



Processo n. 649.222/19

CONVÊNIO N. 2020/054.3

TERCEIRO TERMO ADITIVO AO CONVÊNIO
CELEBRADO ENTRE A CÂMARA DOS
DEPUTADOS E A UNIVERSIDADE DE SÃO PAULO
PARA PESQUISA DE MÉTODOS EM
APRENDIZAGEM DE MÁQUINA E
PROCESSAMENTO DE LINGUAGEM NATURAL

Ao(s) dias do mês de de 2021, a CÂMARA DOS DEPUTADOS, situada na Praça dos Três Poderes, Brasília/DF, CEP: 70.160-900, CNPJ/MF nº 00.530.352/0001-59, neste ato representada por seu Diretor-Geral, o senhor CELSO DE BARROS CORREIA NETO, brasileiro, residente e domiciliado em Brasília-DF, e a UNIVERSIDADE DE SÃO PAULO, autarquia estadual de regime especial, com sede na Rua da Reitoria, nº 374, Cidade Universitária, São Paulo/SP, CEP: 05.508-220, inscrita no CNPJ sob nº 63.025.530/0001-04, por interesse do INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO, doravante denominado ICMC, sediado na Avenida Trabalhador São-Carlense, 400, Centro, na cidade de São Paulo/SP, CEP: 13.566-590, inscrito no CNPJ sob o nº. 63.025.530/0051-73, neste ato representados por seus representantes legais, signatários, em consonância com seus atos constitutivos, acordam em celebrar o presente Aditivo, em conformidade com o processo em referência, e com a Lei nº 8.666/93, observadas as cláusulas e condições a seguir enunciadas.

O presente aditivo decorre da prorrogação do instrumento por 15 (quinze) meses e 15 (quinze) dias, a partir de 30/12/21, com amparo no inciso II do artigo 57 da Lei nº. 8.666/93.

O novo Plano de Trabalho encontra-se anexado ao Anexo Único do presente instrumento.

O Convênio ora aditado, com sua numeração alterada para 2020/054.3, passa a vigorar com o acréscimo da seguinte cláusula:

“.....”

CLÁUSULA TERCEIRA - RECURSOS FINANCEIROS

Os recursos financeiros necessários para o desenvolvimento do projeto, no valor de R\$ 210.348,00 (duzentos e dez mil, trezentos e quarenta e oito reais), serão depositados em conta indicada pela USP, conforme cronograma de desembolso definido no Plano de Trabalho anexo.

**CLÁUSULA QUARTA - VIGÊNCIA**

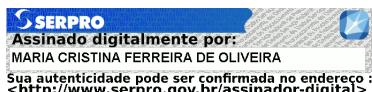
O presente convênio vigorará de 30/12/21 a 13/04/23, prorrogável por iguais ou inferiores períodos. Decorrido o prazo máximo de 05 (cinco) anos, havendo interesse dos partícipes, novo instrumento deverá ser formalizado.

"

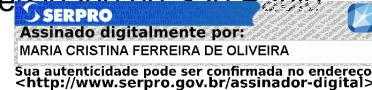
Ficam ratificadas as demais cláusulas e condições vigentes que não tenham sido expressamente modificadas pelo presente Aditivo.

E por estarem assim de acordo, as partes assinam o presente instrumento em 2 (duas) vias de igual teor e forma, para um só efeito, com 2 (duas) páginas cada.

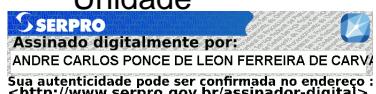
São Paulo-SP, de de 2021.


