

Projeto P&D

Instrução Assistida por Inteligência Artificial
TCU – Tribunal de Contas da União



Proposta Projeto P&D

Versão 1.0

Cliente TCU – Tribunal de Contas da União

Projeto P&D Instrução Assistida por Inteligência Artificial

Indústria Setor Público

Autores Ilan Figueiredo, Luis Quissak, Genaro Costa, Marcelo Valeri

Data de Elaboração	24/5/2022	Nome do Arquivo	9000985371 _ATOS E CIMATEC_CIMATEC_PROJETO_P& D_TCU_NLP_20220218_V1.0
---------------------------	-----------	------------------------	---

Copyright

Os direitos de copyright deste trabalho pertencem a **Atos e Cimatec** e a informação contida nele é confidencial. Este trabalho (integral ou em partes) não deve ser modificado, reproduzido, divulgado para terceiros ou usado para propósitos que não aos que se destina originalmente, sem autorização prévia da **Atos e Cimatec**. Se este trabalho ou suas partes forem fornecidos a um terceiro em virtude de um contrato, o uso deste trabalho bem como suas respectivas partes deverá ser guiado por termos expressos contratualmente entre a **Atos e Cimatec** e o terceiro especificado no contrato.

Índice

Sumário Executivo	4
1 Dados da Atos e SENAI Cimatec.....	7
2 Solução Proposta	8
2.1 Descrição da Solução	8
2.2 Etapas de Pesquisa.....	15
3 Cronograma Físico e Financeiro.....	17
3.1 Cronograma Físico.....	17
4 Gestão do Projeto e Escopo	19
4.1 Escopo de Serviços.....	19
4.2 Horário de Atendimento	20
4.3 Idioma coberto no escopo	20
4.4 Fora do Escopo	20
4.5 Estrutura Organizacional SENAI Cimatec e Atos.....	20
4.6 Governança	21
4.7 Matriz de Responsabilidades	22
4.8 Premissas.....	23
5 Experiencias Anteriores Senai Cimatec e Atos	25
5.1 JBS USA.....	27
5.2 CNI.....	28
5.3 Telconet.....	29
5.4 EMBRAPII P&D	29
6 Tabela de Anexos.....	32
6.1 Institucional Atos	32
6.2 Institucional Senai Cimatec.....	32
6.3 Currículos.....	32
6.4 Cronograma Físico.....	32
6.5 Metodologia Proposta.....	32

Sumário Executivo

Contexto

Este projeto de P&D atende ao escopo da Encomenda Tecnológica solicitada pelo Tribunal de Contas da União que tem por objetivo criar um módulo de Instrução Assistida por Inteligência Artificial, a ser incorporado à solução de Instrução Assistida do TCU, conferindo-lhe as seguintes inovações:

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.

A encomenda terá como escopo os processos de Representações e Denúncias.

E será executada em três ciclos, cada um relacionado a determinado conjunto temático:

- a) Aquisições Públicas.
- b) Demais temas.
- c) Futuras Aquisições Públicas sob a égide da nova Lei de Licitações (Lei 14.133/2021).

Proposição de Valor da Cimatec Atos e Cimatec

Através de uma equipe composta de doutores e mestres do Senai Cimatec e da equipe técnica especializada, flexível e ágil da Atos, propomos a **TCU** um projeto de Pesquisa e Desenvolvimento utilizando o conhecimento de pesquisas e projetos já realizados no Brasil com tecnologia e componentes de software utilizados na comunidade internacional.

O núcleo do atendimento a **TCU** será no Brasil: em São Paulo, Brasília e nos nossos centros de pesquisa em Salvador.

Temos uma equipe especializada em Analytics, Inteligência Artificial com conhecimento de Processamento de Linguagem Natural, que realizará a demanda da Encomenda Tecnológica solicitada pelo Tribunal de Contas da União.

Nossa solução possui uma abordagem por fases, que permite a medição do sucesso em cada etapa de modo a eliminar riscos e assegurar uma medição de resultados clara, viabilizando a implantação da solução inovadora para os serviços prestados pelo **TCU**.

Por que a Atos e Cimatec

Credenciais:

- ▶ O **SENAI CIMATEC** já realizou projetos de Pesquisa e Desenvolvimento envolvendo múltiplas disciplinas para diversas instituições e empresas tais como: Petrobras, Shell, Repsol Sinopec Brasil, NEXA, Ford, JBS, Siemens, Renault, Seara, HP, Huawei, Sanofi, Suzano etc.
- ▶ O **SENAI CIMATEC** tem como prioridade estratégica promover ações integradas, visando o desenvolvimento e à modernização da indústria. Estas ações ocorrem por

meio de atividades de educação, qualificação, atendimento tecnológico e de pesquisa aplicada, as quais buscam contemplar atendimentos às empresas que apresentam problemas de baixa, média e alta complexidade.

- ▶ O **SENAI CIMATEC** participa de Programas e Redes de Cooperação com Instituições de Ensino, Pesquisa e Desenvolvimento no Brasil e no Exterior. Estes programas corroboram para a atualização e compartilhamento de conhecimento entre os recursos humanos. Entre estas parcerias destacam-se: Universidade de Salvador (UNIFACS), Universidade Federal da Bahia (UFBA), Faculdade Metropolitana de Camaçari (FAMEC), Universidade de São Paulo (USP), Universidade Estadual de Campinas (UNICAMP), Faculdade de Tecnologia Empresarial (FTE), Universidade Federal de Santa Catarina (UFSC), Universidade Federal do Rio Grande do Sul (UFRGS) e o Institute für Bildsame Formgebung (Instituto de Conformação Mecânica) da Universidade Técnica da Renânia Vestfálica (RHTW), da Alemanha.
- ▶ O **SENAI CIMATEC** possui cerca 14 anos de experiência com desenvolvimento industrial no estado, em 2003 recebeu o prêmio FINEP de inovação tecnológica como instituição destaque no âmbito nacional.
- ▶ A área de competência de Big Data e Inteligência Artificial do **SENAI CIMATECA** é responsável pela capacitação acadêmica, pelos projetos de pesquisa aplicada à indústria, e pela prestação de serviços nas áreas de Inteligência Artificial, englobando Aprendizado de Máquina (do inglês Machine Learning), e o tratamento, armazenamento e disponibilização de grande volume de dados (Big Data) através das competências relacionadas com a Ciência de Dados (Data Science). Com projetos de sucesso na área de mineração, saúde, petróleo e gás, e manutenção industrial, a área de BigData e IA tem obtido um crescimento vertiginoso nos últimos anos. Recentemente, passou a integrar uma rede internacional de pesquisa no tema de Inteligência Artificial coordenada pela NVIDIA1 - NVIDIA AI Joint Lab, sendo a primeira unidade da América Latina; a Rede MCTI/EMBRAPII de Inovação em Inteligência Artificial; e a rede de Centros de Pesquisas Aplicadas (CPA) em Inteligência Artificial (IA) do MCTI/FAPESP/CGI.org.
- ▶ Já a área de Supercomputação é responsável pelo fornecimento de infraestrutura e profissionais especializados no processamento de alto desempenho contando atualmente com 5 supercomputadores em uma área de 250m², somando uma capacidade de processamento de pico de mais de 4 PFLOPs (quatrilhões de operações de ponto flutuante por segundo) e mais de 5 Petabytes de armazenamento. O laboratório possui computadores com diferentes processadores e arquiteturas: CPU Intel e ARM, GPU e FPGA além de software especializados. Realiza processamento de alto desempenho para simulações de alta fidelidade e resolução de modelos computacionais complexos. Dentre os recursos computacionais há os seguintes supercomputadores: HPC **CIMATEC** Yemoja, apoiado com 405 TFLOPs, equivalente a 17.200 núcleos computacionais, mais de 137.000 GB de memória RAM e mais de 2.000.000 de GB de armazenamento, o supercomputador CIMATEC Yemoja que equivale a mais de 40 mil computadores convencionais conectados por fibra ótica de alta velocidade trabalhando em conjunto para resolver os problemas mais custosos e complexos em pesquisa pura e aplicada; HPC OMOLU com foco no processamento de dados em inteligência artificial para a saúde pública, com cerca de 50 TFLOPs; HPC **CIMATEC** OGUN, com foco nas pesquisas e desenvolvimento para indústria brasileira, capacidade de 220 TFLOPs; HPC CIMATEC AIRIS, visa desenvolvimento de pesquisa em digitalização, sísmica e inteligência artificial para o setor de óleo & gás com capacidade de 900 TFLOPs; e HPC CIMATEC OGBON com capacidade de 2,5 PFLOPs destinado a pesquisas científicas para o setor de petróleo abrangendo desde a exploração até o abandono de campo.
- ▶ Recentemente, foi inaugurado no laboratório de Supercomputação o primeiro simulador quântico da América Latina denominado KUATOMU, destinado a simulação

e emulação aplicações a partir dos elementos quânticos, arquitetura promissora para futuras aplicações de alto desempenho.

