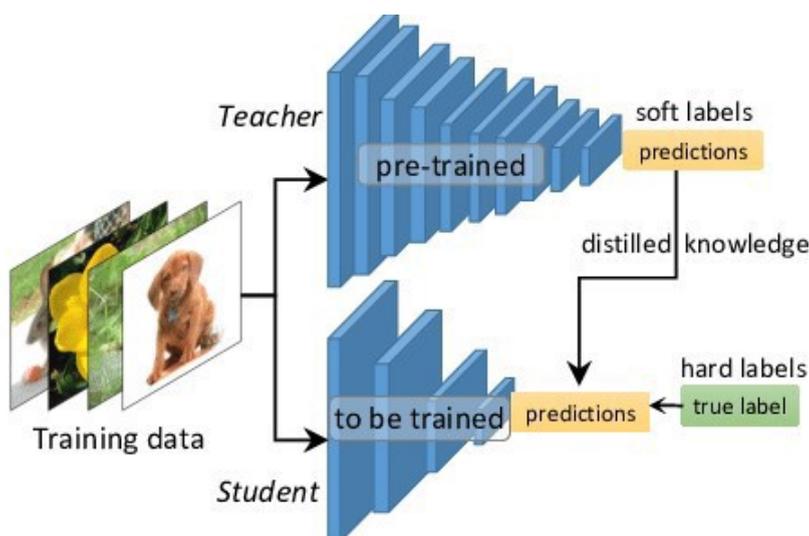
1 {"prompt": "<prompt text>", "completion": "<ideal generated text>"}
2 {"prompt": "<prompt text>", "completion": "<ideal generated text>"}
3 {"prompt": "<prompt text>", "completion": "<ideal generated text>"}
4 ...

```

Deve-se registrar, ainda, que caso o GPT-2 completo se mostre um modelo demasiadamente complexo e pesado, no que diz respeito ao treinamento e à inferência, especialmente se tiver de ser treinado do zero, é possível a utilização da técnica de *knowledge distillation*, que, em síntese, transfere o conhecimento de um modelo de maior tamanho e complexidade para um modelo mais simples, atingindo resultados comparáveis<sup>30</sup>.

A destilação de conhecimento ocorre por meio da utilização do modelo mais robusto como “professor” e o modelo mais simples como “estudante”. Dessa forma os *outputs* do modelo mais robusto são utilizados quando do cálculo da função de custo do modelo mais simples. Desta maneira, o modelo estudante aprende a partir dos dados em conjunção com a previsão do modelo mais robusto, gerando resultados semelhantes a aqueles alcançados pelo modelo professor:

<sup>30</sup>HINTON, Geoffrey et al. Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531, v. 2, n. 7, 2015.



Em resumo, este processo se dá pela utilização do parâmetro de temperatura na função de ativação *softmax*, para a qual é designado um alto valor quando o modelo professor realiza a inferência a ser passada para o modelo estudante, resultando em uma distribuição de probabilidades menos desbalanceada. Dessa forma, as previsões do modelo estudante são comparadas aos dados (*hard labels*) e às previsões do modelo professor (*soft labels*). A função custo do modelo estudante leva em consideração ambas as comparações, e, assim, aprende com o modelo professor, por meio da retropropagação, como usual.

O modelo destilado passa a atingir performances semelhantes ao modelo professor, como mostrado no artigo mencionado, que se utiliza de um dataset de áudio, com cerca de 700 milhões de exemplos, na tarefa de *speech recognition*. Importante registrar também que já existem versões destiladas do GPT-2 para inglês<sup>31</sup>, demonstrando a viabilidade da solução proposta, em caso de necessidade:

System	Test Frame Accuracy	WER
Baseline	58.9%	10.9%
10xEnsemble	61.1%	10.7%
Distilled Single model	60.8%	10.7%

Table 1: Frame classification accuracy and WER showing that the distilled single model performs about as well as the averaged predictions of 10 models that were used to create the soft targets.

Adicionalmente, para buscar uma melhoria do modelo de geração de texto é possível utilizar a classificação do texto realizada nos termos dos capítulos anteriores para auxiliar na geração de textos que seriam vinculados a uma classe específica, o que entra como um *input* no modelo de geração de texto, vinculando a geração do texto à classe indicada como entrada.

<sup>31</sup> <https://huggingface.co/distilgpt2>

A tarefa descrita neste parágrafo é denominada de *controlled text generation*, e pode ser compreendida como uma conjunção no treinamento do modelo de linguagem com um modelo de classificação, na qual o modelo completo será treinado para maximizar a probabilidade da produção das sentenças adequadamente, bem como do pertencimento da sentença gerada a uma determinada categoria<sup>32</sup>, este modelo é denominado de *plug and play language model* - PPLM.

Pode-se compreender o PPLM como uma complementação ao modelo de linguagem, que tenta prever a probabilidade do próximo token, dados os anteriores, com a adição de um classificador, que tenta prever a probabilidade do pertencimento da sequência a uma classe, dados os seus componentes.

No artigo citado utiliza-se do PPLM por meio do GPT-2 em conjunção com um modelo de *bag of words* – BoW, abordagem mais simples, e também em conjunção com um classificador, abordagem mais complexa, para a geração de texto fluente e que, simultaneamente, pertence a um dado tópico, ou classe. No caso do artigo foi utilizada uma combinação entre polaridade (sentimento) e tópicos:

---

[-]	The potato and cauliflower are both in season to make combo breads, mounds, or pads. For an added challenge, try some garlic mashed potatoes.
[Negative]	The potato is a pretty bad idea. It can make you fat, it can cause you to have a terrible immune system, and it can even kill you...
[Positive]	The potato chip recipe you asked for! We love making these, and I've been doing so for years. I've always had a hard time keeping a recipe secret. I think it's the way our kids love to eat them – so many little ones.
[Science]	The potato was once thought to have no health problems and has been promoted as a nutritious food source since the mid-1800s, but recent reports indicate that it has many harmful health issues. In fact, researchers from Johns Hopkins University...
[Politics] [Positive]	To conclude this series of articles, I will present three of the most popular and influential works on this topic. The first article deals with the role of women's political participation in building a political system that is representative of the will of the people.
[Politics] [Negative]	To conclude, the most significant and lasting damage from the economic crisis in 2008 was that many governments, including those in the political center, lost power for the first time in modern history.

---

Nessa toada, a utilização do PPLM pode trazer benefícios à difícil tarefa de geração dos parágrafos da instrução, uma vez que a classificação atribuída ao parágrafo pelos classificadores anteriormente definidos será levada em consideração, permitindo a geração de textos voltadas para o tema do parágrafo analisado na peça antecedente.

Desta feita, pode-se compreender o PPLM como uma ferramenta de combinação de grandes modelos de linguagem pré-treinados com vastas coleções de documentos não rotulados, e modelos de classificação simples:

---

<sup>32</sup>DATHATHRI, Sumanth et al. Plug and play language models: A simple approach to controlled text generation. arXiv preprint arXiv:1912.02164, 2019.

Table 2: Comparison of the different models and distributions. All models in this table are useful in different scenarios. The particular advantage of PPLM is that very small, custom attribute models,  $p(a|x)$ , may be combined with powerful, general pre-trained language models,  $p(x)$ , to create cheap but still powerful conditional generative models,  $p(x|a)$ .

Model type	Form of model	Samples	Example models and number of trainable params
Language Model	$p(x)$	Uncond.	GPT-2 medium: 345M (Radford et al., 2019)
Fine-tuned Language Model	$p(x)$	Uncond.	Fine-tuned GPT-2 medium: 345M (Ziegler et al., 2019)
Conditional Language Model	$p(x a)$	Cond.	CTRL: 1.6B (Keskar et al., 2019)
Plug and Play Language Model (PPLM)	$p(x a) \propto p(x)p(a x)$	Cond.	PPLM-BoW: 0 (curated word list) PPLM-Discrim: $\sim$ 1K/attribute (not counting pretrained $p(x)$ )

O artigo mencionado demonstra a capacidade do método PPLM para a geração de textos com alta precisão em relação ao tópico sobre o qual o texto deve versar. A geração de texto controlada, por meio desta técnica, se mostra uma trilha tecnológica viável para a geração automática de parágrafos, tendo em vista a capacidade de se manter fiel à classificação dada, bem como à fluência do texto gerado. Os resultados a seguir se dão a partir do uso desta técnica, utilizando-se de um modelo de linguagem pré-treinado (GPT-2) em adição a uma rede neural simples de classificação, com apenas uma camada:

Table 6: Evaluation of models/ variants on the sentiment control task, with mean $\pm$ std-dev reported across fluency metrics. Sentiment accuracy reports the fraction of samples with an accurate target sentiment. Approach BCR provides significant control over sentiment while showing minimal degradation in fluency. See Table S9 for full results on individual sentiments. \*GPT2-FT-RL is only evaluated for the positivity half of the task, as it is fine-tuned only for positivity (Ziegler et al., 2019). For human evaluation metrics, we compare the baselines CTRL, GPT2-FT-RL and WD with BCR and perform A/B style testing. We include both numbers for comparison.