Assinado digitalmente por:
MARIA CRISTINA FERREIRA DE OLIVEIRA
Sua autenticidade pode ser confirmada no endereço :
<<http://www.serpro.gov.br/assinador-digital>>

Reitor
Universidade de São Paulo


Assinado digitalmente por:
MARIA CRISTINA FERREIRA DE OLIVEIRA
Sua autenticidade pode ser confirmada no endereço :
<<http://www.serpro.gov.br/assinador-digital>>

Diretor
Unidade


Assinado digitalmente por:
ANDRE CARLOS PONCE DE LEON FERREIRA DE CARVALHO
Sua autenticidade pode ser confirmada no endereço :
<<http://www.serpro.gov.br/assinador-digital>>

Coordenador do Projeto

MAURO LIMEIRA MENA Assinado de forma digital por
BARRETO MAURO LIMEIRA MENA
Dados: 2021.12.23 17:53:16 -03'00'

Diretor-Geral
Câmara dos Deputados

Lz



Anexo Único - PLANO DE TRABALHO

Identificação das partes

Unidade da USP: Instituto de Ciências Matemáticas e de Computação

CNPJ: 63.025.530/0051-73

Endereço: Av. Trabalhador São Carlense, 400, São Carlos, SP

Representante Legal: Maria Cristina Ferreira de Oliveira

Instituição Parceira: Câmara dos Deputados

CNPJ: 00.530.352/0001-59

Endereço: Praça dos Três Poderes - Brasília - DF, CEP 70160-900

Representante Legal: Sr. Sergio Sampaio Contreiras de Almeida

Projetos

Tabela 1 - Objetivos de pesquisa

Objetivo 1: aprimorar desempenho de algoritmos existentes

Investigar e implementar solução de processamento de linguagem natural (NLP) capaz de aprimorar o desempenho dos seguintes algoritmos do Programa Ulysses:

- U3 - Algoritmo para pesquisa por estudos técnicos e projetos de leis similares às solicitações recebidas do Gabinete Digital;
- U6 - Algoritmo de análise geral de argumentações sobre temas.

São exemplos de tarefas de NLP alvos da investigação: *Named-entity recognition, query expansion, sentiment analysis* etc.

Após a investigação, os pesquisadores devem modelar e implementar a solução mais promissora, envolvendo uma ou mais tarefas de NLP investigadas, bem como:

- anotação dos *datasets* necessários ao treinamento e validação nos casos de utilização de aprendizado supervisionado;
- definição de métricas para quantificar e otimizar o desempenho dos modelos gerados;
- definição de métricas para quantificar e limitar o viés dos *datasets* anotados e dos modelos gerados, quando aplicável;
- documentar a arquitetura da solução e automatizar o *pipeline* de melhoria contínua dos modelos (*pipeline de retreinamento*);
- pesquisa e proposição de formas adequadas de visualização dos resultados dos modelos gerados.

Objetivo 2: desenvolver algoritmos para comparar e aplicar emendas ao texto legislativo

Investigar e implementar solução de processamento de linguagem natural (NLP) capaz de **reproduzir o efeito da aplicação de uma emenda** (proposição destinada a modificar outra proposição) sobre o texto legislativo da proposição afetada, considerando a localização e alteração dos trechos apropriados. As emendas podem ser supressivas, aglutinativas, substitutivas, modificativas ou aditivas. Além destes cinco tipos, o Regimento Interno da Câmara dos Deputados acrescenta o conceito de subemenda e de emenda de redação [15].



Essa pesquisa visa apoiar o desenvolvimento do seguinte algoritmo do Programa Ulysses:

- U4 - Algoritmo para identificação de impacto das emendas no texto da proposição.

São exemplos de tarefas de NLP alvos da investigação: *Named-entity recognition, sentence similarity* etc.

Após a investigação, os pesquisadores devem modelar e implementar a solução mais promissora, envolvendo uma ou mais tarefas de NLP investigadas, bem como:

- anotação dos *datasets* necessários ao treinamento e validação nos casos de utilização de aprendizado supervisionado;
- definição de métricas para quantificar e otimizar o desempenho dos modelos gerados;
- definição de métricas para quantificar e limitar o viés dos *datasets* anotados e dos modelos gerados, quando aplicável;
- documentar a arquitetura da solução e automatizar o *pipeline* de melhoria contínua dos modelos (*pipeline de retreinamento*);
- pesquisa e proposição de formas adequadas de visualização dos resultados dos modelos gerados.