- ▶ A **Atos realizou** projetos de Inovação envolvendo múltiplas tecnologias para diversas empresas tais como: Petrobras, VW CO, Seara, JBS, Siemens e Renault,
- ▶ A **Atos** possui um Centro de Operações, especializado em Analytics e Inteligência Artificial, que utiliza boas práticas de mercado e possui uma ampla cobertura de tecnologias e apoio global da AWS, Microsoft e Google,
- ▶ A **Atos** fornece serviços e soluções para empresa dos mesmos segmentos de atuação do TCU em diversas áreas, tais como Cyber Security, IT consulting, IT outsourcing, Automação, Visão Computacional, IA, ML, IOT entre outras,
- ▶ A **Atos** está presente em 73 países e possui cobertura de 100% em relação aos países onde a TCU atua o que nos dá tranquilidade e proximidade cultural para atender uma ampla gama de iniciativas para todo o grupo,
- ▶ A **Atos** é líder em sua indústria em Descarbonização, tendo adquirido a EcoAct, empresa de consultoria de estratégia de redução de carbono reconhecida internacionalmente,
- ▶ A **Atos** é reconhecida com o selo "Great Place to Work". Possuímos ações sociais integradas com nossa operação, as quais viabilizam inclusão social.

Como um parceiro de serviços de longa data, nosso contínuo crescimento global e na América do Sul, acreditamos que estamos bem-posicionados para proativamente trabalhar com a **TCU** para tornar esta iniciativa de grande sucesso.

Através desta proposta, eu e comitê executivo da Atos e Cimatec, assumimos o compromisso de entregar um serviço de alta qualidade com melhoria contínua para a **TCU**.

Atenciosamente,

Andre Alexandre Lopes de Souza
Account Manager

1 Dados da Atos e SENAI Cimatec

Matriz



São Paulo - SP

Avenida das Nações Unidas, N.º 12.901 – Torre Norte – 19.º Andar
Brooklin, São Paulo – SP – CEP: 04578-910
Tel: +55 11 3550 2000

Account Manager



Andre Alexandre Lopes de Souza
Celular: + +5561998084913
E-mail: andre.alexandre@atos.net

Solution Manager



Marcelo Valeri
Celular : +55 (11) 98765 2043
Email: marcelo.valeri@Atos e Cimatec.net



Sede Global da Atos e
Cimatec em Bezons -
França

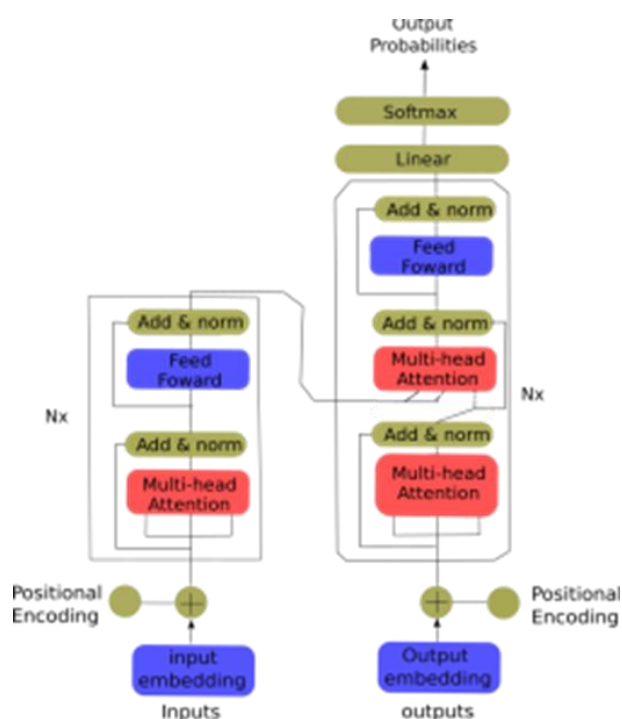
2 Solução Proposta

Visando atender ETEC, nossa estratégia para entendimento das necessidades da TCU é conforme apresentado abaixo.

2.1 Descrição da Solução

Nossa solução objetiva elevar ainda mais os padrões de prestação de serviços a população do Tribunal de Contas da União.

O desenvolvimento de modelos de aprendizado de máquina robustos, construídos a partir da análise de grandes volumes de dados, pode trazer consigo algumas demandas técnicas que, muitas vezes, tornam-se verdadeiros empecilhos a depender das características da investigação que se pretende realizar. Além da quantidade de dados na entrada do modelo, o uso de alguns tipos de redes neurais de aprendizado profundo pode levar a um tempo de processamento e custo computacional bastante alto. Esse é o caso das redes sequenciais com grandes quantidades de unidades ocultas para treinar, por exemplo. Uma forma de solucionar esse impasse é a utilização de transfer learning a partir de modelos pré-treinados, extraído conhecimento de um modelo treinado para uma tarefa e aplicando-o a uma nova tarefa por transferência de aprendizado. A utilização desse tipo de técnica reduz a necessidade de uma nova coleta e modelagem de dados para o endereçamento de problemas similares (PAN & YANG, 2009). Essa será a abordagem a ser empregada no projeto de pesquisa e desenvolvimento do TCU.



2.1.1 Características Gerais

Em termos gerais, conforme definido por VASWANI et al. (2017) e visualizado na Figura acima, a arquitetura Transformer utiliza uma abordagem encoder-decoder e um mecanismo de atenção denominado Multi Head Attention, em que cada head aprende uma distribuição de atenção diferente. Assim, aumenta-se a capacidade do modelo focar em diferentes posições, possuindo, então, cada atenção seu próprio subespaço de representações. O encoder e decoder são compostos por um empilhamento de camadas. Cada camada do encoder é formada por duas subcamadas, a primeira responsável pelo mecanismo de Multi-Head Self-Attention e a segunda formada por uma rede neural do tipo Feed Forward. São implementadas conexões residuais entre as subcamadas, e estas são seguidas por uma camada de normalização. Em relação ao decoder, além de ser formado por duas subcamadas especificadas no encoder, possui uma terceira camada que aplica o mecanismo de Multi-Head Self-Attention na saída do encoder.

Multi Head Attention trabalha com três vetores: *query* (Q), *keys* (K) e *values* (V). A técnica de *Attention* consiste na aplicação da função *softmax* no resultado do produto escalar do vetor *query* com todas as *keys*, dividido pela raiz da dimensão de *key*. Para cada *head*, é aplicado o mecanismo de atenção, e no final, ocorre uma concatenação desses resultados. Todo esse processo é sintetizado nas expressões abaixo.

$$\begin{aligned} \text{MultiHead}(Q, K, V) &= \text{Concat}(\text{head}_1, \dots, \text{head}_n)W^O, \\ \text{head}_i &= \text{Attention}(QW_i^Q, KW_i^K, VW_i^V), \\ \text{Attention}(Q, K, V) &= \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \end{aligned}$$

Similarmente a outros modelos de tradução de sequência, no *Transformer*, os *inputs* são representados por *word embeddings*. É comum também utilizar representação com *tokens* para cada palavra. Para essa situação, o início e o fim de sequências são sinalizados por *tokens* específicos. Na aplicação de *machine translation*, é essencial realizar uma representação da ordem do texto e, por esse motivo, após a codificação por *embeddings*, adiciona-se uma codificação posicional (*positional encoding*), resultando em uma representação *embedding with time signal*.

Paralelamente ao modo como os *Transformers* alavancaram o mecanismo de *self-attention* para modelar dependências de longo e curto alcance em texto e voz, novos trabalhos têm apresentado técnicas que utilizam o mecanismo de *self-attention* para superar as limitações apresentadas por métodos tradicionais de aprendizagem profunda (*deep learning*), com aplicações em diferentes áreas (p.e., CARION, 2020; LIM, 2021).

Neste sentido, o projeto aqui apresentado tem como base a seguinte questão de pesquisa: é possível obter o mesmo sucesso observado em tarefas de processamento de linguagem natural com a aplicação da arquitetura de *Transformers* a outras áreas de *deep learning*, aplicando, testando e tornando seu uso mais intuitivo e acessível para o desenvolvimento de novas soluções?

Com o objetivo de analisar os benefícios do uso de *Transformers* em modelos de *deep learning*, a proposta deste projeto é implementar a arquitetura para problemas em três

áreas, a saber: Detecção de significado nas peças processuais, Painel de jurimetria, Redação de peças;

Serão avaliados os possíveis ganhos em performance, custo computacional, paralelização, extração de características e representação de fenômenos com comportamento de mais longo prazo, bem como todo e qualquer novo aspecto resultante da pesquisa, já que se trata, também, de uma investigação exploratória. Além disso, os modelos desenvolvidos serão disponibilizados para futuras implementações em tarefas diversas via *transfer learning*, visando tornar seu uso mais acessível e intuitivo.