Method	Sentiment Acc. (%) (human)	Sentiment Acc. (%) (external classifier)	Perplexity ( $\downarrow$ better)	Dist-1 ( $\uparrow$ better)	Dist-2 ( $\uparrow$ better)	Dist-3 ( $\uparrow$ better)	Human Evaluation Fluency ( $\uparrow$ better)
B	19.3	52.2	42.1 $\pm$ 33.14	0.37	0.75	0.86	3.54 $\pm$ 1.08
BR	41.5	62.2	44.6 $\pm$ 34.72	0.37	0.76	0.87	3.65 $\pm$ 1.07
BC	39.6	64.4	41.8 $\pm$ 34.87	0.33	0.70	0.86	2.79 $\pm$ 1.17
BCR	<b>73.7</b>	<b>78.8</b>	46.6 $\pm$ 40.24	0.36	0.77	0.91	3.29 $\pm$ 1.07
CTRL	<b>76.7</b>	96.6	37.4 $\pm$ 16.89	0.35	0.78	0.89	3.54 $\pm$ 0.77
BCR	70.0	–	–	–	–	–	3.36 $\pm$ 0.82
GPT2-FT-RL*	13.3	77.8	217.3 $\pm$ 176.4	0.54	0.91	0.94	3.31 $\pm$ 0.84
BCR	<b>84.4</b>	–	–	–	–	–	3.68 $\pm$ 0.83
WD	18.9	52.2	31.7 $\pm$ 28.0	0.33	0.69	0.83	3.67 $\pm$ 0.89
BCR	<b>61.1</b>	–	–	–	–	–	3.75 $\pm$ 0.66

Pode-se perceber que a acurácia de aderência ao tópico controlado (no caso, sentimento) se mostra compatível com o estado-da-arte, enquanto a medida de fluência permanece semelhante à dos modelos nos quais não há controle sobre o tópico.

Importante ressaltar que existem pesquisas no âmbito da geração automática de textos jurídicos, especificamente para a geração de parágrafos de decisões. A diferença essencial na pesquisa apontada a seguir é a utilização dos textos anteriores da decisão para a previsão dos parágrafos subsequentes, enquanto neste projeto, a previsão deverá ser feita a partir do texto da peça

processual antecedente. Porém, a pesquisa configura evidência científica da possibilidade de execução da tarefa proposta<sup>33</sup>.

Os resultados alcançados na pesquisa acima se mostram promissores, uma vez que o modelo GPT-2 treinado para linguagem jurídica foi capaz de produzir textos com o mesmo nível de fluência dos trechos verdadeiros a partir de apenas 50.000 decisões para treinamento, bem como os textos mencionados não foram identificados por humanos como textos gerados automaticamente, que apresentaram uma taxa de acerto de apenas 49% na identificação de parágrafos verdadeiros e gerados pelo modelo:

(a) **Prompt:** Appellant urges that he acted diligently to present his claim to the court and that this diligence should justify the one-day delay. Without reciting all of the details of his efforts we note that the 90-day period for filing a complaint expired on May 14, 1973.  
 (b) **Genuine Ending:** On March 6, 1973, he had sought the services of an attorney to draft a formal complaint for him. The complaint was prepared and delivered to him prior to the filing deadline but the exact date is not clear from the record. Appellant's Memorandum is Support of Motion to Appeal would indicate this date as April 13, 1972.  
 (c) **Transformer-XL:** In this connection, the Nygren's argument is without support. Appellant argues, for example, for a period not exceeding five days after filing an indictment on October 15. Appellee argues, for example for an offense involving his activities under §506, which was committed on October 22 and April 22.  
 (d) **Vanilla GPT-2:** Appellant was able to present his claim in the court in less than 10 days, which is not a trivial amount of time to provide for a review of a case in which the complaint was filed within the 90-day period. Moreover, it is undisputed that appellant did not have any prior knowledge of the nature of the allegations and the nature of the issues raised in the complaint.  
 (e) **Fine-tuned GPT-2 on Legal Text:** The Government's failure to disclose the facts of the case, and the failure of the court to address the underlying factual issues is not at issue here. On November 19, 1973, respondent filed his civil suit in the District Court. The District Court denied the claim for \$ 100 million, arguing that the District Court had not acted to protect his interest in the case, and that there was no constitutional or statutory impediment to enforcing the suit.

Figure 4. Generated text example.

Necessário citar a recente adição de um sumário abstrativo à ferramenta Google Docs<sup>34</sup>, que permite apontar para outra solução semelhante à proposta neste projeto, já em fase de produção. O sumário automático do Google Docs também se utilizou de uma arquitetura de *transformers* para gerar a sumarização abstrativa de documentos constantes da ferramenta Google Docs.

Da mesma forma proposta neste projeto, o modelo é treinado com pares de sentenças, porém, o modelo utilizado pela Google neste caso se utiliza da técnica de *Masked Language Modelling*, diferentemente do GPT-2. Porém, deve-se apontar o resultado como indício da possibilidade de utilizar modelos com a mesma função, bem como mostrar a possibilidade de utilização em produção de modelos semelhantes ao proposto neste projeto. Ademais, o modelo utilizado pela Google se utiliza da técnica da destilação do conhecimento para que as inferências possam ser realizadas por um modelo mais enxuto, fato também relevante para a pesquisa a ser desenvolvida a partir deste projeto, como se apontou acima.

<sup>33</sup>PERIC, Lazar et al. Legal Language Modeling with Transformers. In: Proceedings of the Fourth Workshop on Automated Semantic Analysis of Information in Legal Text (ASAIL 2020) held online in conjunction with the 33rd International Conference on Legal Knowledge and Information Systems (JURIX 2020) December 9, 2020. CEUR-WS, 2020.

<sup>34</sup> <https://ai.googleblog.com/2022/03/auto-generated-summaries-in-google-docs.html>

Nessa toada, a maturidade tecnológica desta solução pode ser considerada atualmente como TRL<sup>35</sup> (Technology Readiness Levels) de nível 3, uma vez que a prova do conceito já existe em publicações acadêmicas, porém a sua validação em ambiente controlado, simulado e relevante ainda estão pendentes, uma vez que a tarefa final de geração de textos jurídicos em português, da maneira solicitada pelo TCU, é inédita no país.

Desta maneira, a partir da descrição da solução dada neste capítulo, é possível perceber que essa se mostra viável para que seja atingido o nível de TRL 9, tendo em vista utilização de métodos já testados em publicações acadêmicas relevantes e que produzem resultados semelhantes aos buscados pelo TCU, conforme apresentado acima, que demonstram a correção da trilha tecnológica e da solução proposta neste projeto, havendo apenas maiores incertezas do grau de precisão do resultado, já equacionada pelo TCU, o que justifica a edição desta ETEC.

Desta feita, demonstrada a rota tecnológica a ser seguida, que se revela novamente na fronteira do conhecimento técnico, e resta demonstrada a adequação da solução proposta ao problema levantado por esta ETEC, bem como sua viabilidade.

### **Etapas de desenvolvimento do projeto**

Etapa 1: **Desenho final da solução** – Definição de diretrizes: Definição das formas de relacionamento entre os parágrafos, dado o entendimento dos especialistas do TCU;

Etapa 2: **Corpus de dados** – Anotação manual dos pares de parágrafos de acordo com os correspondentes pares de peças processuais;

Etapa 3: **Pesquisa e desenvolvimento** – Seleção final e treinamento dos modelos de geração automática de texto (*sequence-to-sequence*), teste final do PPLM como modelo de linguagem condicional;

Etapa 4: **Construção do *pipeline* de inferência** – Testagem final do pipeline completo – Modelo de linguagem - Geração de texto e sumarização – Teste final com petições reais recebidas pelo TCU para observação da acurácia em produção;

Etapa 5: **Integração com os sistemas do TCU e transferência do conhecimento - *Deployment*** dos modelos treinados, a partir das orientações do TCU quanto à melhor forma de integração dos modelos ao sistema já utilizado pelo tribunal de contas, transferência do conhecimento ao TCU.

---

<sup>35</sup> Technology Readiness Levels (TRL) é utilizado para apoiar gestores na tomada de decisões relativas aos investimentos e esforços por etapa de desenvolvimento, visando a transição dos graus de maturidade da tecnologia. Pode-se utilizar o TRL como uma ferramenta para gerenciar o progresso da atividade de pesquisa e desenvolvimento de um projeto, programa, produto ou processo. Através do TRL é possível definir a maturidade tecnológica de diferentes elementos.