Objeto

Cooperação entre as partes na execução de Projeto de Pesquisa.

Objetivo

Desenvolver pesquisa para aplicar modernas tecnologias de inteligência artificial e de processamento de linguagem natural no aprimoramento de algoritmos do Programa Ulysses - a Inteligência Artificial da Câmara dos Deputados - priorizados na Estratégia Digital da Câmara dos Deputados, instituída pela [Portaria DG nº 268, de 14 de setembro de 2021](#) [1, 16].

Justificativa

As tarefas de processamento de linguagem natural (NLP) necessárias ao atendimento dos objetivos de pesquisa elencadas na Tabela 1, em linha com a Estratégia Digital da Câmara dos Deputados, representam desafios tratados recentemente pelo meio acadêmico utilizando majoritariamente *corpora* e bases de dados em língua inglesa. Assim, muitos dos resultados das pesquisas publicadas não são diretamente aplicáveis ao contexto da língua portuguesa, sendo necessário investimento em pesquisa e inovação para compilação de grandes conjuntos de dados de qualidade nesta língua, desenvolvimento e validação de métodos capazes de lidar com as peculiaridades linguísticas do português do Brasil.

A inferioridade no desempenho em tarefas NLP em relação ao inglês não é uma exclusividade da língua portuguesa. Apesar do esforço de pesquisa em modelos de linguagem *cross-lingual* [13], o desempenho para língua inglesa ainda é consistentemente superior (Tabela 2). A maioria dos estudos recentes não inclui métricas para a língua portuguesa, mas é seguro inferir que o desempenho para o nosso idioma seria no máximo comparável ao de idiomas como o francês e o espanhol, dada a menor quantidade de conjuntos de dados anotados.

**Tabela 2 - Comparação de desempenho de tarefas NLP para diferentes idiomas**

| | en | fr | es | de | el | bg | ru |
|--|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| <i>Machine translation baselines (TRANSLATE-TRAIN)</i> | | | | | | | |
| Devlin et al. (2018) | 81.9 | - | 77.8 | 75.9 | - | - | - |
| XLM (MLM+TLM) | <u>85.0</u> | 80.2 | 80.8 | <u>80.3</u> | <u>78.1</u> | <u>79.3</u> | <u>78.1</u> |
| <i>Machine translation baselines (TRANSLATE-TEST)</i> | | | | | | | |
| Devlin et al. (2018) | 81.4 | - | 74.9 | 74.4 | - | - | - |
| XLM (MLM+TLM) | <u>85.0</u> | 79.0 | 79.5 | 78.1 | 77.8 | 77.6 | 75.5 |
| <i>Evaluation of cross-lingual sentence encoders</i> | | | | | | | |
| Conneau et al. (2018b) | 73.7 | 67.7 | 68.7 | 67.7 | 68.9 | 67.9 | 65.4 |
| Devlin et al. (2018) | 81.4 | - | 74.3 | 70.5 | - | - | - |
| Artetxe and Schwenk (2018) | 73.9 | 71.9 | 72.9 | 72.6 | 73.1 | 74.2 | 71.5 |
| XLM (MLM) | 83.2 | 76.5 | 76.3 | 74.2 | 73.1 | 74.0 | 73.1 |
| XLM (MLM+TLM) | <u>85.0</u> | <u>78.7</u> | <u>78.9</u> | <u>77.8</u> | <u>76.6</u> | <u>77.4</u> | <u>75.3</u> |

Tabela adaptada de [13].