Processamento de Linguagem Natural (NLP)

Proposto inicialmente para tarefas de tradução (VASWANI et al., 2017), o uso de Transformers revolucionou a capacidade do Processamento de Linguagem Natural (NLP), que envolve inúmeros desafios em diferentes subáreas, tais como classificação de documentos, perguntas e respostas, interpretação textual, tradução automática e análise de similaridade semântica. O desempenho de modelos neurais sequenciais comumente utilizados em problemas relacionados à linguagem como LSTM, GRU e as RNNs, foram superados por modelos desenvolvidos com base em Transformers. Atualmente, os principais modelos baseados neste tipo de arquitetura são GPT, Turing-NLG, XLnet e BERT, considerando ainda as diferentes versões do GPT e variações do BERT.

A primeira versão do modelo GPT (Generative Pre-Training Transformer) foi apresentada em 2018 por pesquisadores da Open AI. Trata-se de um modelo autorregressivo, cujos primeiros experimentos tinham o objetivo de dar conta de um desafio enfrentado pelas tarefas de linguagem natural: a necessidade de grandes volumes de dados anotados para o treinamento de modelos dedicados a cada atividade em específico (como classificação de texto, análise de sentimento etc.). A proposta foi a utilização de um modelo pré-treinado com um corpus não rotulado e diverso, passando posteriormente por um fine-tuning para cada tarefa em específico (RADFORD et al., 2018). Os autores demonstraram a eficiência do modelo e, mais tarde, apresentaram uma versão ainda mais robusta chamada GPT-2, modelo treinado com um dataset maior e com a adição de mais parâmetros (RADFORD et al., 2019). Há uma versão do GPT-2 treinada especificamente para o português: o GPorTuguese-2, que utiliza o modelo GPT-2 reduzido pré-treinado com textos da Wikipédia em português (GUILLOU, 2020).

O GPT-3 é atualmente a versão mais completa da série, com 175 bilhões de parâmetros, e dimensão de word embeddings acrescida para 12.288 (contra 1600 do GPT-2). O modelo foi lançado em 2020 e demonstrou alta capacidade de lidar com tarefas de NLP, além de possuir recursos como a escrita de artigos e a realização de tarefas para as quais não foi treinado explicitamente, como a escrita de códigos computacionais, por exemplo (BROWN et al., 2020).

No mesmo ano em que a OpenAI lançava a primeira versão do GPT, o Google lançou o BERT (Bidirectional Encoder Representations from Transformers) em código aberto (DEVILIN et al., 2019). O BERT é um codificador Transformer bidirecional multicamadas, desenvolvido para pré-treinar representações bidirecionais profundas de texto não rotulado que pode ser ajustado com apenas uma camada de saída adicional, de modo a criar modelos para diferentes tarefas de NLP, com mínimas modificações na arquitetura de tarefas específicas. Uma variedade de modelos já foi criada para diferentes aplicações a partir da arquitetura do BERT, como o ALBERT (LAN et al., 2019), versão mais leve para

aprendizado auto supervisionado com menor número de parâmetros, e o RoBERTa (LIU et al., 2019), que propõe algumas modificações de treinamento do BERT de modo a torná-lo ainda mais robusto.

Aplicações para domínios específicos de linguagem também têm sido desenvolvidas. Este é o caso do BioBERT (BERT for Biomedical Text Mining), modelo com a mesma estrutura do BERT, pré-treinado com textos biomédicos e ajustado para três tarefas de mineração de textos da área médica: NER (Named Entity Recognition), RE (Relation Extraction) e QA (Question Answering). O desempenho do modelo superou modelos anteriores de mineração de textos biomédicos (LEE et al., 2020). Em sua versão específica para a língua portuguesa, o BioBERTpt foi proposto por Schneider et al. (2020), a partir do ajuste de modelos baseados em BERT treinados com textos da área médica em português. Os autores realizaram experimentos de NER, a partir dos quais o modelo treinado com o conjunto de textos clínicos apresentou melhor desempenho (SCHNEIDER et al., 2020). O estudo demonstra o benefício do transfer learning em processamento de linguagem natural para atividades clínicas.

Com o objetivo de avaliar as capacidades do transfer learning para língua portuguesa, Souza et al. (2020) propuseram o BERTimbau a partir do treinamento de modelos BERT para português Brasileiro. Os autores avaliam três tarefas finais de processamento de linguagem natural desempenhadas pelo modelo: similaridade semântica, inferência textual e reconhecimento de entidades nomeadas, comparando a performance dos modelos treinados com corpus em português com os modelos multilíngues disponíveis. Os resultados apresentados mostram que o BERTimbau possui melhor performance que o BERT multilíngue quando aplicados a tarefas em português e demonstram a grande utilidade de modelos pré-treinados em língua portuguesa para o desempenho de tarefas de NLP na língua.

Apesar de haver evoluções em muitos modelos de Transformers, há poucos avanços no que diz respeito à aplicação da técnica para o português Brasileiro. Como já testado para alguns procedimentos, um modelo pré-treinado em língua portuguesa tende a oferecer melhor desempenho que um modelo multilíngue. No entanto, os estudos desenvolvidos (supracitados) estão limitados a: uma área particular; modelo pouco robusto; ou tarefas específicas. Novos experimentos e investigações mais aprofundadas tornam-se, portanto, extremamente importantes para o avanço do campo em um domínio monolíngue focado em português Brasileiro. Assim, um dos objetivos, é implementar modelos mais recentes e modernos de NLP, realizando um pré-treinamento com corpus em língua portuguesa Brasileira, a fim de desenvolver um modelo que possa ser utilizado em diferentes tarefas de NLP em português Brasileiro.

Sistemas de Recomendação

A área de sistemas de recomendação investiga abordagens para aliviar a sobrecarga de informações das pessoas (especialmente usuários de serviços digitais) por meio de sugestões automatizadas e personalizadas. O uso de sistemas de recomendação potencializa o acréscimo das vendas em serviços online, aumento da satisfação geral do usuário, assim como um maior entendimento dos seus interesses e necessidades. Serviços de recomendação são hoje ubíquos abarcando domínios tão diversos quanto turismo, saúde, e-commerce e direito, por exemplo. Os métodos hoje mais bem sucedidos nessa área são métodos baseados em aprendizagem de máquina, notadamente modelos baseados em redes neurais profundas (Deep Neural Networks).

No contexto de recomendações baseadas em sessões de usuários, como é o caso em muitos sites de e-commerce, os métodos de melhor desempenho atualmente são modelos baseados em redes neurais recorrentes em conjunto com abordagens híbridas (tais como Wide & Deep) e, mais recentemente, métodos baseados em Transformers. Moreira et al. (2021) realizaram uma comparação entre várias arquiteturas de Transformers, originalmente propostas para tarefas de NLP, e baselines baseados em redes neurais recorrentes e KNN para a tarefa de recomendação. Eles adaptaram essas arquiteturas (Transformers-XL, BERT, ELECTRA, XLNet) para o problema da recomendação do próximo item, ou seja, dados os itens já consumidos dentro de uma sessão, a tarefa é recomendar o item que o usuário gostaria de consumir em seguida.

Considerando quatro conjuntos de dados de larga escala (dois no domínio de e-commerce e dois no domínio de notícias), os métodos baseados em Transformers apresentaram os melhores resultados na maioria dos cenários avaliados. O uso de Transformers também proporciona o aumento da capacidade de interpretabilidade e de explicabilidade de modelos de deep learning, uma vez que os pesos construídos a partir do mecanismo de Attention não apenas dão a compreensão do contexto subjacente dos dados de entrada, mas também são capazes de fornecer resultados mais explicáveis aos usuários, através da representação vetorial densa proporcionada pelos embeddings construídos, algo de grande relevância para uma área que utiliza dados do comportamento de usuários, a fim de mitigar aspectos de viés e injustiça dos modelos desenvolvidos (Khan et al., 2020). Esses resultados são recentes e trazem boas evidências que, embora incipientes, Transformers é uma tecnologia que promete trazer avanços significativos para aplicações baseadas em Sistemas de Recomendação.

Referências:

BROWN, Tom B. et al. Language models are few-shot learners. In: NeurIPS Proceedings - Advances in Neural Information Processing Systems 33, 2020.

CARION, Nicolas et al. End-to-end object detection with transformers. In: European Conference on Computer Vision. Springer, Cham, 2020. p. 213-229.

CORDONNIER, Jean-Baptiste; LOUKAS, Andreas; JAGGI, Martin. On the relationship between self-attention and convolutional layers. arXiv preprint arXiv:1911.03584, 2019.

DEVLIN, J., CHANG, M. W., LEE, K., & TOUTANOVA, K. Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL-HLT 2019, pages 4171–4186, 2019.

GUILLOU, Pierre. GPorTuguese-2 (Portuguese GPT-2 small): a Language Model for Portuguese text generation (and more NLP tasks...). (2020). In: pierre2020gpt2smallportuguese. Disponível em: <https://huggingface.co/pierreguillou/gpt2-small-portuguese> Acesso em: 07 jul. 2021.