## Cronograma

Período	Atividade	Entregável
<b>Etapa 1 - Desenho final da solução</b>		
Sem 1	Definição das diretrizes de anotação em conjunto com os especialistas do TCU	Guidelines de anotação contendo as diretrizes dos especialistas da área fim TCU
Sem 2	Treinamento dos anotadores – Primeira rodada de anotação em fase de teste – Medida de concordância entre os anotadores	Relatório da anotação teste - Medida de concordância (kappa)
Sem 3	Treinamento dos anotadores – Segunda rodada de anotação em fase de teste – Medida de concordância entre os anotadores	Relatório da anotação teste 2 - Medida de concordância 2 (kappa)
Sem 4	Treinamento dos anotadores - Terceira rodada de anotação em fase de teste – Medida de concordância entre os anotadores	Relatório da anotação teste 3 - Medida de concordância 3 (kappa)
<b>Etapa 2 - Corpus de dados</b>		
Sem 5	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização	8,33% dos parágrafos anotados - Modelos v1 (Relatório de performance)
Sem 6	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização	16,66% dos parágrafos anotados - Relatório da curadoria - Modelos v2 (Relatório de performance)
Sem 7	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização	24,99% dos parágrafos anotados - Relatório da curadoria - Modelos v3 (Relatório de performance)
Sem 8	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização	33,32% dos parágrafos anotados - Relatório da curadoria - Modelos v4 (Relatório de performance)
Sem 9	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização	41,65% dos parágrafos anotados - Relatório da curadoria - Modelos v5 (Relatório de performance)

Sem 10	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização	49,98% dos parágrafos anotados - Relatório da curadoria - Modelos v6 (Relatório de performance)
Sem 11	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização	58,31% dos parágrafos anotados - Relatório da curadoria - Modelos v7 (Relatório de performance)
Sem 12	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização	66,64% dos parágrafos anotados - Relatório da curadoria - Modelos v8 (Relatório de performance)
Sem 13	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização	74,97% dos parágrafos anotados - Relatório da curadoria - Modelos v9 (Relatório de performance)
Sem 14	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização	83,3% dos parágrafos anotados - Relatório da curadoria - Modelos v10 (Relatório de performance)
Sem 15	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização	91,63% dos parágrafos anotados - Relatório da curadoria - Modelos v11 (Relatório de performance)
Sem 16	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização - FIM DA FASE DE SANEAMENTO	100% dos parágrafos anotados - Relatório da curadoria - Modelos v12 (Relatório de performance) - Dataset completo <b>MACRO ENTREGA</b>
<b>Etapa 3 - Pesquisa e desenvolvimento</b>		
Sem 17	Pesquisa final, a partir dos resultados obtidos nas semanas anteriores, utilizando o dataset completo.	Relatório de pesquisa referente aos testes feitos com o dataset completo
Sem 18	Seleção final das técnicas de <i>machine learning</i> que produziram os melhores resultados na pesquisa final e nos testes anteriores, transferência do conhecimento	Entrega final dos modelos de geração de texto/sumarização selecionados <b>MACRO ENTREGA</b>
<b>Etapa 4 - Construção do <i>pipeline</i> de inferência</b>		
Sem 19	Testagem final do pipeline completo – Modelo de linguagem - Geração de texto e sumarização – Teste final com petições reais	Relatório detalhando o funcionamento do <i>pipeline</i> , incluindo eventuais falhas,

	recebidas pelo TCU para observação da acurácia em produção	dificuldades e as respectivas soluções
<b>Etapa 5 - Integração com os sistemas do TCU e transferência do conhecimento</b>		
Sem 20	Definição dos métodos de integração para a utilização do modelo treinado pelo TCU – Testes para consumo do modelo com os métodos de integração definidos pelo TCU	Relatório de definição dos métodos de integração a serem utilizados e contendo o resultado dos testes preliminares
Sem 21	Desenvolvimento da plataforma inferência a ser consumida pelo TCU	Relatório de desenvolvimento a partir das necessidades levantadas na semana anterior
Sem 22	Finalização da plataforma de inferência com o modelo treinado para utilização em produção pelo TCU, minicurso para transferência do conhecimento ao TCU	Entrega final da solução para ser utilizada em produção pelo TCU <b>MACRO ENTREGA</b>

### 3) CICLOS TEMÁTICOS

Conforme delimitado no Anexo I do Termo de Referência desta ETEC, no que diz respeito ao segundo ciclo temático, deve-se entender que há um menor nível de estruturação das ferramentas acessórias ao projeto de instrução assistida, especificamente: não há listagem completa dos encaminhamentos possíveis, deve ser elaborado o formulário de petição inicial, devem ser definidas as classes próprias do referido ciclo (em substituição à classificação quanto ao tipo de irregularidade), as classificações propostas para o primeiro e terceiro ciclo deverão ser analisadas, uma a uma, tendo em vista a possível necessidade de adequação das classes.

As classificações, bem como as classes que as compõem, que as proponentes julgaram serem incompatíveis foram indicadas no capítulo referente às classificações, e serão substituídas ou removidas na fase de definição de diretrizes, em conjunto com os especialistas do TCU.

Já em relação ao terceiro ciclo temático, dada a proximidade dos temas, acredita-se que o mesmo período atribuído ao primeiro ciclo para definições iniciais e diretrizes de anotação, já será possível sanar eventuais alterações, sendo mantido o mesmo cronograma proposto para o primeiro ciclo temático.

Deve-se ressaltar, ainda, que a partir da finalização do primeiro ciclo, é possível fazer tentativas de *transfer learning*<sup>36</sup>, isto é, utilizar o primeiro modelo treinado para produzir as

<sup>36</sup> SHI, Zhiyuan; SIVA, Parthipan; XIANG, Tao. Transfer learning by ranking for weakly supervised object annotation. arXiv preprint arXiv:1705.00873, 2017.

anotações iniciais dos parágrafos dos ciclos subsequentes<sup>37</sup>. Acredita-se que essa abordagem poderá acelerar especialmente o terceiro ciclo temático, tendo em vista a proximidade ao primeiro ciclo. Já o segundo ciclo temático provavelmente apenas será beneficiado pelas técnicas de *transfer learning* para eventuais classes que se repitam.

Nesse sentido, apenas o cronograma apresentado para o segundo ciclo temático será dilatado em duas semanas, a serem utilizadas no início do ciclo temático, para que haja tempo hábil para o desenvolvimento das ferramentas acessórias à instrução assistida, conforme mencionado acima. No mais, o cronograma é igual ao referente às classificações e reconhecimento de entidades nomeadas para o ciclo temático de aquisições públicas.

### Cronograma do segundo ciclo temático

Período	Atividade	Entregável
<b>Etapa 1 - Desenho final da solução</b>		
Sem 1	Desenvolvimento do formulário de petição inicial e definição final das classificações a serem realizadas e entidades a serem extraídas	Formulário de petição inicial e documento de definição final das classificações e entidades nomeadas
Sem 2	Listagem exaustiva dos encaminhamentos e análises técnicas possíveis para sugestão automática do módulo de IA	Atualização do documento de definição das classificações incluindo os encaminhamentos e análises técnicas
Sem 3	Versão inicial das classes e entidades – Guidelines de anotação – Validação com especialistas do TCU	Documento de definição das guidelines de anotação
Sem 4	Definição dos encaminhamentos e análises técnicas a serem utilizados como classificação, conforme detalhado neste projeto	Relatório com a lista final de encaminhamentos
Sem 5	Treinamento dos anotadores – Anotações em fase de teste – Medida de concordância entre os anotadores	Relatório da anotação teste - Medida de concordância (kappa)
Sem 6	Treinamento dos anotadores - Segunda rodada de anotação em fase de teste – Medida de concordância entre os anotadores	Relatório da anotação teste 2 - Medida de concordância 2 (kappa)

<sup>37</sup> MOZAFARI, Marzieh; FARAHBAKHS, Reza; CRESPI, Noel. A BERT-based transfer learning approach for hate speech detection in online social media. In: International Conference on Complex Networks and Their Applications. Springer, Cham, 2019. p. 928-940.

<b>Etapa 2 - Corpus de dados</b>		
Sem 7	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem, classificador e NER	8,33% dos parágrafos anotados - Modelos v1 (Relatório de performance)
Sem 8	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	16,66% dos parágrafos anotados - Relatório da curadoria - Modelos v2 (Relatório de performance)
Sem 9	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	24,99% dos parágrafos anotados - Relatório da curadoria - Modelos v3 (Relatório de performance)
Sem 10	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	33,32% dos parágrafos anotados - Relatório da curadoria - Modelos v4 (Relatório de performance)
Sem 11	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	41,65% dos parágrafos anotados - Relatório da curadoria - Modelos v5 (Relatório de performance)
Sem 12	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	49,98% dos parágrafos anotados - Relatório da curadoria - Modelos v6 (Relatório de performance)
Sem 13	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	58,31% dos parágrafos anotados - Relatório da curadoria - Modelos v7 (Relatório de performance)
Sem 14	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	66,64% dos parágrafos anotados - Relatório da curadoria - Modelos v8 (Relatório de performance)
Sem 15	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	74,97% dos parágrafos anotados - Relatório da curadoria - Modelos v9 (Relatório de performance)

Sem 16	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	83,3% dos parágrafos anotados - Relatório da curadoria - Modelos v10 (Relatório de performance)
Sem 17	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	91,63% dos parágrafos anotados - Relatório da curadoria - Modelos v11 (Relatório de performance)
Sem 18	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER - FIM DA FASE DE SANEAMENTO	100% dos parágrafos anotados - Relatório da curadoria - Modelos v12 (Relatório de performance) - Dataset completo <b>MACRO ENTREGA</b>
<b>Etapa 3 - Pesquisa e desenvolvimento</b>		
Sem 19	Pesquisa final, a partir dos resultados obtidos nas semanas anteriores, utilizando o dataset completo.	Relatório de pesquisa referente aos testes feitos com o dataset completo
Sem 20	Seleção final das técnicas de <i>machine learning</i> que produziram os melhores resultados na pesquisa final e nos testes anteriores	Entrega final dos modelos de classificação e NER selecionados <b>MACRO ENTREGA</b>
<b>Etapa 4 - Construção do <i>pipeline</i> de inferência</b>		
Sem 21	Testagem final do pipeline completo – extração de parágrafos, classificação e NER – Teste final com petições reais recebidas pelo TCU para observação da acurácia em produção, transferência do conhecimento	Relatório detalhando o funcionamento do <i>pipeline</i> , incluindo eventuais falhas, dificuldades e as respectivas soluções
<b>Etapa 5 - Integração com os sistemas do TCU e transferência do conhecimento</b>		
Sem 22	Definição dos métodos de integração para a utilização do modelo treinado pelo TCU – Testes para consumo do modelo com os métodos de integração definidos pelo TCU	Relatório de definição dos métodos de integração a serem utilizados e contendo o resultado dos testes preliminares
Sem 23	Desenvolvimento da plataforma inferência a ser consumida pelo TCU	Relatório de desenvolvimento a partir das necessidades levantadas na semana anterior

Sem 24	Finalização da plataforma de inferência com o modelo treinado para utilização em produção pelo TCU, minicurso para transferência do conhecimento ao TCU	Entrega final da solução para ser utilizada em produção pelo TCU <b>MACRO ENTREGA</b>
--------	---	--

#### 4) DO ESTUDO TÉCNICO PRELIMINAR - ETP

Também se faz necessário apontar e justificar os pontos deste projeto que indicam pequenas inadequações nas sugestões constantes do ETP, uma vez que essas sugestões não se coadunam com a realidade do projeto, uma vez que a complexidade exigida das soluções acaba por ser incompatível com as referidas sugestões.