Em virtude da complexidade inerente às tarefas de NLP e da crescente produção científica que vem modificando radicalmente o panorama do estado da arte nos últimos sete anos, é prudente que a Administração Pública firme parcerias com pesquisadores que estejam acompanhando os avanços da área de NLP, potencializados pelo aprendizado baseado em redes neurais profundas (*deep learning*), com o objetivo de selecionar e aplicar tecnologias que ofereçam melhores resultados, com menores taxas de erro e de viés, no atendimento dos objetivos de pesquisa relacionados na Tabela 1.

Adicionalmente ao elevado nível da pesquisa tecnológica, o presente plano de trabalho se apresenta em um momento em que a Câmara dos Deputados não tem expectativa de crescimento ou reposição de seu quadro pessoal, condição de forte recomendação por inovação tecnológica. Além de ser uma tendência entre as principais organizações do país e objeto de atenção da Estratégia Brasileira de Inteligência Artificial (EBIA) [2], instituída pela Portaria MCTI nº 4.617, de 6 de abril de 2021, o uso estratégico da Inteligência Artificial, por meio do Programa Ulysses, atende os objetivos delineados na Estratégia Digital da Câmara dos Deputados, instituída pela Portaria DG nº 268, de 14 de setembro de 2021.

Ressalte-se que, além de contribuir para o atendimento dos objetivos estratégicos, o presente plano de trabalho proporciona o repasse de conhecimento necessário ao atendimento, pelas equipes internas da Diretoria de Inovação e Tecnologia da Informação (Ditec), de demandas futuras, similares às tratadas pelos objetivos de pesquisa relacionados na Tabela 1.

O plano de trabalho anterior, executado no período de outubro de 2020 até outubro de 2021, proporcionou a pesquisa e o desenvolvimento de algoritmos de inteligência artificial que estão sendo integrados aos serviços digitais da Casa, previstos na Estratégia Digital da Câmara dos Deputados [1, 16], que apoiam os processos de trabalho da Diretoria Executiva de Comunicação e Mídias Digitais (Direx) e da Consultoria Legislativa (Conle). No caso da Direx, os algoritmos auxiliam na análise e mineração de argumentos expressos nos comentários das



enquetes sobre projetos de leis em tramitação na Casa. No caso da Conle, os algoritmos auxiliam a etapa de **pesquisa prévia** sugerindo proposições e solicitações relacionadas ou similares ao conteúdo das solicitações de trabalho. Os algoritmos desenvolvidos são passíveis de melhoria contínua a partir do feedback dos usuários desses serviços digitais, técnica conhecida por *Active Learning*. Com o repasse do conhecimento recebido dos pesquisadores da Universidade, a Ditec pretende manter, evoluir e reutilizar os algoritmos em outras aplicações no âmbito da Câmara dos Deputados.

Resultados esperados

- Entrega, com repasse de conhecimento, de solução baseada em uma ou mais tarefas de NLP capaz de aprimorar o desempenho dos algoritmos do Programa Ulysses apontados na tabela 1;
- Entrega, com repasse de conhecimento, de solução baseada em uma ou mais tarefas de NLP capaz de lidar com a reprodução do efeito resultante da aplicação de emendas ao texto legislativo afetado.

Coordenação e Equipe Técnica

Professor Dr. André Ponce de Leon F. de Carvalho (ICMC/USP)

Patrícia Gomes de Rego de Almeida (Câmara dos Deputados)

Francisco Edmundo de Andrade (Câmara dos Deputados)

Michael Shigeki Onishi (Câmara dos Deputados)



Requisitos gerais

No contexto da cooperação, os **pesquisadores externos deverão**:

- observar as disposições legais sobre o manuseio dos dados coletados, especialmente o que especifica a Lei Geral de Proteção de Dados Pessoais (LGPD) - Lei nº 13.709, de 14 de agosto de 2018;
- apresentar no início dos trabalhos, para fins de validação e aprovação, proposta metodológica e cronograma de entregas; a metodologia deve incluir as tarefas de NLP a serem investigadas, acompanhadas da especificação dos respectivos dados/metadados de treinamento e das métricas de avaliação do desempenho;
- utilizar apenas ferramentas e bibliotecas de código *open source*;
- disponibilizar sob licenciamento permissivo (ex: MIT, Apache, BSD) todo o código dos algoritmos produzidos no âmbito da cooperação, bem como os modelos treinados resultantes;
- documentar e automatizar os passos utilizados na coleta e preparação dos conjuntos de dados utilizados, preferencialmente sob a forma de *notebooks* (ex: *Jupyter Notebook*);
- documentar e automatizar os passos utilizados no treinamento e/ou *fine-tuning* dos modelos e demais resultados das tarefas de NLP;
- apresentar relatórios de progresso em formato digital a cada 15 dias e estarem disponíveis para participar de teleconferência com a equipe da célula de inovação, quando esta requisitar.

Considerando os objetivos de pesquisa identificados na Tabela 1, os pesquisadores externos deverão entregar os códigos de treinamento/retreinamento e de inferência (para uso em produção) juntamente com eventuais *datasets* anotados, conforme o caso. A Diretoria de Inovação e Tecnologia da Informação (Ditec), da Câmara dos Deputados, será responsável pela validação das entregas e pelo desenvolvimento das interfaces destinadas ao usuário final de cada solução entregue.

Objetivo 1: aprimorar desempenho de algoritmos existentes

A solução para o atingimento do objetivo de pesquisa n. 1 deve atender o especificado na Tabela 1. Os algoritmos devem ser implementados em linguagem Python observando-se os requisitos gerais apontados anteriormente. Os algoritmos a serem implantados em ambiente de produção devem ser parametrizáveis e implementados como microsserviços RESTFUL, passíveis de operar em ambientes de container (preferencialmente Docker).

Para o objetivo de pesquisa n. 1, necessita-se alocar pesquisadores externos, sendo no mínimo:

1. Um pesquisador doutor que atenda a pelo menos um dos requisitos enumerados na Tabela 3;
2. Dois estudantes universitários, bolsistas de iniciação científica.

**Tabela 3 - Requisitos mínimos para pesquisador doutor**

| Requisito | Detalhamento |
|---|--|
| Doutorado em Ciência da Computação, Engenharia de Computação, Engenharia Elétrica, Estatística, Física Aplicada ou Matemática Aplicada. | O doutorado deve ter sido feito em instituição brasileira classificada como nível 6 ou 7 pela Capes ou em instituição estrangeira de reconhecida excelência acadêmica. |
| Publicação acadêmica em periódico científico. | Ao menos uma publicação em periódicos científicos classificados como nível A1 pela Capes, nos últimos cinco anos, nas áreas estabelecidas no requisito anterior e, além disso, associado a uma ou mais das seguintes temáticas: Inteligência Artificial, Aprendizado de Máquina (<i>Machine Learning</i>), Aprendizado Profundo (<i>Deep Learning</i>) e Processamento de Linguagem Natural (NLP). |

A Tabela 4 apresenta o cronograma de entregas previstas para o objetivo n. 1.

Tabela 4 - Cronograma de entregas previstas para o objetivo n. 1

| Marco | Produto |
|---------------------|--|
| A cada 15 dias | Relatório de progresso (em formato digital). |
| Ao longo do 1º mês | Reuniões de alinhamento das equipes envolvidas. |
| Ao final do 2º mês | Proposta metodológica e cronograma de entregas. |
| Ao final do 4º mês | Apresentação da primeira versão da solução e entrega dos seguintes artefatos: - documentação da arquitetura de solução; - códigos usados para treinamento e retreinamento de modelos, com instruções; - códigos usados para execução em ambiente de produção, com instruções de implantação dos microsserviços; - datasets anotados, conforme o caso, com instruções para anotação. |
| Ao final do 7º mês | Apresentação da segunda versão da solução e entrega dos seguintes artefatos: - documentação da arquitetura de solução; - códigos usados para treinamento e retreinamento de modelos, com instruções; - códigos usados para execução em ambiente de produção, com instruções de implantação dos microsserviços; - datasets anotados, conforme o caso, com instruções para anotação. |
| Ao final do 10º mês | Apresentação da terceira versão da solução e entrega dos seguintes artefatos: |