KLINGENBRUNN, N. Transformers for Time-series Forecasting. Disponível em: <https://medium.com/mlearning-ai/transformer-implementation-for-time-series-forecasting-a9db2db5c820>. Acesso em: 30 jun. 2021.

LAN, Zhenzhong et al. Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942, 2019.

LECUN, Yann et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, v. 86, n. 11, p. 2278-2324, 1998.

LUONG, Minh-Thang et al. Addressing the rare word problem in neural machine translation. arXiv preprint arXiv:1410.8206, 2014.

MITCHELL, Benjamin R. et al. The Spatial Inductive Bias of Deep Learning. 2017. Tese de Doutorado. Johns Hopkins University.

PAN, Sinno Jialin; YANG, Qiang. A survey on transfer learning. IEEE Transactions on knowledge and data engineering, v. 22, n. 10, p. 1345-1359, 2009.

PARMAR, Niki et al. Image transformer. In: International Conference on Machine Learning. PMLR, 2018. p. 4055-4064.

PARMEZAN, Antonio Rafael Sabino. Predição de Séries Temporais por Similaridade. 2016. 221 f. Dissertação (Mestrado) - Curso de Ciências de Computação e Matemática Computacional, USP, São Carlos, 2016.

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

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

RAJPURKAR, Pranav; JIA, Robin; LIANG, Percy. Know what you don't know: Unanswerable questions for SQuAD. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15- 20, 2018, Volume 2: Short Papers, pages 784–789.

SCHNEIDER, Elisa Terumi Rubel et al. Biobertpt-a portuguese neural language model for clinical named entity recognition. In: Proceedings of the 3rd Clinical Natural Language Processing Workshop. 2020. p. 65-72.

SOUZA, F., NOGUEIRA, R., & LOTUFO, R. (2020). BERTimbau: pretrained BERT models for Brazilian Portuguese. In Brazilian Conference on Intelligent Systems (pp. 403-417). Springer, Cham.

VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A. N., ... & POLOSUKHIN, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).

VASWANI, A. et al. Attention is all you need. In Advances in Neural Information Processing Systems, pages 6000–6010, 2017.

WANG, Alex et al. GLUE: A multi-task benchmark and analysis platform for natural language understanding. In Proceedings of the Workshop: Analyzing and Interpreting Neural Networks for NLP, BlackboxNLP@EMNLP 2018, Brussels, Belgium, November 1, 2018, pages 353–355.

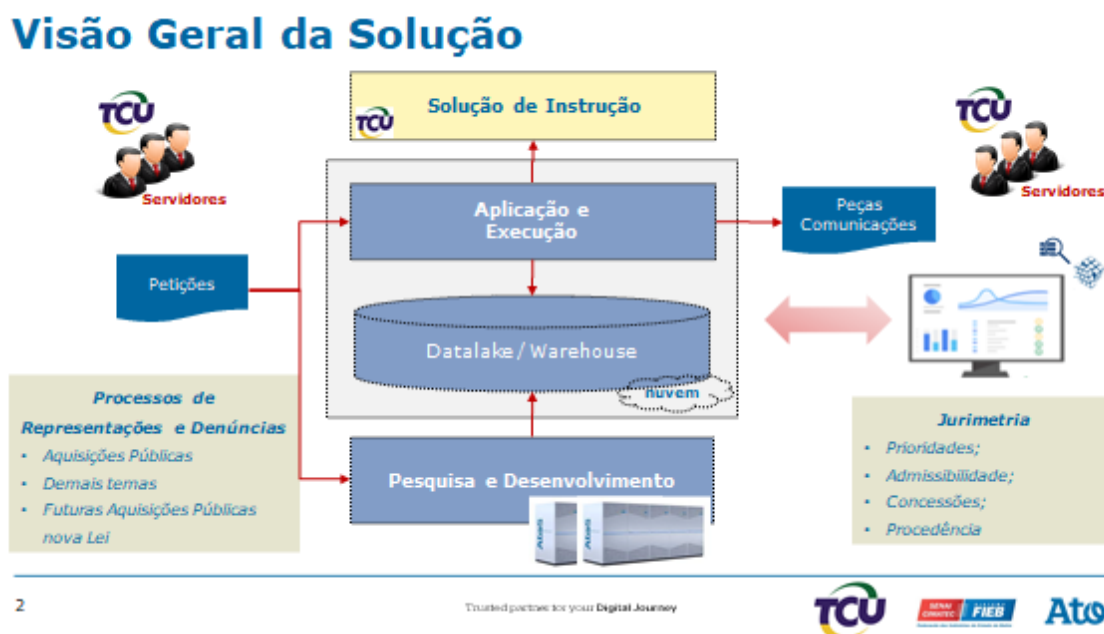
WU, N. et al. Deep transformer models for time series forecasting: The influenza prevalence case. arXiv preprint arXiv:2001.08317, 2020.

ZELLERS, Rowan et al. Swag: A large-scale adversarial dataset for grounded commonsense inference. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018, pages 93– 104.

Khan, Z. Y., Niu, Z., Sandiwarno, S., & Prince, R. Deep learning techniques for rating prediction: a survey of the state-of-the-art. Artificial Intelligence Review, 54, pgs. 95-135, 2020. doi:10.1007/s10462-020-09892-9

2.1.2 Visão Geral da Solução

Abaixo apresentamos uma visão inicial da solução técnica geral sugerida para a encomenda tecnológica:



A solução contempla dois momentos ou fases, a primeira sendo a da pesquisa e desenvolvimento e o segundo a da sustentação. A Arquitetura aqui apresentada será implementada em ambiente de desenvolvimento e será ampliada de forma elástica, utilizando nuvem computacional, para a realização do ambiente de produção, na fase de sustentação.

Nessa proposta contemplamos somente o ambiente para desenvolvimento, seja para o ambiente em nuvem, seja para o ambiente do Senai Cimatec de Pesquisa e Desenvolvimento, onde realizaremos o processamento de alta performance de treinamento profundo para obter os algoritmos, usaremos o servidor HPC Cimatec Ogbon.

O uso do HPC Cimatec Ogbon se dará nas etapas de simulação e em cada sprint de pesquisa e desenvolvimento, onde faremos de forma evolutiva a extração e anotação dos textos jurídicos e o treinamento para instruções e comunicações, usando modelos baseados em redes neurais recorrentes em conjunto com abordagens híbridas.

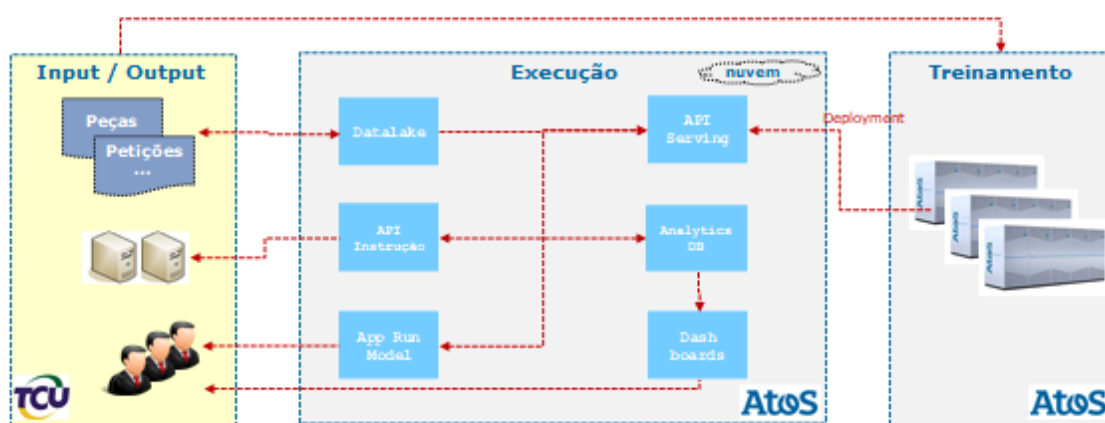
Os algoritmos gerados serão implementados na arquitetura em blocos acima apresentada, para interpretação dos textos jurídicos e os algoritmos para a redação das instruções serão construídos com as técnicas apresentadas no capítulo acima e serão implementadas no ambiente da nuvem representados pela caixa "Aplicação e Execução".

O processamento realizado pela "Aplicação e Execução" terá por missão fazer a ingestão de cada petição, a sua interpretação, a instrução, a comunicação e a geração dos dados para o painel de jurimetria. Todos os dados gerados pela execução dessa processo ficarão armazenados em um datalake para consulta e análise.

Considerando o ambiente futuro de produção, a sua execução diária e sua marcação de sucesso ou fracasso será objeto de entrada para realização do treinamento periódico dos Algoritmos, visando o aumento da sua acuracidade e precisão.