Em primeiro lugar, o ETP sugere uma remuneração mensal de R\$ 1.000,00 (mil reais) para os anotadores que serão encarregados de rotular os exemplos que servirão de treinamento para os modelos a serem utilizados pelo TCU. Importante ressaltar, que apesar de ser tarefa repetitiva e relativamente maçante, a anotação é a etapa mais relevante do projeto, uma vez que independentemente das rotas tecnológicas apresentadas e dos modelos de IA mais sofisticados, o que dita a qualidade da solução produzida é a qualidade dos dados utilizados no treinamento.

Nesse sentido, o valor de R\$ 1.000,00 (mil reais) se mostra inferior ao necessário para a contratação de anotadores suficientemente qualificados para a rotulagem de uma tarefa extremamente complexa, uma vez que deverá ser pessoa com conhecimentos sobre a atividade jurisdicional do TCU, bem como seus entendimentos e jurisprudência, ademais, o valor sugerido é compatível com a contratação apenas de estagiários de baixa qualificação ou em início da graduação, o que se é um risco muito grande a toda a performance do projeto. A experiência das proponentes em projetos anteriores (mais de 50 projetos em I.A.) corrobora a necessidade do cuidado e devida atenção às etapas de anotação, com a composição de equipe qualificada.

Sugere-se, portanto, o valor de R\$ 2.200,00 (dois mil e duzentos reais) por mês para cada anotador, com o objetivo de formar uma equipe extremamente qualificada de estudantes de Direito das melhores universidades, com os melhores históricos e experiência com a operação do Direito, de preferência, em fase final do curso de Direito e com experiência em escritório de advocacia, ou órgão público relacionado aos temas desta ETEC, como a jurisdição do TCU ou temas afetos ao direito administrativo.

Em segundo lugar, o ETP sugere o prazo de um mês para a rotulagem dos exemplos a serem utilizados no treinamento do modelo. Ocorre que, primeiramente, com a segmentação das peças processuais em parágrafos, o corpus a ser anotado tende a se multiplicar significativamente. Deve-se lembrar também que as anotações a serem feitas para este projeto não são simples, devendo levar em consideração elementos jurídicos, como já mencionado, o que acarreta em um maior dispêndio de tempo na anotação e no treinamento/alinhamento dos anotadores para que alcancem uma meta de concordância aceitável.

Ademais, a anotação para a geração automática de texto é ainda mais complexa, uma vez que o anotador deverá identificar quais parágrafos da peça processual antecedente se ligam a quais parágrafos da peça processual consequente. Novamente com um alto volume de parágrafos a serem anotados.

Nesse sentido, compreendendo que a aceleração demasiada do processo de anotação pode prejudicar de maneira crítica todos os resultados do projeto, as proponentes apontam como suficiente o prazo de três meses de anotação. Apesar desta extensão no período de rotulagem, as proponentes ainda se atêm ao prazo de quatro meses para a finalização da fase de saneamento.

Por fim, as proponentes apontam para a separação entre a fase de pesquisa e desenvolvimento e a construção do corpus de dados. Deve-se compreender que os testes realizados para pesquisa e desenvolvimento de rotas tecnológicas se dá, neste projeto de P&D, de forma iterativa e concomitante à construção do corpus de dados. Isto é, a cada semana de entregas de exemplos rotulados, são realizados treinamentos de modelos para um acompanhamento paulatino das performances das rotas tecnológicas selecionadas.

Dessa maneira, ao fim da fase de construção do corpus de dados, estas proponentes esperam já ter realizado ao menos doze rodadas de testes, das quais espera-se que ao menos as oito finais possuam relevância estatística, isto é, com um volume de dados razoavelmente adequado para a medida de performance dos modelos. Portanto, ao fim da fase de construção do corpus de dados, já haverá conhecimento suficiente para finalizar a pesquisa e desenvolvimento em duas semanas após a entrega do dataset totalmente rotulado.

## **5) DA METODOLOGIA DE GESTÃO DO PROJETO**

O projeto será executado sob os conceitos de gestão baseados na metodologia SCRUM. Serão realizados pequenos ciclos de atividades organizados em sprints semanais, com reunião de alinhamento no início e reunião de fechamento no último dia de cada semana ao longo de todo o projeto. Os ciclos farão parte do grupo maior de entregas previstos como MACRO ENTREGA

Ao fim de cada ciclo de MACRO ENTREGA prevista no cronograma serão apresentadas as evidências do trabalho realizado.

## **6) PROPOSTA**

Demonstradas todas as soluções e como as rotas tecnológicas a serem seguidas para cada uma das soluções, percebe-se claramente que este projeto de P&D atende a todos os critérios de seleção determinados pelo TCU para avaliação dos projetos enviados para solução da ETEC.

As proponentes apresentam o valor total para desenvolvimento do presente projeto, consoante o detalhamento previsto nos anexos I e II - **CRONOGRAMA FÍSICO FINANCEIRO** e **PLANILHA DE FORMAÇÃO DE PREÇO** respectivamente.

DESCRIÇÃO	VALOR
Projeto de P&D de Encomenda Tecnológica de Módulo de Instrução Assistida Por Inteligência Artificial	
<b>SUSTENTAÇÃO</b>	
Suporte técnico para tratamento de incidentes (sustentação) e de Sprints de Manutenção Evolutiva	
<b>VALOR TOTAL DO PROJETO</b>	

## 7) CRONOGRAMA

Dá-se ao presente projeto o prazo total de 38 (trinta e oito) meses.

## 8) CONCLUSÃO

Demonstradas todas as soluções propostas, bem como as rotas tecnológicas a serem seguidas para cada uma das soluções, percebe-se claramente que este projeto de P&D atende a todos os critérios de seleção determinados pelo TCU para avaliação dos projetos enviados para solução da ETEC.

Em primeiro lugar, as soluções apresentadas se mostram adequadas à solução dos problemas levantados, uma vez que com a segmentação das peças em parágrafos, classificação e extração de entidades nomeadas, e geração automática de texto, todos os itens levantados pelo TCU enquanto problemas a serem solucionados são atendidos pelas propostas realizadas. Da mesma maneira, todas as rotas tecnológicas sugeridas são capazes de gerar resultados compatíveis com o estado-da-arte, utilizando técnicas de *machine learning* de vanguarda, conforme determinado no Edital desta ETEC.

No que diz respeito à viabilidade das atividades de pesquisa, todas as soluções e rotas tecnológicas propostas neste projeto já foram utilizadas em cenários ao menos similares ao da presente ETEC. No mesmo sentido, os resultados apresentados apontam para a viabilidade das soluções propostas, tendo em vista que seguem, com rigor científico, as publicações que definem métodos já validados para a implementação da solução requerida nesta ETEC

O detalhamento da solução se mostra adequado, uma vez que cada passo do seu desenvolvimento foi listado e analisado detidamente neste projeto. Todas as soluções propostas contam com amplo apoio de pesquisas científicas recentes, com publicações nos mais conceituados periódicos. Ademais, todas as soluções indicadas foram apresentadas juntamente aos resultados já obtidos utilizando a solução, o que demonstra a plena capacidade de detalhamento dos processos a serem realizados, tendo em vista o apontamento de outras pesquisas científicas semelhantes.

Adicionalmente, este projeto também indica especificamente as metodologias de anotação, bem como quais são os *datasets* a serem produzidos.

Por fim, as proponentes que apresentam este projeto já desenvolvem um projeto praticamente idêntico ao proposto neste projeto, sendo a **UE CEIA/UFG** referência nacional em produção de P&D e soluções em inteligência artificial, enquanto a **COREJUR** é referência nacional na automação de documentos e na aplicação de tecnologia ao âmbito jurídico.

Dessa forma, o presente projeto de P&D se mostra adequado aos requisitos técnicos definidos no edital desta ETEC, apontando soluções adequadas aos problemas propostos, com demonstração de sua viabilidade e detalhamento em nível semanal das atividades a serem desenvolvidas.

## 9) REFERÊNCIAS

- BRADBURY, James et al. Quasi-recurrent neural networks. arXiv preprint arXiv:1611.01576, 2016
- BROWN, Tom et al. Language models are few-shot learners. Advances in neural information processing systems, v. 33, p. 1877-1901, 2020
- CHALKIDIS, Ilias et al. Extreme multi-label legal text classification: A case study in EU legislation. arXiv preprint arXiv:1905.10892, 2019.
- CHALKIDIS, Ilias et al. LEGAL-BERT: The muppets straight out of law school. arXiv preprint arXiv:2010.02559, 2020.
- CHALKIDIS, Ilias; ANDROUTSOPOULOS, Ion; ALETRAS, Nikolaos. Neural legal judgment prediction in English. arXiv preprint arXiv:1906.02059, 2019b.
- CHALKIDIS, Ilias; ANDROUTSOPOULOS, Ion; MICHOS, Achilleas. Extracting contract elements. In: Proceedings of the 16th edition of the International Conference on Artificial Intelligence and Law. 2017. p. 19-28.
- DATHATHRI, Sumanth et al. Plug and play language models: A simple approach to controlled text generation. arXiv preprint arXiv:1912.02164, 2019.
- DEVLIN, Jacob et al. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.
- EISENSCHLOS, Julian Martin et al. MultiFiT: Efficient multi-lingual language model fine-tuning. arXiv preprint arXiv:1909.04761, 2019.
- HE, Kaiming et al. Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. 2017. p. 2961-2969.
- HILL, Felix et al. The goldilocks principle: Reading children's books with explicit memory representations. arXiv preprint arXiv:1511.02301, 2015.
- HINTON, Geoffrey et al. Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531, v. 2, n. 7, 2015.
- IN, Tsung-Yi et al. Microsoft coco: Common objects in context. In: European conference on computer vision. Springer, Cham, 2014. p. 740-755.
- KOLESNYK, Vladyslav; ROCKTÄSCHEL, Tim; RIEDEL, Sebastian. Generating natural language inference chains. arXiv preprint arXiv:1606.01404, 2016.