| | |
|---------------------|---|
| | <ul style="list-style-type: none">- documentação da arquitetura de solução;- códigos usados para treinamento e retreinamento de modelos, com instruções;- códigos usados para execução em ambiente de produção, com instruções de implantação dos microsserviços;- datasets anotados, conforme o caso, com instruções para anotação. |
| Ao final do 12º mês | <p>Relatório final em formato de artigo científico segundo normas ABNT.</p> <p>Repasso de conhecimento técnico dos pesquisadores externos para os servidores da Ditec.</p> |

Objetivo 2: desenvolver algoritmos para comparar e aplicar emendas ao texto legislativo

A solução para o atingimento do objetivo de pesquisa n. 2 deve atender o especificado na Tabela 1. Os algoritmos devem ser implementados em linguagem Python observando-se os requisitos gerais apontados anteriormente. Os algoritmos a serem implantados em ambiente de produção devem ser parametrizáveis e implementados como microsserviços RESTFUL, passíveis de operar em ambientes de container (preferencialmente Docker).

Para o objetivo de pesquisa n. 2, necessita-se alocar pesquisadores externos, sendo no mínimo:

1. Um pesquisador doutor que atenda aos requisitos enumerados na Tabela 3;
2. Dois estudantes universitários, bolsistas de iniciação científica.

A Tabela 5 apresenta o cronograma de entregas previstas para o objetivo n. 2.

Tabela 5 - Cronograma de entregas previstas para o objetivo n. 2

| Marco | Produto |
|--------------------|--|
| A cada 15 dias | Relatório de progresso (em formato digital). |
| Ao longo do 1º mês | Reuniões de alinhamento das equipes envolvidas. |
| Ao final do 2º mês | Proposta metodológica e cronograma de entregas. |
| Ao final do 4º mês | Apresentação da primeira versão da solução e entrega dos seguintes artefatos: <ul style="list-style-type: none">- documentação da arquitetura de solução;- códigos usados para treinamento e retreinamento de modelos, com instruções;- códigos usados para execução em ambiente de produção, com instruções de implantação dos microsserviços;- datasets anotados, conforme o caso, com instruções para anotação. |



| | |
|---------------------|--|
| Ao final do 7º mês | Apresentação da segunda versão da solução e entrega dos seguintes artefatos: - documentação da arquitetura de solução; - códigos usados para treinamento e retreinamento de modelos, com instruções; - códigos usados para execução em ambiente de produção, com instruções de implantação dos microsserviços; - datasets anotados, conforme o caso, com instruções para anotação. |
| Ao final do 10º mês | Apresentação da terceira versão da solução e entrega dos seguintes artefatos: - documentação da arquitetura de solução; - códigos usados para treinamento e retreinamento de modelos, com instruções; - códigos usados para execução em ambiente de produção, com instruções de implantação dos microsserviços; - datasets anotados, conforme o caso, com instruções para anotação. |
| Ao final do 12º mês | Relatório final em formato de artigo científico segundo normas ABNT. Repasso de conhecimento técnico dos pesquisadores externos para os servidores da Ditec. |

Recomendações técnicas

Na busca por melhores resultados de **desempenho** dos modelos preditivos, recomenda-se que os pesquisadores avaliem a aplicabilidade das técnicas e estratégias indicadas pela literatura da área, dentre elas:

- *Transfer learning*, técnica de AM que permite a reutilização de modelos treinados previamente com domínio de dados ou tarefas NLP diferentes do domínio ou tarefa almejados, com vistas à redução do tempo de processamento e ao aumento da acurácia (redução da taxa de erros) em cenários com reduzido volume de dados anotados; a partir de 2018, o uso de modelos pré-treinados vem superando o estado da arte em diversas tarefas NLP: ULMFiT [3], ELMo[4], OpenAI Transformer [5], BERT [6], GPT-2 [7] e MultiFiT [8];
- *Active learning*, que permite que especialistas da área de negócio contribuam para o aprimoramento contínuo e incremental do desempenho da solução. O algoritmo de aprendizagem seleciona criteriosamente questões a serem respondidas pelo especialista [9, 10]. Os pesquisadores deverão definir no contexto da metodologia por eles proposta que tipos de questão/anotação serão coletados para treinamento das redes neurais, podendo utilizar inicialmente qualquer ferramenta de anotação disponível (ex: Doccano [11]) até que a Ditec conclua o desenvolvimento de uma interface customizada para melhor conforto dos especialistas da área de negócio;
- *Explainable/Interpretable Machine Learning*, de modo a ampliar o grau de interpretabilidade dos modelos gerados, na busca de explicações humanamente compreensíveis para suas saídas/inferências [12];



- Tratamento anti-viés (*debias*) tanto dos conjuntos de dados utilizados para treinamento de modelos quanto dos próprios modelos resultantes do treinamento [14].

Propriedade Intelectual (PI)

1. Divisão de PI

50% para a USP
50% para a Câmara dos Deputados

2. Justificativa para divisão da Propriedade Intelectual

Como está havendo contrapartida financeira por parte da Câmara dos Deputados, entende-se que a propriedade intelectual deve ser repartida igualmente.

3. Instituição responsável pelas ações de redação, registro ou proteção, acompanhamento e manutenção da Propriedade Intelectual

Câmara dos Deputados

4. Instituição responsável pelos custos de registro e manutenção da Propriedade Intelectual

Câmara dos Deputados

5. Instituição com prioridade na produção e exploração comercial da Propriedade Intelectual, conforme regulado em instrumento específico

Não se aplica, pois não haverá exploração comercial do que for desenvolvido nesta parceria.

6. Direitos Autorais sobre obras científicas ou literárias

50% para a USP
50% para a Câmara dos Deputados

Plano de Aplicação dos Recursos

Item I - Pesquisa para aprimorar desempenho de algoritmos existentes

| | Valor mensal | Valor por 12 meses | Total |
|----------------------------------|--------------|--------------------|---------------|
| 1 bolsa de pós-doutorado | R\$ 7.373,10 | R\$ 88.477,20 | R\$ 88.477,20 |
| 2 bolsas de iniciação científica | R\$ 695,70 | R\$ 8.348,40 | R\$ 16.696,80 |



| | |
|---------------------|----------------|
| Total para o item I | R\$ 105.174,00 |
|---------------------|----------------|

Item II - Pesquisa de algoritmos para aplicação de emendas ao texto legislativo

| | Valor mensal | Valor por 12 meses | Total |
|----------------------------------|--------------|--------------------|----------------|
| 1 bolsa de pós-doutorado | R\$ 7.373,10 | R\$ 88.477,20 | R\$ 88.477,20 |
| 2 bolsas de iniciação científica | R\$ 695,70 | R\$ 8.348,40 | R\$ 16.696,80 |
| Total para o item II | | | R\$ 105.174,00 |

| | |
|------------------------------|----------------|
| Total para o convênio | R\$ 210.348,00 |
|------------------------------|----------------|

Cronograma de desembolso

Os pagamentos serão feitos em 12 parcelas mensais de iguais valores iniciando 30 dias após a seleção da equipe de pesquisa, que deverá ser realizada pela Universidade de São Paulo em até 3 meses, contados da assinatura do presente aditivo.

Vigência

O convênio terá sua duração estendida por 15 meses e 15 dias, sendo:

- 45 dias para a seleção da equipe, pela USP, indicando à Câmara dos Deputados o nome e currículo dos pesquisadores;
- 12 meses para desenvolvimento da pesquisa, segundo cronograma definido acima;
- 2 meses para trâmites finais de pagamento e repasse das bolsas.