Os componentes planejados para serem implementados na solução técnica sugerida são apresentados abaixo:

Componentes Execução



2.2 Etapas de Pesquisa

Etapa 1. Estabelecimento da Rota Tecnológica

O objetivo desta etapa é realizar uma revisão do estado da arte, casos de uso para os diferentes desafios, prospecção e estabelecimento de rotas tecnológicas. O acompanhamento do cumprimento desta etapa será realizado através de relatório contendo a revisão bibliográfica e as rotas tecnológicas decorrentes desta revisão.

Etapa 2. Implementar Arquitetura de transformers

Esta etapa visa definir e implementar arquiteturas de transformers para diferentes domínios considerados neste projeto. Para acompanhamento do projeto será escrito relatório técnico contendo justificativas para escolha das arquiteturas como também a descrição e o funcionamento das arquiteturas.

Etapa 3. Treinamento e fine-tuning dos modelos (P&D)

Nesta etapa o foco é a construção, treinamento e validação dos modelos construídos na Etapa 2. Além disso, esta etapa visa realizar testes e fine-tuning dos modelos validados para aplicações específicas. O acompanhamento desta etapa será realizado através da disponibilização de repositório de código para treinamento e validação dos modelos. Além disso, serão escritos relatórios técnicos com a descrição do processo de construção, treinamento, validação, testes, e fine-tuning dos modelos.

Etapa 4. Aplicação da tecnologia em ambientes e escopos específicos (Sprints conforme definido na encomenda) (P&D)

Esta etapa visa a aplicação dos modelos validados em provas de conceitos alinhados com as empresas parceiras, refinamentos das arquiteturas e modelos, como também o desenvolvimento de APIs para uso dos modelos pelos parceiros. No mais, será disponibilizada documentação dos artefatos gerados a fim de realizar a transferência de tecnologia entre os parceiros. Para acompanhamento desta etapa, serão escritos relatórios técnicos contendo os resultados das aplicações dos modelos e justificativas dos ajustes realizados. Além disso, serão realizadas palestras entre os parceiros para transferência de tecnologia e workshop de finalização de projeto.

3 Cronograma Físico e Financeiro

3.1 Cronograma Físico

A figura abaixo apresenta um resumo de nossa solução de cronograma físico para atendimento a **TCU**.

Ciclos	Etapas	Tarefas	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18		
	Startup	Seleção - Início	S01	S04	S08	S12	S16	S20	S24	S28	S32	S36	S40	S44	S48	S52	S56	S60	S64	S68	S72	
Aquisições Públicas	Saneamento	Fornecimento de todas as peças	*																			
		Extração, rotulagem e estruturação dos dados para treino		*																		
		Parecer sobre a viabilidade		*																		
	P&D	Backlog do projeto																				
		Ajuste da fórmula de aferição																				
		Pesquisa - análise da petição inicial			1	3																
		Pesquisa - painel de jurimetria					5															
		Pesquisa - esboço da instrução						7	9													
		Prototipação - análise da petição inicial			1																	
		Prototipação - painel de jurimetria					3															
		Prototipação - esboço da instrução							5													
		Dev IA - análise da petição inicial				2	4															
		Dev IA - painel de jurimetria						6														
Dev IA - esboço da instrução							8	10														
Integração à Solução de Instrução Assistida									2													
Treinamento / KT						1	2		3	4												
Demais Temas	Saneamento	Fornecimento de todas as peças		*																		
		Extração, rotulagem e estruturação dos dados para treino			*																	
		Parecer sobre a viabilidade			*																	
	P&D	Backlog do projeto																				
		Ajuste da fórmula de aferição																				
		Pesquisa - análise da petição inicial								1	3											
		Pesquisa - painel de jurimetria										5										
		Pesquisa - esboço da instrução											7	9								
		Prototipação - análise da petição inicial									1											
		Prototipação - painel de jurimetria											3									
		Prototipação - esboço da instrução												5								
		Dev IA - análise da petição inicial										2	4									
		Dev IA - painel de jurimetria												6								
Dev IA - esboço da instrução													8	10								
Integração à Solução de Instrução Assistida																				2		
Treinamento / KT										1	2			3	4							
Nova Lei	Saneamento	Fornecimento de todas as peças																				
		Extração, rotulagem e estruturação dos dados para treino																				
		Parecer sobre a viabilidade																				
	P&D	Backlog do projeto																				
		Ajuste da fórmula de aferição																				
		Pesquisa - análise da petição inicial														1	3					
		Pesquisa - painel de jurimetria															5					
		Pesquisa - esboço da instrução																7	9			
		Prototipação - análise da petição inicial																1				
		Prototipação - painel de jurimetria																	3			
		Prototipação - esboço da instrução																				5
		Dev IA - análise da petição inicial																	2	4		
		Dev IA - painel de jurimetria																		6		
Dev IA - esboço da instrução																			8	10		
Integração à Solução de Instrução Assistida																						
Treinamento / KT															1	2				3		
Perfil	Empresa																					

Na figura acima M1, M2, Mn representam os meses de execução da pesquisa e desenvolvimento da solução para o TCU, S01, S04, Snn representam as semanas de execução da pesquisa.

A linha do tempo de cada mês foi consolidada para uma melhor visualização da página do documento, dessa forma cada mês sempre é representado pela primeira e quarta semana na figura acima, anexamos um arquivo do tipo .xls com o detalhe de todo cronograma físico.

O cronograma físico foi construído considerando o planejamento sugerido na Encomenda Tecnológica, considerando as fases Inicial, Saneamento e Pesquisa e Desenvolvimento.

Na fase inicial faremos a construção da equipe de pesquisa, o entendimento inicial de cada tipo de documento a ser tratado, os critérios para estabelecer a viabilidade da pesquisa e a revisão do planejamento de trabalho em conjunto com o TCU, essa fase

aparece nesse cronograma como sendo o Startup com a tarefa chamada de Seleção Início e com a duração de duas semanas.

A etapa de Saneamento foi planejada para ser realizada para os 3 grandes tipos considerados na pesquisa, quais sejam, Aquisições Públicas, Demais Temas e Nova Lei, com a duração total de 4 meses.

Na etapa de Saneamento será executada a extração, rotulação e estruturação dos dados para realização do treinamento usando aprendizado de máquina, obedecendo as técnicas e métodos acima apresentados no capítulo de solução de nossa proposta.

O treinamento será realizado no supercomputador do Senai Cimatec, localizado na cidade de Salvador, com esse aprendizado realizado em primeira rodada na fase de Saneamento, serão gerados os algoritmos para o processamento de linguagem natural. Após essa atividade é realizado o processamento na massa de documentos escolhido, seja Aquisições Públicas ou demais, sendo emitido um parecer técnico para seguimento ou não para as Etapas de Pesquisa e Desenvolvimento.

Caso o parecer técnico seja favorável, preparado pela Atos e Cimatec, com os resultados de testes de cada algoritmo e avaliado pelos técnicos e especialistas do TCU, o backlog ou um detalhamento dos futuros produtos a serem desenvolvidos nas sprints, serão definidos e em paralelo será executado os eventuais ajustes de aferição de cada fórmula.

Para cada tipo de documento será a seguir realizado os trabalhos de pesquisa, prototipação e desenvolvimento das soluções para execução de análise da petição inicial, esboço de instrução e construção do painel de jurimetria.

As equipes trabalharão obedecendo a metodologia SCRUM e estão planejados 10 sprints para cada tipo de documento, sendo então o cronograma de trabalho terminado em 18 meses.

Após o 18º mês de trabalho sugerimos uma atividade de implantação em produção e operação assistida, seguida da atividade de sustentação.

Está planejado nesse projeto físico o uso da plataforma de execução dos algoritmos de NLP na nuvem, privilegiando plataforma como serviço para realização da ingestão dos documentos, processamento, guarda e apresentação de dados e dashboards.

Para cada etapa de Pesquisa e Desenvolvimento está planejado também 4 sessões de transferência de conhecimento para garantir um entendimento de forma contínua.

4 Gestão do Projeto e Escopo

Nesta seção apresentamos nosso entendimento sobre o escopo solicitado na ETEC. Iniciamos com o escopo de tecnologias, seguido do escopo de serviços, idiomas, horário e SLA solicitados.

4.1 Escopo de Serviços

O SENAI CIMATEC será o responsável por liderar as atividades na área de Processamento de Linguagem Natural (Natural Language Processing - NLP) e previsão e classificação de séries temporais. Entretanto, o SENAI CIMATEC atuará em todas as tres frentes do projeto, sendo responsável por liderar tecnicamente a iniciativa. Na área de NLP o desafio reside em desenvolver modelos que recebem sequência de palavras ou letras para realizar tarefas tais como classificação de documentos, perguntas e respostas, interpretação textual, tradução automática e análise de similaridade semântica. A previsão e classificação de séries temporais possuem desafios como diagnóstico e prognóstico de eventos indesejáveis. O professor doutor Erick Giovani Sperandio Nascimento atuará como o principal nome e responsável pelo SENAI CIMATEC.