- LONG, Jonathan; SHELHAMER, Evan; DARRELL, Trevor. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 3431-3440.

MEDVEDEVA, Masha; WIELING, Martijn; VOLS, Michel. Rethinking the field of automatic prediction of court decisions. *Artificial Intelligence and Law*, p. 1-18, 2022.

MERITY, Stephen et al. Pointer sentinel mixture models. *arXiv preprint arXiv:1609.07843*, 2016.

MOZAFARI, Marzieh; FARAHBAKHS, Reza; CRESPI, Noel. A BERT-based transfer learning approach for hate speech detection in online social media. In: *International Conference on Complex Networks and Their Applications*. Springer, Cham, 2019. p. 928-940.

PAPERNO, Denis et al. The LAMBADA dataset: Word prediction requiring a broad discourse context. *arXiv preprint arXiv:1606.06031*, 2016.

PERIC, Lazar et al. Legal Language Modeling with Transformers. In: *Proceedings of the Fourth Workshop on Automated Semantic Analysis of Information in Legal Text (ASAIL 2020) held online in conjunction with the 33rd International Conference on Legal Knowledge and Information Systems (JURIX 2020) December 9, 2020*. CEUR-WS, 2020.

POLIAK, Adam. A survey on recognizing textual entailment as an NLP evaluation. *arXiv preprint arXiv:2010.03061*, 2020.

PRETTENHOFER, Peter; STEIN, Benno. Cross-language text classification using structural correspondence learning. In: *Proceedings of the 48th annual meeting of the association for computational linguistics*. 2010. p. 1118-1127

RADFORD, Alec et al. Improving language understanding by generative pre-training. 2018

RADFORD, Alec et al. Language models are unsupervised multitask learners. *OpenAI blog*, v. 1, n. 8, p. 9, 2019.

REN, Shaoqing et al. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, v. 28, 2015.

SAYS, JURI. Prediction System for the European Court of Human Rights. In: *Legal Knowledge and Information Systems: JURIX 2020: The Thirty-third Annual Conference, Brno, Czech Republic, December 9-11, 2020*. IOS Press, 2020. p. 277.

SHI, Zhiyuan; SIVA, Parthipan; XIANG, Tao. Transfer learning by ranking for weakly supervised object annotation. *arXiv preprint arXiv:1705.00873*, 2017.

SOCHER, Richard et al. Recursive deep models for semantic compositionality over a sentiment treebank. In: *Proceedings of the 2013 conference on empirical methods in natural language processing*. 2013. p. 1631-1642

VASWANI, Ashish et al. Attention is all you need. *Advances in neural information processing systems*, v. 30, 2017.

WARSTADT, Alex; SINGH, Amanpreet; BOWMAN, Samuel R. Neural network acceptability judgments. *Transactions of the Association for Computational Linguistics*, v. 7, p. 625-641, 2019.

**ANEXO I**  
**CRONOGRAMA FÍSICO FINANCEIRO**  
**PROJETO DE PESQUISA E DESENVOLVIMENTO DE ENCOMENDA TECNOLÓGICA DE MÓDULO DE INSTRUÇÃO ASSISTIDA**  
**POR INTELIGÊNCIA ARTIFICIAL**

Período	Atividades	Entregáveis	Valor
<b>DETECÇÃO DE OBJETOS PARA EXTRAÇÃO DOS PARÁGRAFOS</b>			
<b>MÊS 1</b>	Diretrizes, anotação e treinamento do modelo de detecção de objetos	Entrega final do modelo de extração de parágrafos treinado - Extração dos parágrafos necessários à anotação das próximas fases do projeto - Entrega do dataset não rotulado	██████████
<b>PRIMEIRO CICLO TEMÁTICO</b>			
<b>CLASSIFICADORES E RECONHECIMENTO DE ENTIDADES NOMEADAS</b>			
<b>MÊS 2</b>	Diretrizes, definição das análises técnicas e encaminhamentos, treinamento dos anotadores e anotação em fase de teste	Documento de definição das guidelines de anotação, lista final de encaminhamentos possíveis, relatório final de anotação em teste (2 Rodadas)	██████████
<b>MÊS 3</b>	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	33,32% dos parágrafos anotados - Relatórios da curadoria - Modelos em quarta versão de teste (Relatórios de performance)	██████████

<b>MÊS 4</b>	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	66,64% dos parágrafos anotados - Relatórios da curadoria - Modelos em oitava versão de teste (Relatórios de performance)	██████████
<b>MÊS 5</b>	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	100% dos parágrafos anotados - Relatórios da curadoria - Modelos em décima primeira versão de teste (Relatórios de performance) - Dataset anotado completo	██████████
<b>MÊS 6</b>	Definição final do pipeline completo e Desenvolvimento da plataforma inferência a ser consumida pelo TCU	Relatório de pesquisa dos testes feitos com o dataset completo, Entrega final dos modelos de classificação e NER selecionados, Relatório detalhando o funcionamento do pipeline, incluindo eventuais falhas, dificuldades e as respectivas soluções e Relatório de definição dos métodos de integração a serem utilizados e contendo o resultado dos testes preliminares, entrega final da solução para ser utilizada em produção pelo TCU	██████████
<b>GERAÇÃO AUTOMÁTICA DE TEXTO E SUMARIZAÇÃO</b>			
<b>MÊS 7</b>	Diretrizes, definição das análises técnicas e encaminhamentos, treinamento dos anotadores e anotação em fase de teste	Documento de definição das guidelines de anotação, lista final de encaminhamentos possíveis, relatório final de anotação em teste (3 Rodadas)	██████████

<b>MÊS 8</b>	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização	33,32% dos parágrafos anotados - Relatórios da curadoria - Modelos em quarta versão de teste (Relatórios de performance)	
<b>MÊS 9</b>	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização	66,64% dos parágrafos anotados - Relatórios da curadoria - Modelos em oitava versão de teste (Relatórios de performance)	
<b>MÊS 10</b>	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização	100% dos parágrafos anotados - Relatórios da curadoria - Modelos em décima primeira versão de teste (Relatórios de performance) - Dataset anotado completo	
<b>MÊS 11</b>	Pesquisa final com o dataset completo, seleção das técnicas com melhor performance e testagem final do pipeline completo (Extração de parágrafos e geração de texto/sumarização)   Definição dos métodos de integração para a utilização do modelo treinado pelo TCU, desenvolvimento da plataforma inferência a ser consumida pelo TCU	Relatório de pesquisa dos testes feitos com o dataset completo, Entrega final dos modelos de classificação e NER selecionados, Relatório detalhando o funcionamento do pipeline, incluindo eventuais falhas, dificuldades e as respectivas soluções   Relatório de definição dos métodos de integração a serem utilizados e contendo o resultado dos testes preliminares, entrega final da solução para ser utilizada em produção pelo TC	
<b>SEGUNDO CICLO TEMÁTICO</b>			

<b>CLASSIFICADORES E RECONHECIMENTO DE ENTIDADES NOMEADAS</b>			
<b>MÊS 12</b>	Desenvolvimento do formulário de petição inicial e definição final das classificações a serem realizadas e entidades a serem extraídas, listagem exaustiva dos encaminhamentos e análises técnicas possíveis para sugestão automática do módulo de IA	Formulário de petição inicial e documento de definição final das classificações e entidades nomeadas	██████████
<b>MÊS 13</b>	Diretrizes, definição das análises técnicas e encaminhamentos, treinamento dos anotadores e anotação em fase de teste	Documento de definição das guidelines de anotação, lista final de encaminhamentos possíveis, relatório final de anotação em teste (2 Rodadas)	██████████
<b>MÊS 14</b>	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	33,32% dos parágrafos anotados - Relatórios da curadoria - Modelos em quarta versão de teste (Relatórios de performance)	██████████
<b>MÊS 15</b>	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	66,64% dos parágrafos anotados - Relatórios da curadoria - Modelos em oitava versão de teste (Relatórios de performance)	██████████
<b>MÊS 16</b>	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	100% dos parágrafos anotados - Relatórios da curadoria - Modelos em décima primeira versão de teste (Relatórios de performance) - Dataset anotado completo	██████████

<b>MÊS 17</b>	Pesquisa final com o dataset completo, seleção das técnicas com melhor performance e testagem final do pipeline completo (Extração de parágrafos, classificação e entidades nomeadas)   Definição dos métodos de integração para a utilização do modelo treinado pelo TCU, desenvolvimento da plataforma inferência a ser consumida pelo TCU	Relatório de pesquisa dos testes feitos com o dataset completo, Entrega final dos modelos de classificação e NER selecionados, Relatório detalhando o funcionamento do pipeline, incluindo eventuais falhas, dificuldades e as respectivas soluções   Relatório de definição dos métodos de integração a serem utilizados e contendo o resultado dos testes preliminares, entrega final da solução para ser utilizada em produção pelo TCU	██████████
<b>GERAÇÃO AUTOMÁTICA DE TEXTO E SUMARIZAÇÃO</b>			
<b>MÊS 18</b>	Diretrizes, definição das análises técnicas e encaminhamentos, treinamento dos anotadores e anotação em fase de teste	Documento de definição das guidelines de anotação, lista final de encaminhamentos possíveis, relatório final de anotação em teste (3 Rodadas)	██████████
<b>MÊS 19</b>	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização	33,32% dos parágrafos anotados - Relatórios da curadoria - Modelos em quarta versão de teste (Relatórios de performance)	██████████