Referências

[1] BRASIL. Câmara dos Deputados. Gabinete do Diretor-Geral. **Portaria DG nº 268, de 14 de setembro de 2021. Ementa: Institui a Estratégia Digital da Câmara dos Deputados.** Brasília, 2021. Disponível em

<<https://www2.camara.leg.br/legin/int/portar/2021/portaria-268-14-setembro-2021-791753-norma-cd-dg.html>>

[2] BRASIL. Ministério da Ciência, Tecnologia e Inovações. Gabinete do Ministro. **Portaria MCTI nº 4.617, de 6 de abril de 2021.** Brasília, 2021. Disponível em:

<https://www.gov.br/mcti/pt-br/accompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ia_estrategia_portaria_mcti_4-617_2021.pdf>

[3] Howard, Jeremy; Ruder, Sebastian. **Universal Language Model Fine-tuning for Text Classification.** Disponível em <<https://arxiv.org/abs/1801.06146>>

[4] Peters, Matthew E.; Neumann, Mark; Iyyer, Mohit; Gardner, Matt; Clark, Christopher; Lee, Kenton; Zettlemoyer, Luke. **Deep contextualized word representations.** Disponível em <<https://arxiv.org/abs/1802.05365>>

[5] Radford, Alec; Narasimhan, Karthik; Salimans, Tim; Sutskever, Ilya. **Improving Language Understanding by Generative Pre-Training.** Disponível em <https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf>

[6] Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton; Toutanova, Kristina. **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.** Disponível em <<https://arxiv.org/abs/1810.04805>>

[7] Radford, Alec; Wu, Jeffrey; Child, Rewon; Luan, David; Amodei, Dario; Sutskever, Ilya. **Language Models are Unsupervised Multitask Learners.** Disponível em <https://d4mucfpksywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf>

[8] Eisenschlos, Julian; Ruder, Sebastian; Czapla, Piotr; Kardas, Marcin; Gugger, Sylvain; Howard, Jeremy. **MultiFiT: Efficient Multi-lingual Language Model Fine-tuning.** Disponível em <<https://arxiv.org/abs/1909.04761>>

[9] Settles, Burr. **Active Learning Literature Survey.** Disponível em <<http://burrsettles.com/pub/settles.activelearning.pdf>>

[10] dos Santos, Davi P.; Prudêncio, Ricardo Bastos Cavalcante; de Carvalho, André Carlos Ponce de Leon Ferreira. 2019. **Empirical Investigation of Active Learning Strategies.** Neurocomputing 326-327: 15-27.

[11] Doccano annotation tool. **Text Classification Demonstration.** Disponível em: <<https://doccano.herokuapp.com/demo/text-classification/>>



[12] Molnar, Christoph. **Interpretable Machine Learning: a Guide for Making Black Box Models Explainable.** Disponível em: <<https://christophm.github.io/interpretable-ml-book/>>

[13] Lample, Guillaume; Conneau, Alexis. **Cross-lingual language model pretraining.** Disponível em: <<https://arxiv.org/abs/1901.07291>>

[14] Wey, Jerry. **Bias in Natural Language Processing (NLP): A Dangerous But Fixable Problem.** Disponível em: <<https://towardsdatascience.com/bias-in-natural-language-processing-nlp-a-dangerous-but-fixable-problem-7d01a12cf0f7>>

[15] BRASIL. Câmara dos Deputados. Clique Regimento. **Tipos de Emendas.** Disponível em: <https://educacaoadistancia.camara.leg.br/clique_regimento/card/23>

[16] BRASIL. Câmara dos Deputados. **Estratégia Digital da Câmara dos Deputados.** <<https://www2.camara.leg.br/a-camara/estruturaadm/gestao-na-camara-dos-deputados/gestao-estrategica-na-camara-dos-deputados/gestao-estrategica-de-tic/estrategia-digital>>