A tabela abaixo apresenta os serviços que estão dentro do escopo desta proposta.

Escopo
Iniciação
Realização KT do negócio
Refinamento do Product Backlog
Estabelecimento da Rota Tecnológica
Implementar Arquitetura de Transformers
Montagem do ambiente de CI/CD
Simulação
Treinamento e fine-tuning dos modelos
Realização dos algoritmos
Execução das provas e medição do sucesso
P&D
Definição dos ambientes de trabalho dos usuários finais

Realizar Sprints para Detecção de significado nas peças processuais
Realizar Sprints para painéis de jurimetria
Realizar Sprints para redação de peças
Execução dos testes integrados
Implementação
Suporte

4.2 Horário de Atendimento

De Segunda a Sexta das 09h00 às 18h00. Horário de Brasília – DF.

4.3 Idioma coberto no escopo

O suporte e toda prestação do serviço serão em português.

4.4 Fora do Escopo

Abaixo apresentamos a tabela de itens que não consideramos no escopo da ETEC.

Fora do escopo
Implantação do solução em produção
Integração da solução com sistemas do TCU
Operação Assistida
Sustentação e Suporte da Solução
Trabalhos nas dependências do TCU
Trabalhos fora do horário comercial

4.5 Estrutura Organizacional SENAI Cimatec e Atos

O diagrama abaixo apresenta a organização de nosso time para atendimento das demandas do ETEC.

Gestão	<ul style="list-style-type: none"> • Gerência do Projeto e Metodologia Scrum • Responsabilidade Atos
Assessoria Jurídica	<ul style="list-style-type: none"> • Assessoria para interpretação do significado e montagem da base de conhecimento • Responsabilidade Senai Cimatec
Anotação de Dados	<ul style="list-style-type: none"> • Estruturação e Anotação dos documentos jurídicos para aprendizado de máquina • Responsabilidade Senai Cimatec com Supervisão Atos
Análise Computacional	<ul style="list-style-type: none"> • Equipe especialista em Ciência e Análise de Dados • Responsabilidade compartilhada Senai Cimatec e Atos
Desenvolvimento Modelos de Inteligência Artificial	<ul style="list-style-type: none"> • Desenvolvimento dos Algoritmos usando o Supercomputador do Senai Cimatec • Responsabilidade Senai Cimatec
Engenharia de Dados	<ul style="list-style-type: none"> • Engenheiro de Dados • Responsabilidade Atos
Desenvolvimento Aplicação	<ul style="list-style-type: none"> • Criação das aplicações específicas do Tribunal considerando Front-end e Back-end • Responsabilidade Senai Cimatec
Testes	<ul style="list-style-type: none"> • Análise e Testes do Sistema • Responsabilidade Atos

Figura 1 – Organização e Responsabilidades das Equipes da Atos e Cimatec

A Estrutura acima é baseada em diversos papéis que podem ser desempenhados de forma dedicada ou compartilhada, com as responsabilidades indicadas acima.

4.6 Governança

A **Atos e Cimatec** irá estabelecer uma governança do contrato que irá assegurar o cumprimento de todas as atividades propostas. Os monitoramentos dos serviços executados pela **Atos e Cimatec** serão desempenhados em conformidade com as regras e atividades estabelecidas na ETEC pela **TCU**.

4.6.1.1 Plano de Comunicação

No plano de comunicação abaixo evidenciamos quem são os interlocutores por parte da **Atos e Cimatec** e do **TCU**, bem com o a frequência, escopo, modo e periodicidade da interação entre eles (sugerida).

Durante a iniciação serão determinados os nomes dos interlocutores, bem como a periodicidade das interações por nível hierárquico do plano de comunicação.

4.6.1.2 Relatórios – Gestão de Serviços

A **Atos e Cimatec** irá disponibilizar periodicamente relatórios de acompanhamento dos serviços contratados à **TCU**, de modo a possibilitar uma ampla visão de todas as iniciativas e frentes da prestação dos serviços. Segue a lista de relatórios:

Tabela 1 - Mapa de Relatórios a serem gerados

Veículos de Comunicação Escrita	Responsável	Público	Periodicidade	Meio de distribuição
Relatório de status do projeto	Scrum Master	Comitê Diretor; Gestores do Projeto	Semanal - após a reunião de acompanhamento do projeto.	E-mail
Relatório Executivo do projeto (parte do relatório de Status do Projeto)	Gestor do Projeto	Comitê Diretor; Alta Gerência e Diretora	Mensal	E-mail
Cronograma Físico e Financeiro	Gestor do Projeto	Comitê Diretor; Gestores do Projeto	Mensal	E-mail
Sprints e Produtividade	Scrum Master	Gestores do Projeto, Equipe de Projeto	Semanal - após a reunião de acompanhamento do projeto.	E-mail e Agile Dashboard
Product Backlog	Scrum Master	Gestores do Projeto, Equipe de Projeto	Semanal - após a reunião de acompanhamento do projeto.	E-mail e Agile Dashboard
Registro das Reuniões Diárias	Scrum Master	Gestores do Projeto, Equipe de Projeto	Diária	Agile Dashboard

4.7 Matriz de Responsabilidades

Atividades	Etapa	Atos e Cimatic	TCU
Gestão e liderança da Transição para atividades sob responsabilidade da Atos e Cimatic	Iniciação	X	
Facilitação para extração de dados para estruturação Atos e Cimatic	Iniciação		X
Liberação de acesso e disponibilização de espaço físico para os profissionais Atos e Cimatic	Iniciação		X
Liberação de acesso via VPN ao ambiente da TCU	Iniciação		X
Implementação da conectividade (VPN)	Iniciação	X	X
Realizar o papel de facilitação do processo de iniciação junto aos provedores de TI em situações de impasse	Iniciação		X
Disponibilizar documentações existentes de sistemas que por acaso não tenham sido compartilhadas	Iniciação		X
Colaborar na resolução de problemas que envolvam outros provedores em situações limítrofes em que há dúvida quanto à responsabilidade	Iniciação	X	

Fornecimento do ambiente computacional para geração dos algoritmos	Simulação	X	
Identifica o grupo correspondente ao escopo do problema e encaminhamento do chamado quando não for aplicável à Atos e Cimatec	Simulação	X	
Escalonamento dos chamados recebido conforme o escopo	Simulação	X	
Proposição de melhoria nos processos e ferramentas visando melhoria contínua	P&D	X	
Realizar testes de acordo com os procedimentos pré-acordados	P&D	X	
Solicitar homologação de cada sprint pela TCU	P&D	X	
Informar com antecedência quanto às mudanças (realizadas pela TCU e/ou terceiros) que impactem na prestação de serviços da Atos e Cimatec	Gestão		X
Garantir que todas as solicitações e requisições de serviços sejam feitas através do canal correto	Gestão		X
Aprovar em até 2 semanas as modificações, melhorias e adaptações que estiverem sujeitas à aprovação da TCU	Gestão		X
No caso de da transferência de conhecimento, deverá ser designado um ou vários “especialistas-chave” que exercerá o papel de multiplicador das práticas definidas como solução	Gestão		X
Gestão de KPIs e critérios de aceite acordados	Gestão	X	
Gestão dos Serviços sob sua responsabilidade e apresentação de relatórios periódicos	Gestão	X	
Assegurar que as reuniões do Plano de Comunicação ocorram e liderar as mesmas conforme a necessidade para o bom andamento do contrato	Gestão	X	
Participação nas reuniões conf. plano de comunicação previamente acordado	Gestão	X	X
Garantir a alocação de profissionais com os perfis necessários e adequados para a execução dos serviços;	Gestão	X	
Garantir a confidencialidade de dados e informações referentes à operação da TCU	Gestão	X	
Garantir o atendimento dos serviços contratados dentro dos níveis de serviço acordados entre as empresas.	Gestão	X	
Respeitar e fazer com que o seu pessoal respeite as normas de segurança e políticas internas em vigor na TCU	Gestão	X	
Manter a documentação referente aos serviços entregues, durante a vigência do contrato, com o devido zelo, segurança e sigilo;	Gestão	X	

4.8 Premissas

#	Premissa
1	Prazo de início: o prazo para início dos serviços será de até 1 mês após o aceite da proposta;
5	Não considerado despesas de viagem. Caso seja necessário deslocamentos, as despesas ficarão à cargo da TCU .
6	Os serviços (Running) serão prestados de forma remota.
7	Vale ressaltar que o escopo, premissas e outros elementos essenciais (incluindo, mas não limitado a, dependências, requisitos técnicos e