<b>MÊS 20</b>	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização	66,64% dos parágrafos anotados - Relatórios da curadoria - Modelos em oitava versão de teste (Relatórios de performance)	██████████

<b>MÊS 21</b>	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização	100% dos parágrafos anotados - Relatórios da curadoria - Modelos em décima primeira versão de teste (Relatórios de performance) - Dataset anotado completo	██████████
<b>MÊS 22</b>	Pesquisa final com o dataset completo, seleção das técnicas com melhor performance e testagem final do pipeline completo (Extração de parágrafos e geração de texto/sumarização)   Definição dos métodos de integração para a utilização do modelo treinado pelo TCU, desenvolvimento da plataforma inferência a ser consumida pelo TCU	Relatório de pesquisa dos testes feitos com o dataset completo, Entrega final dos modelos de classificação e NER selecionados, Relatório detalhando o funcionamento do pipeline, incluindo eventuais falhas, dificuldades e as respectivas soluções   Relatório de definição dos métodos de integração a serem utilizados e contendo o resultado dos testes preliminares, entrega final da solução para ser utilizada em produção pelo TCU	██████████
<b>TERCEIRO CICLO TEMÁTICO</b>			
<b>CLASSIFICADORES E RECONHECIMENTO DE ENTIDADES NOMEADAS</b>			
<b>MÊS 23</b>	Diretrizes, definição das análises técnicas e encaminhamentos, treinamento dos anotadores e anotação em fase de teste	Documento de definição das guidelines de anotação, lista final de encaminhamentos possíveis, relatório final de anotação em teste (2 Rodadas)	██████████
<b>MÊS 24</b>	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	33,32% dos parágrafos anotados - Relatórios da curadoria - Modelos em quarta versão de teste (Relatórios de	██████████

		performance)	
<b>MÊS 25</b>	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	66,64% dos parágrafos anotados - Relatórios da curadoria - Modelos em oitava versão de teste (Relatórios de performance)	██████████
<b>MÊS 26</b>	Anotação dos parágrafos em fase de produção – Curadoria dos dados produzidos na semana anterior – Treinamento do modelo de linguagem, classificador e NER	100% dos parágrafos anotados - Relatórios da curadoria - Modelos em décima primeira versão de teste (Relatórios de performance) - Dataset anotado completo	██████████
<b>MÊS 27</b>	Pesquisa final com o dataset completo, seleção das técnicas com melhor performance e testagem final do pipeline completo (Extração de parágrafos, classificação e entidades nomeadas)   Definição dos métodos de integração para a utilização do modelo treinado pelo TCU, desenvolvimento da plataforma inferência a ser consumida pelo TCU	Relatório de pesquisa dos testes feitos com o dataset completo, Entrega final dos modelos de classificação e NER selecionados, Relatório detalhando o funcionamento do pipeline, incluindo eventuais falhas, dificuldades e as respectivas soluções   Relatório de definição dos métodos de integração a serem utilizados e contendo o resultado dos testes preliminares, entrega final da solução para ser utilizada em produção pelo TCU	██████████
<b>GERAÇÃO AUTOMÁTICA DE TEXTO E SUMARIZAÇÃO</b>			
<b>MÊS 28</b>	Diretrizes, definição das análises técnicas e encaminhamentos, treinamento dos anotadores e anotação em fase de teste	Documento de definição das guidelines de anotação, lista final de encaminhamentos possíveis, relatório final de anotação em teste (3 Rodadas)	██████████

<b>MÊS 29</b>	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização	33,32% dos parágrafos anotados - Relatórios da curadoria - Modelos em quarta versão de teste (Relatórios de performance)	██████████
<b>MÊS 30</b>	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização	66,64% dos parágrafos anotados - Relatórios da curadoria - Modelos em oitava versão de teste (Relatórios de performance)	██████████
<b>MÊS 31</b>	Anotação dos parágrafos em fase de produção – Treinamento do modelo de linguagem - Geração de texto e sumarização	100% dos parágrafos anotados - Relatórios da curadoria - Modelos em décima primeira versão de teste (Relatórios de performance) - Dataset anotado completo	██████████
<b>MÊS 32</b>	Pesquisa final com o dataset completo, seleção das técnicas com melhor performance e testagem final do pipeline completo (Extração de parágrafos e geração de texto/sumarização)   Definição dos métodos de integração para a utilização do modelo treinado pelo TCU, Entrega da plataforma inferência a ser consumida pelo TCU	Relatório de pesquisa dos testes feitos com o dataset completo, Entrega final dos modelos de classificação e NER selecionados, Relatório detalhando o funcionamento do pipeline, incluindo eventuais falhas, dificuldades e as respectivas soluções   Relatório de definição dos métodos de integração a serem utilizados e contendo o resultado dos testes preliminares, entrega final da solução para ser utilizada em produção pelo TCU	██████████
<b>SUBTOTAL PROJETO</b>			██████████

<b>SUSTENTAÇÃO</b>		
<b>MÊS 33</b>	Suporte técnico para tratamento de incidentes (sustentação) e de Sprints de Manutenção Evolutiva	
<b>MÊS 34</b>	Suporte técnico para tratamento de incidentes (sustentação) e de Sprints de Manutenção Evolutiva	
<b>MÊS 35</b>	Suporte técnico para tratamento de incidentes (sustentação) e de Sprints de Manutenção Evolutiva	
<b>MÊS 36</b>	Suporte técnico para tratamento de incidentes (sustentação) e de Sprints de Manutenção Evolutiva	
<b>MÊS 37</b>	Suporte técnico para tratamento de incidentes (sustentação) e de Sprints de Manutenção Evolutiva	
<b>MÊS 38</b>	Suporte técnico para tratamento de incidentes (sustentação) e de Sprints de Manutenção Evolutiva	
<b>SUBTOTAL SUSTENTAÇÃO</b>		
<b>TOTAL PROJETO</b>		<b>R</b>

**ANEXO II**  
**PLANILHA DE FORMAÇÃO DE PREÇO**  
**PROJETO DE PESQUISA E DESENVOLVIMENTO DE ENCOMENDA**  
**TECNOLÓGICA DE MÓDULO DE INSTRUÇÃO ASSISTIDA POR**  
**INTELIGÊNCIA ARTIFICIAL**

<b>CUSTOS DE EQUIPE</b>				
<b>ITEM 1 - Profissionais especializados para atuação durante todo o projeto</b>				
Descrição	Unidade de medida	Quantidade	Valor unitário (R\$)	Total
Cientista de dados	Valor Mensal	1		
Analista de desenvolvimento em I.A.	Valor Mensal	2		
<b>Quantidade (Meses)</b>				
<b>Subtotal</b>				R

<b>ITEM 2 - Profissionais especializados para atuação parcial</b>				
Descrição	Unidade de medida	Quantidade	Valor unitário (R\$)	Total
Desenvolvedor JAVA	Valor Mensal	2		
Analista de automação	Valor Mensal	2		
<b>Quantidade (Meses)</b>				
<b>Subtotal</b>				

<b>ITEM 3 - Profissionais de apoio</b>				
Descrição	Unidade de medida	Quantidade	Valor unitário (R\$)	Total
Anotador Pleno	Valor Mensal	16		
<b>Quantidade (Meses)</b>				
<b>Subtotal</b>				

<b>CUSTOS OPERACIONAIS</b>				
<b>ITEM 4 - Infra estrutura</b>				
Descrição do serviço	Unidade de medida	Quantidade	Valor unitário (R\$)	Total
Computação em nuvem	Valor mensal	1		
Manutenção de estrutura física	Valor mensal	1		

<b>Quantidade (Meses)</b>				
<b>Subtotal</b>				
			<b>Total global</b>	
Impostos estimados	Porcentual	1	16%	
Cálculo de lucratividade	Porcentual	1	20%	
<b>Valor Total Global</b>				

## ANEXO III

### PORTFÓLIO DE SOLUÇÕES DAS EMPRESAS CONSORCIADAS PARTICIPANTES DO PROJETO DE PESQUISA E DESENVOLVIMENTO DE ENCOMENDA TECNOLÓGICA DE MÓDULO DE INSTRUÇÃO ASSISTIDA POR INTELIGÊNCIA ARTIFICIAL



A **COREJUR** é a mais avançada **Plataforma de Automação para o Universo Jurídico**. Seu objetivo é promover a Hiperautomação de escritórios, organizações públicas e empresas através de **Automação de Processos (BPM)**, **Automação Documental**, **Robôs autônomos (RPA)** e **Inteligência Artificial**. Cada nível de automação é oferecido em módulos que podem ser integrados entre si e serviços externos.

Presente em mais de 90 organizações de todo o Brasil, a **COREJUR** promove máxima produtividade, eficiência e redução de custos.

Saiba mais

<https://corejur.com.br>

#### CASES



MASCARENHAS BARBOSA ADVOGADOS

#### - Automação Documental com Inteligência Artificial -

##### Descrição

**Automação documental** para geração de mais de 15.000 petições/mês, dentre Contestações, Recursos, Manifestações e demais tipos de peças jurídicas.

Desenvolvimento de **Projeto de Inteligência Artificial** para a classificação de textos da peça inicial e geração da contestação sem intervenção humana, com resultados de 90% de acurácia em relação à avaliação humana.

## Resultados

- Redução de até 95% do tempo de elaboração de peças jurídicas
  - Elaboração automática da contestação através da leitura da inicial pela Inteligência Artificial
  - Geração de mais de 15.000 peças/mês
- 



STF - SUPREMO TRIBUNAL FEDERAL

### - Automação Documental -

#### Descrição

**Automação documental** para a assessoria jurídica (AJU).

Desenvolvimento de Modelos de Documentos Automatizados com a implementação do raciocínio jurídico de alta complexidade na peça.

#### Resultados

- Padronização dos pareceres e análise jurídicas
  - Redução do tempo de elaboração de documentos em mais de 70%.