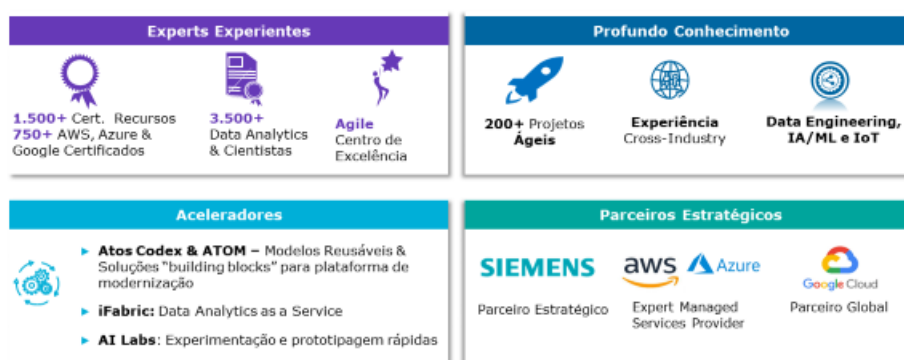
	<p>obrigações do cliente) desta Proposta foram definidos levando em consideração o status "as is", dos efeitos do COVID-19, no(s) local(is) em que a entrega irá ocorrer e/ou no momento da assinatura e aceite da Proposta (doravante, o "Status atual"). No caso de qualquer mudança no Status Atual, a Atos e Cimatec terá o direito de enviar uma alteração apropriada de sua Oferta.</p>
8	<p>O projeto será executado 100% de forma remota. Caso sejam necessárias viagens fora dessa base, os períodos serão acordados entre as partes e a TCU será responsável por arcar com tais despesas, assim como eventuais atrasos pela necessidade e constantes deslocamentos, seguindo as condições a seguir:</p> <p>A Atos e Cimatec irá dar ciência aos seus profissionais sobre a necessidade de deslocamento e caso o profissional não esteja de acordo em se locomover até o local do projeto, continuará a atuar de forma remota, não sendo passível de solicitação de substituição do profissional;</p> <p>A Atos e Cimatec irá dar ciência aos seus profissionais sobre os procedimentos de segurança a serem adotados nas instalações da TCU , tais como uso obrigatório de máscara e realização do teste de COVID-19. Caso o profissional Atos e Cimatec não esteja de acordo em realizar o teste, caberá à Atos e Cimatec buscar alternativas para solucionar a questão de forma a não impactar o andamento do projeto;</p> <p>Os custos referentes à realização dos testes de COVID-19 serão de responsabilidade da TCU ;</p> <p>Caberá à TCU arcar com todas as despesas de viagem, incluindo deslocamentos aéreo e terrestre, hospedagem, água, despesas de alimentação, despesas de lavanderia;</p> <p>A fim de otimizar os deslocamentos, o planejamento do projeto irá considerar "fly back" dos profissionais da Atos e Cimatec a cada 2 semanas.</p>
9	<p>A taxa hora apresentada na proposta comercial refere-se a horas normais de trabalho. Horas extras terão um multiplicador de 2x e horas de sobreaviso terão uma taxa de 1/3 do preço/hora apresentado na tabela da proposta comercial.</p>

5 Experiencias Anteriores Senai Cimatec e Atos

A Atos e Cimatec possui uma rede global de competência com vários centros de excelência entre eles o centro do Brazil, distribuídos em São Paulo, Londrina, Recife e Salvador, temos especialistas em visão computacional, engenheiros de dados, arquitetos de dados, cientistas de dados, arquitetos de soluções em analytics certificados nas principais nuvens como AWS, Google e Microsoft.

Abaixo apresentamos em um quadro nossas principais capacidades da Atos em projetos de Inovação e de Pesquisa e Desenvolvimento.

Atos em Inovação, IA e Video Analytics

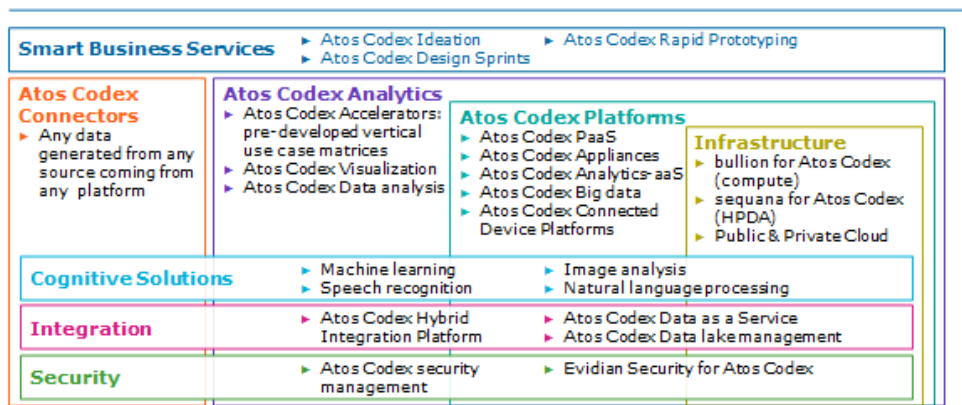


Trusted partner for your Digital Journey

Temos um modelo de referência que nos orienta na implantação e sustentação das plataformas de BigData de nossos clientes, chamada de Atos e Cimatec Codex, cujos principais blocos de ofertas e de funções está abaixo apresentado.

Atos Codex Digital Factory:

A reference model with a complete set of products and skills to design, build and run digital business platforms



10 © Atos

Nossas entregas e modelo são avaliados periodicamente por vários institutos de pesquisa e a seguir apresentamos as últimas avaliações.

Atos Codex Overview "LEADER" According to the Key Analysts



A seguir apresentamos projetos de inovação e de P&D desenvolvidos pela Atos e Cimatic nos últimos anos.

5.1 JBS USA

Com seis décadas de história, a JBS é atualmente o maior produtor de proteínas do mundo e a segunda maior empresa de alimentos do mundo¹. A companhia opera no processamento de carnes bovina, suína, ovina e de frango, e atua na produção de alimentos de conveniência e valor agregado. Além disso, comercializa produtos de couros, higiene e limpeza, colágeno, embalagens metálicas, biodiesel, entre outros.

Atualmente, a JBS possui mais de 400 unidades no mundo, sendo mais de 230 diretamente relacionadas à produção de carnes e produtos de maior valor agregado e conveniência. A Companhia conta com mais de 240 mil colaboradores e com capacidade para processar, por dia, mais de 75 mil bovinos, em torno de 14 milhões de aves, 115 mil suínos e 60 mil peças de couro.

P&D realizado em prova de conceito que está em TRL-9:

- ▶ Os seguintes casos de uso:
 - Esterilização de facas no corte
 - Assertividade do Abate de Animais
 - Análise de Couro
 - Redução de descarte/sobras de carnes nos ossos

Período: 2019 – atual

Fases realizadas:

- ▶ Saneamento de Dados e Criação dos Modelos computacionais
- ▶ Pesquisa e Desenvolvimento
- ▶ Implantação em Produção

Escopo:

- ▶ Obtenção, tagueamento e marcação de vídeos
- ▶ Treinamento de modelos computacionais e geração de algoritmo nos supercomputadores do Cimatec
- ▶ Instalação física e lógica de servidores de borda (Edge), câmeras e demais equipamentos
- ▶ Implantação dos algoritmos nos computadores de borda
- ▶ Testes cíclicos e retreinamento dos modelos computacionais através da análise em tempo real de cada caso de uso
- ▶ Desenvolvimento Sistema de Indicadores e Alertas para cada tipo de caso de uso, residente na nuvem computacional

5.2 CNI

A CNI é a principal representante da indústria brasileira na defesa e na promoção de políticas públicas que favoreçam o empreendedorismo e a produção industrial, num setor que reúne mais de 476 mil indústrias no país.

P&D realizado em prova de conceito:

- ▶ Os seguintes casos de uso:
 - Entendimento das trilhas de formação do CNI
 - Entendimento dos currículos dos alunos em formação no CNI
 - Entendimento das vagas oferecidas em mercado no linkedin
 - Descobrimto do gap de trilhas de formação em comparação com o mercado
 - Descobrimto do gap de currículos dos alunos em comparação com o mercado

Período: 2021 – atual

Fases realizadas:

- ▶ Entendimento Inicial
- ▶ Saneamento de Dados para uma trilha de formação para efeitos de demonstração
- ▶ Início do contrato

Escopo:

- ▶ Extração e rotulação dos textos das trilhas de formação – currículos de alunos – vagas oferecidas no linkedin
- ▶ Marcar e estruturar dados para fins de treinamento de modelos computacionais
- ▶ Realizar o treinamento do modelo computacional utilizando tokenização, NLTK e AutoML do Bigquery
- ▶ Implantação dos algoritmos na nuvem
- ▶ Testes cíclicos e retreinamento dos modelos computacionais através da análise em tempo real de cada caso de uso
- ▶ Desenvolvimento Sistema de Indicadores para cada trilha de formação, residente na nuvem computacional

5.3 Telconet

Telconet é uma empresa com mais de vinte e seis anos de experiência no Equador, desenvolvendo soluções tecnológicas para Rede, Conectividade, Nuvem, Segurança, Colaboração, Segurança Eletrônica e Trânsito para o segmento corporativo, com base em uma sólida plataforma e infraestrutura de alto Fibra Óptica de nível de capilaridade que permite desenvolver negócios, juntamente com as subsidiárias pertencentes à Holding Telconet, bem como a presença em outros países da América Latina, como Panamá, Guatemala e Colômbia.