-



PGE-GO - PROCURADORIA GERAL DO ESTADO DE GOIÁS

**- Automação Documental -**

**Descrição**

**Automação de iniciais e peças processuais** para os processos de execução fiscal.

**Resultados**

- Redução de até 95% do tempo de elaboração de peças jurídicas
- Elaboração automática de petições iniciais e recursos



MP-RJ - MINISTÉRIO PÚBLICO DO RIO DE JANEIRO

**- Inteligência Artificial para leitura e classificação de publicações -**

**Descrição**

**Desenvolvimento de modelo de Inteligência Artificial** para leitura e classificação de intimações judiciais.

**Resultados**

- Máxima produtividade na triagem de intimações e despacho de atividades da promotoria

## UE CEIA/UFG



### Centro de Excelência em Inteligência Artificial

#### Visão geral

O centro de excelência em Inteligência Artificial, sediado na Universidade Federal de Goiás, desenvolve projetos nas áreas de Inteligência Artificial, Ciência de dados e Big Data.

O CEIA atua em: 1) formação e treinamento, 2) consultoria, 3) desenvolvimento de projetos de pesquisa e desenvolvimento (P&D) e 4) co-criação de startups. Nos últimos anos, além de láureas científicas tradicionais como (best paper award, concursos de teses e dissertações, etc), os pesquisadores envolvidos no centro conquistaram importantes reconhecimentos internacionais de empresas e organizações.

#### Reconhecimentos internacionais

O centro possui um amplo portfólio de projetos com atuação em diversos segmentos tais como, saúde, varejo, energia, eletrônicos, logística, dentre outros.

Nas últimas três edições do Challenge de IA da RSNA (Radiological Society of North American) sempre houve um membro do CEIA no pódio, com destaque para a terceira posição geral em 2017 entre mais de 300 competidores.

No campo de NLP (Natural Language Processing) destacamos:

- Winner IberLEF: Iberian Languages Evaluation Forum, Iberian Languages Evaluation Forum.
- Winner Assin2 - Evaluation of Semantic Textual Similarity and Textual Inference in Portuguese (Avaliação de Similaridade Semântica e Inferência Textual), STIL - Symposium in Information and Human Language Technology (Propor 2019)
- Reconhecimento do Allen Institute for Artificial Intelligence (<https://allennlp.org/elmo>), mantido por Paul Allen fundador da Microsoft, que definiu a nossa rede neural como referência de resultado para problemas envolvendo a língua portuguesa.
- Terceiro lugar geral no Pediatric Bone Age Challenge da Sociedade Americana de

Radiologia - RSNA em 2017.

- Duas vezes (2019 e 2021) projeto de destaque (spotlight talks) no principal evento de IA do mundo ( NeurIPS).
- Em 2021, pesquisadores do CEIA conquistaram o primeiro lugar no COUGH Challenge do INTERSPEECH 2021, principal evento de processamento de voz do mundo

Recentemente, um projeto na área de saúde superou os resultados reportados pelo time da Google Health. Atualmente este projeto monitora cerca de 3 milhões de pessoas no Brasil possibilitando uma medicina preventiva eficaz e que reduz entre 40% a 60% os custos com grupos de pacientes crônicos.

	<b>Google Health (hospitals)</b>	<b>CEIA/Deep Learning Brasil (health plan providers)</b>
Data	<b>Clinical data</b>	<b>Financial records</b>
Task	<b>Death prediction</b>	<b>Complication prediction</b>
AUC	92%	94%

Este projeto na área da saúde recebeu em dezembro de 2019 o prêmio de “Champions of Science for Latin American” concedido pela Jhonson & Jhonson Innovation

*Johnson & Johnson* INNOVATION

**CHAMPIONS**  


---

**OF**  


---

**SCIENCE**

### Patrocinadores

O CEIA gerencia uma carteira de projetos de inovação de mais de 40 milhões de reais com cerca de 33 empresas, dos quais destacamos abaixo as principais:

### **Startups spin-off**

O centro possui um relevante portfólio de startups spin-off.



## Proposta de trabalho

O centro propõe o estabelecimento de projeto de pesquisa e inovação que busque promover a competitividade tecnológica da empresa ao mesmo tempo que promove a formação de recursos humanos que poderão ser absorvidos pelas empresas patrocinadoras.

O CEIA é uma unidade Embrapii, com autonomia financeira e administrativa para mitigar custos de projetos de inovação, aportando cerca de 40% do valor financeiro de um investimento em inovação para projetos com empresas que possuam CNAE de atividade econômica de 5 a 33, 62.01-5 e 62.03-1.

### Reportagens jornalísticas do CEIA e seus produtos ou patrocinadores:

<https://g1.globo.com/go/goias/noticia/2019/05/12/estudante-da-ufg-desenvolve-programa-que-preve-piora-no-quadro-de-diabeticos-com-96percent-de-precisao.ghtml>

<https://g1.globo.com/goias/noticia/instituto-de-identificacao-lanca-programa-de-reconhecimento-facial-para-ajudar-a-esclarecer-crimes-em-goias.ghtml>

<http://tribunadoplanalto.com.br/2020/10/14/plataforma-faz-correcao-automatica-de-redacao/>

<https://exame.com/carreira/universidade-vai-oferecer-o-1o-curso-de-inteligencia-artificial-do-brasil/>

<https://www.segs.com.br/seguros/259821-em-iniciativa-inedita-tokio-marine-adota-inteligencia-artificial-na-regulacao-de-sinistros>

[Startup brasileira anuncia a exportação de veículos autônomos a partir de 2022 | TI INSIDE Online](#)

Depoimento de patrocinadores: <https://youtu.be/RKssoSMcYWs>

## ANEXO IV

### **CURRÍCULOS RESUMIDOS DOS PRINCIPAIS PROFISSIONAIS ENVOLVIDOS NO PROJETO DE PESQUISA E DESENVOLVIMENTO DE ENCOMENDA TECNOLÓGICA DE MÓDULO DE INSTRUÇÃO ASSISTIDA POR INTELIGÊNCIA ARTIFICIAL**

UNIÃO DE VALÊNCIAS E EXPERTISES PARA CRIAR A MAIS COMPLETA E PREPARADA EQUIPE PARA A COMPLEXIDADE DO DESAFIO!

Doutores, Mestres e Especialistas com vasta experiência em projetos de mercado em Inteligência Artificial (NLP) para o Universo Jurídico, Automação Documental de Peças Processuais e Automação Avançada de Processos Judiciais, para garantir total aderência e entregas satisfatórias.

#### **ANDERSON DA SILVA SOARES, DR**



Reconhecido como uma das maiores autoridades no desenvolvimento de Inteligência Artificial no Brasil. Fundador e atual coordenador científico do Centro de Excelência em Inteligência Artificial. Fundador e atual coordenador do Bacharelado em Inteligência Artificial da UFG. Representante brasileiro no GPAI (Global Partnership on Artificial Intelligence). Carreira de inovação com sólida formação acadêmica (cum laude) e destacada experiência em projetos *top-tier* com empresas privadas tais como Samsung, Data-H Artificial Intelligence, Americas Health Group, Cyberlabs, iFood Food Delivery, Acordo Certo, Hospital Albert Einstein, Cilia Software, Copel Energy, Falconi Consultants, dentre outros. Membro de conselho de startups de base tecnológica. Professor de Inteligência Artificial da Universidade Federal de Goiás.

**Lattes**

<http://lattes.cnpq.br/1096941114079527>

**Linkedin**

<https://www.linkedin.com/in/inteligenciaartificialbrasil/>

## ARLINDO RODRIGUES GALVÃO FILHO, DR



Researcher Professor in Machine Learning and Deep Learning  
| Universidade Federal de Goiás

Lattes

<http://lattes.cnpq.br/7744765287200890>

Linkedin

<https://www.linkedin.com/in/arlindogalv%C3%A3o/>

Pesquisador Líder do CEIA-UFG (Centro de Excelência em Inteligência Artificial) em Processamento de Linguagem Natural, é professor do Instituto de Informática da Universidade Federal de Goiás. Possui graduação em Engenharia de Computação pela Pontifícia Universidade Católica de Goiás (2007), Mestrado em Engenharia Eletrônica e Computação pelo Instituto Tecnológico de Aeronáutica (2010) e Doutorado em Ciências da Computação pela Universidade Federal de Goiás (2019). Atualmente atua como coordenador do Projeto Classificação de Documentos e Extração Inteligente de Informações em Textos Jurídicos do CEIA-UFG/COREJUR. Possui sólida experiência em inovação obtidas em projetos financiados por empresas privadas como Americas Health Group, Aneel, Petrobras, Natura, dentre outros.

## CLÁUDIA DE ABREU MARTINS E NASCIMENTO



Analista de Automação Legal na COREJUR

**OAB/GO 45.952**

Linkedin

<https://www.linkedin.com/in/claudia-de-abreu/>

Especialista em Automação de Documentos Processuais, Advogada e Analista de Processos Especialista (BPM), realiza a gestão e implementação de Projetos de Automação Jurídica, além de ministrar treinamentos pela COREJUR da ferramenta de automação de documentos e de processos para os escritórios de advocacia;

Analista de processos - BPM.

Responsável pela implantação de Sistemas de Automação Avançada de Documentos Processuais no STF (Fluxo de licitações) e PGE - GO (Fluxo de execução fiscal).

## **DUÍLIO CALAÇA JUNIOR**



Diretor de Produção na Redeempresarial

**Linkedin**

<https://www.linkedin.com/in/du%C3%ADlio-cala%C3%A7a-2830b531/>

Analista de Sistemas, Analista de Processos (MBA) e Gestor de projetos, gerencia a operação de desenvolvedores da equipe COREJUR. Com experiência de mais de 25 anos de desenvolvimento, domina a sintaxe e execução em diversas linguagens, sendo atualmente gestor em projetos JAVA e PHYTON predominantemente.