P&D realizado em prova de conceito que está em TRL-8:

- ▶ Os seguintes casos de uso:
 - Reconhecimento de Placas de Automóveis
 - Reconhecimento de Eventos Complexos como Carro abandonado, etc
 - Reconhecimento Facial

Período: 2020 – atual

Fases realizadas:

- ▶ Saneamento de Dados e Criação dos Modelos computacionais
- ▶ Pesquisa e Desenvolvimento
- ▶ Preparação para implantação em Produção

Escopo:

- ▶ Obtenção, tagueamento e marcação de vídeos
- ▶ Treinamento de modelos computacionais e geração de algoritmo nos supercomputadores do Cimatec
- ▶ Instalação física e lógica de servidores de borda (Edge), câmeras e demais equipamentos
- ▶ Implantação dos algoritmos nos computadores de borda
- ▶ Testes cíclicos e retreinamento dos modelos computacionais através da análise em tempo real de cada caso de uso
- ▶ Desenvolvimento Sistema de Indicadores e Alertas para cada tipo de caso de uso, residente na nuvem computacional

5.4 EMBRAPII P&D

A EMBRAPII (Associação Brasileira de Pesquisa e Inovação Industrial) é uma Organização Social qualificada pelo Poder Público Federal que, desde 2013, apoia instituições de pesquisa tecnológica fomentando a inovação na indústria brasileira.

A assinatura do Contrato de Gestão com o Ministério da Ciência, Tecnologia, Inovações e Comunicações – MCTIC ocorreu em 2 de dezembro de 2013, tendo o Ministério da Educação – MEC como instituição interveniente. Os dois órgãos federais repartem igualmente a responsabilidade pelo seu financiamento.

A contratação da EMBRAPII parte do reconhecimento das oportunidades de exploração das sinergias entre instituições de pesquisa tecnológica e empresas industriais, em prol do fortalecimento da capacidade de inovação brasileira. Ela tem por missão apoiar instituições de pesquisa tecnológica, em selecionadas áreas de competência, para que executem projetos de desenvolvimento de pesquisa tecnológica para inovação, em cooperação com empresas do setor industrial.

A EMBRAPII atua por meio da cooperação com instituições de pesquisa científica e tecnológica, públicas ou privadas, tendo como foco as demandas empresariais e como alvo o compartilhamento de risco na fase pré-competitiva da inovação. Ao compartilhar riscos de projetos com as empresas, tem objetivo de estimular o setor industrial a inovar mais e com maior intensidade tecnológica para, assim, potencializar a força competitiva das empresas tanto no mercado interno como no mercado internacional..

P&D em execução:

- ▶ Os seguintes casos de uso:
 - Área-foco 1: “Transformers4All: Um Framework baseado em Transformers para Domínios Diversos”
 - Área-foco 3: “Plataforma de desenvolvimento de few-shot learning para visão computacional e processamento de linguagem natural [FEWSHOT]”

Período: 2021 – atual

Fases realizadas:

- ▶ Saneamento de Dados e Criação dos Modelos computacionais
- ▶ Pesquisa e Desenvolvimento em curso

Escopo:

- ▶ A expectativa ao final do é a disponibilização de um repositório de modelos de Deep Learning baseados na arquitetura Transformers para toda a comunidade (e.g.: GitHub). Os modelos desenvolvidos e disponibilizados possuem finalidades distintas nas seguintes áreas estratégicas:
 - ▶ (i) Visão Computacional,
 - ▶ (ii) Séries Temporais,
 - ▶ (iii) Sistemas de Recomendação e
 - ▶ (iv) Processamento de Linguagem Natural. Obtenção, tagueamento e marcação de vídeos

A iniciativa abrangerá 4 Unidades EMBRAPII da Rede MCTI/EMBRAPII de Inovação em Inteligência Artificial, com CATI vigente (SENAI CIMATEC, CEEI/UFMG, Eldorado e EDGE UFAL). O desenvolvimento do projeto contará com os recursos do PPI IoT/Manufatura 4.0, sob a coordenação da EMBRAPII, e será executado seguindo as regras estabelecidas na Lei de TICs e em dois instrumentos de acompanhamento utilizados pela EMBRAPII: o Manual de Operação e a Orientação Operacional específica sobre essa ação.

No campo da aprendizagem profunda de máquina, ou deep learning, os modelos baseados em mecanismos de atenção, conhecidos como transformers, têm revolucionado a área de processamento de linguagem natural nos últimos anos. Mais recentemente, diversas propostas de arquiteturas voltadas para outros tipos de aplicações, como previsão e análise de séries temporais, visão computacional e sistemas de recomendação, têm

surgido na literatura científica sobre o tema. No campo da visão computacional, por exemplo, estes estudos já mostram um futuro bastante promissor desta tecnologia devido à sua simplicidade e escalabilidade. Além disso, estudos mostram que em visão computacional os transformers também abordam limitações conhecidas das principais arquiteturas utilizadas hoje em dia, tais como as dependências de longo prazo e alcance presentes nas redes convolucionais.

Da mesma forma, estudos recentes têm mostrado que os transformers são promissores em outras áreas. Em séries temporais, por exemplo, são até oito vezes mais resistentes; em termos de acurácia, apresentam uma melhora de 20% se comparados com outros métodos, além da redução da complexidade computacional. No campo de sistemas de recomendação também já existem trabalhos mostrando o potencial de adaptações de arquiteturas de transformers, originalmente utilizadas em processamento de linguagem natural, para trazer performances melhores do que as técnicas utilizadas atualmente. Além disso, a sua facilidade de interpretação das decisões dos modelos também contribui para a transparência das decisões tomadas. Já existem muitos avanços na aplicação de transformers para processamento de linguagem natural da língua inglesa. No entanto, a tecnologia e o seu potencial de aplicação ainda estão em evolução para o contexto da língua Portuguesa no Brasil.

Desta forma, o objetivo deste projeto é estudar e implementar arquiteturas de transformers para problemas de visão computacional, séries temporais, sistemas de recomendação e processamento de linguagem natural para a língua Portuguesa no Brasil. O estudo será realizado através de levantamento bibliográfico, casos de uso das áreas alvo, implementação de arquiteturas, fine-tuning de modelos e disponibilização de modelos de transformers em um framework através de APIs. O projeto contribuirá com o enriquecimento do conhecimento sobre esta tecnologia emergente no Brasil e a simplificação da sua utilização em casos de uso comuns de deep learning.

6 Tabela de Anexos

6.1 Institucional Atos

Nome
Anexo I – Institucional Atos

6.2 Institucional Senai Cimatec

Nome
Anexo II – Senai Cimatec Apresentação Institucional

6.3 Currículos

Nome
Anexo III – Currículos Atos e Senai Cimatec

6.4 Cronograma Físico

Nome
Anexo IV – Cronograma Físico

6.5 Metodologia Proposta

Nome
Anexo V – Metodologia de Trabalho

Sobre a Atos

A **Atos** é líder global em transformação digital com 105.000 funcionários em 71 países e receita anual de € 11 bilhões. Européia número um em Cloud, Cybersecurity e High-Performance Computing, o Grupo fornece soluções de nuvem híbrida, Big Data, Aplicativos de Negócios e Workplace Digital por meio de sua Fábrica de Transformação Digital. Com suas tecnologias de ponta e conhecimento do setor, a **Atos** apóia a transformação digital de seus clientes em todos os setores de negócios. O Grupo é o Parceiro Mundial de Tecnologia da Informação para os Jogos Olímpicos e Paralímpicos e opera sob as marcas **Atos**, **Atos Syntel**, e Unify. A **Atos** está listada no índice de ações CAC40 Paris.

Saiba mais sobre nós
[Atos.net](https://atos.net)
[Atos.net/blog](https://atos.net/blog)

Vamos começar uma discussão juntos



Todas as marcas registradas são de propriedade de seus respectivos donos. **Atos e Cimatec**, o logotipo da **Atos e Cimatec**, **Atos e Cimatec Syntel**, e Unify são marcas registradas do grupo **Atos e Cimatec**. A **Atos e Cimatec** se reserva o direito de modificar este documento a qualquer momento, sem aviso prévio. Algumas ofertas ou partes de ofertas descritas neste documento podem não estar disponíveis localmente. Entre em contato com o escritório local da **Atos e Cimatec** para obter informações sobre as ofertas disponíveis em seu país. Este documento não representa um compromisso contratual. Janeiro de 2021. © 2021 **Atos e Cimatec**.