## **MANOEL VERÍSSIMO DOS SANTOS NETO, MS**



Machine Learning Engineer - CEIA/UFG

**Linkedin**

<https://www.linkedin.com/in/manoel-verissimo/>

Senior Software Engineer | Machine Learning Engineer | Deep Learning Researcher @ Deep Learning Brazil

## MATEUS SOARES JORGE



Analista jurídico - Legal AI na COREJUR

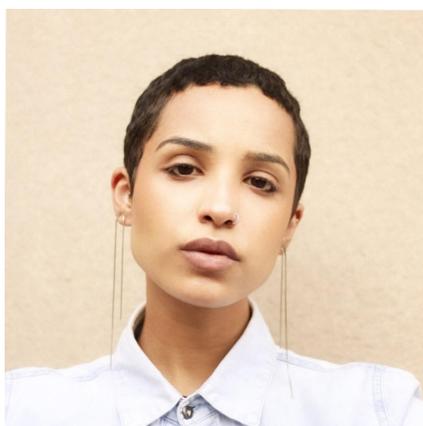
**OAB/DF 67.731**

**Linkedin**

<https://www.linkedin.com/in/mateus-soares-180ab0124>

Advogado, Programador Python e Pesquisador Líder da equipe de anotação do Projeto Classificação de Documentos e Extração Inteligente de Informações em Textos Jurídicos do CEIA-UFG/COREJUR, é responsável pela coordenação, estruturação e execução do processo de anotação de entidades e classes em estruturas documentais complexas.

## MARCELLA HERRANA BACELAR



Analista de Automação na COREJUR

**OAB/GO 59.241**

**Linkedin**

<https://www.linkedin.com/in/marcella-bacelar-812bb1164/>

Advogada, Analista de Processos e Especialista em Automação de Documentos Processuais, desenvolve Modelos Automatizados de Documentos Complexos para escritórios de advocacia nas áreas Cível, Tributária e Trabalhista, além de integrar o corpo de anotadores do Projeto Classificação de Documentos e Extração Inteligente de Informações em Textos Jurídicos.

## ANEXO V

### CARTA DE INTENÇÃO PARA CONSTITUIÇÃO DE CONSÓRCIO

#### 1. DAS PARTES

**REDEEMPRESARIAL SERVIÇOS WEB LTDA**, com sede na Av. Marechal Deodoro, 247, Sobreloja Gal. Pasteur, Pq. Anhanguera, Goiânia, GO, CEP 74340-065 inscrita no CNPJ sob o nº 05.907.755/0001-98, detentora das marcas e plataformas de automação CORE bpa e COREJUR, neste ato representada por RENATO NAVES DE OLIVEIRA SANTOS, portador do CPF nº 704.470.711-49, Sócio da REDEEMPRESARIAL, e

**O CENTRO DE EXCELÊNCIA EM INTELIGÊNCIA ARTIFICIAL - Unidade Embrapii CEIA/UFG**, neste ato representada por TELMA WOERLE DE LIMA SOARES, portadora do CPF nº 296.868.878-22, Diretora executiva da UE CEIA/UFG, vinculada à **UNIVERSIDADE FEDERAL DE GOIÁS - UFG** – instituição federal de ensino e pesquisa de nível superior, constituída como autarquia educacional de regime especial e vinculada ao Ministério da Educação, criada pela Lei n.º 3.834-C, de 14/12/1960, e reestruturada pelo Decreto n.º 63.817, de 16/12/1968, inscrita no CNPJ sob o nº 01.567.601/0001-43, com sede no Campus Samambaia, Goiânia-GO, CEP: 74.001-970, através de seu CENTRO, ambos formalizam, pela presente, a intenção e o compromisso de constituir consórcio, com a finalidade de apresentar, em conjunto, PROJETO DE PESQUISA E DESENVOLVIMENTO DE ENCOMENDA TECNOLÓGICA DE MÓDULO DE INSTRUÇÃO ASSISTIDA POR INTELIGÊNCIA ARTIFICIAL, com a participação efetiva das empresas ora associadas.

Considerando que o Edital de Chamamento Público de processo nº TC 036.798/2020-7, permite a participação de empresas em consórcio para a apresentação conjunta de proposta;

Considerando que as empresas acima qualificadas têm interesse em participar desse processo de seleção em consórcio formado por elas, têm entre si pactuado, e para os fins nele previstos, a presente CARTA DE INTENÇÃO PARA CONSTITUIÇÃO DE CONSÓRCIO, que ajustam segundo as cláusulas e condições adiante dispostas:

#### 2. OBJETO

Pelo presente instrumento particular de CARTA DE INTENÇÃO, as PARTES comprometem-se a se consorciar para participar do processo nº TC 036.798/2020-7, promovido pelo TRIBUNAL DE CONTAS DA UNIÃO, em todas as suas etapas, apresentando proposta, e, caso seja esta adjudicada, a assinar o respectivo CONTRATO, para o que firmarão CONTRATO DE CONSTITUIÇÃO DE CONSÓRCIO, observados os termos do que dispõem as Leis nº 6.404/76 e nº 13.243/2016, comprometendo se a dar cabal cumprimento a todas as obrigações assumidas por força deste instrumento, que celebram em caráter irrevogável e irretratável.

#### 3. DA LIDERANÇA DO CONSÓRCIO

O consórcio usará a denominação de **COREJUR.CEIA**, tendo como líder a empresa

REDEEMPRESARIAL SERVIÇOS WEB LTDA e como representante do consórcio o Sr. Renato Naves de Oliveira Santos, Brasileiro, portador do RG de nº 4017283 DGPC-GO e CPF de nº 704.470.711-49, com plenos poderes para tratar de assuntos técnicos, administrativos, econômico-financeiro e outros julgados de interesse do TRIBUNAL DE CONTAS DA UNIÃO.

#### **4. DA EXCLUSIVIDADE**

As PARTES que compõem o CONSÓRCIO obrigam-se, por este instrumento, a não integrar outro consórcio, nem tampouco participar isoladamente, neste processo licitatório.

#### **5. DA RESPONSABILIDADE SOLIDÁRIA**

As empresas que formam o CONSÓRCIO responderão solidariamente, por todos os atos praticados pelas PARTES, seja durante as fases da licitação ou durante a execução do Contrato, que dela eventualmente decorra.

#### **6. DA INALTERABILIDADE DO AJUSTE**

Declararam as PARTES que não alterarão a constituição ou composição do consórcio sem prévia e expressa anuência do TRIBUNAL DE CONTAS DA UNIÃO, obrigando se a manter sempre presentes as condições que assegurarem a habilitação do CONSÓRCIO, até a conclusão dos serviços a serem contratados, exceto na hipótese de as PARTES virem a se fundir numa só, que as suceda para todos os fins e efeitos legais.

#### **7. DA NATUREZA DO CONSÓRCIO**

Para a proposta apresentada pelo CONSÓRCIO, ajustam as PARTES que a execução dos serviços será distribuída nos termos a seguir apresentados.

#### **8. DO INSTRUMENTO DEFINITIVO**

Caso a proposta apresentada pelo CONSÓRCIO venha a ser adjudicada, obrigam-se as PARTES a promover, no prazo de até 05 (cinco) dias antes da celebração do Contrato, a constituição e o registro do consórcio, cuja duração será, no mínimo, igual ao prazo necessário para a conclusão da encomenda tecnológica objeto da seleção referida, até sua definitiva aceitação, que deverá observar os dispositivos legais aplicáveis, as cláusulas do Edital acima referido e todos os termos deste COMPROMISSO.

#### **9. DO PRAZO DE VIGÊNCIA**

Este Compromisso de Constituição de Consórcio é firmado por prazo indeterminado, vigendo a partir da data de sua assinatura e ficando, automaticamente, rescindido caso ocorra qualquer dos seguintes fatos:

- ser proferida decisão, de que não caiba recurso administrativo ou judicial, de inabilitação do consórcio;
- ser proferida decisão, de que não caiba recurso administrativo ou judicial, de desclassificação do consórcio;
- após esgotados todos os recursos, administrativos e judiciais, na hipótese de adjudicação de proposta ofertada por outro concorrente ou no caso de anulação/revogação do referido Chamamento Público;

- após celebrado e registrado o instrumento de constituição de consórcio a que se refere a Cláusula Sétima, que substituirá este para os fins de direito.

## 10. DO ENDEREÇO

O CONSÓRCIO, para os fins da licitação, adotará como endereço o da LÍDER, situado na Av. Marechal Deodoro, 247, Sobreloja Gal. Pasteur, Pq. Anhanguera, Goiânia, GO, CEP 74340-065.

## 11. DO FORO

Elegem, as PARTES, para dirimir quaisquer litígios decorrentes deste instrumento, o Foro da comarca de Goiânia, Estado de Goiás, com expressa renúncia a qualquer outro, por mais privilegiado que seja.

E por estarem, assim, justas e contratadas as partes firmam este instrumento para um só efeito, juntamente com duas testemunhas a tudo presentes.

Goiânia, 22 de março de 2022

RS  
Assinado de forma digital por  
RENATO NAVES DE OLIVEIRA  
SANTOS:70447071149  
Dados: 2022.03.25 17:47:08  
-03'00'

Assinado de forma digital por  
TELMA WOERLE DE LIMA  
SOARES:29686887822  
Dados: 2022.03.25 17:35:30 -03'00'

REDEEMPRESARIAL SERVIÇOS WEB  
LTDA

Renato Naves de Oliveira Santos

CENTRO DE EXCELÊNCIA EM  
INTELIGÊNCIA ARTIFICIAL - UE  
CEIA/UFG

Telma Woerle de Lima Soares

Testemunhas

Nome:

CPF nº

Nome:

CPF nº