

RESULTADO DA SELEÇÃO
E-TEC DE INSTRUÇÃO ASSISTIDA POR INTELIGÊNCIA ARTIFICIAL
CHAMAMENTO PÚBLICO DO TRIBUNAL DE CONTAS DA UNIÃO Nº 1/2022

A **Comissão de Seleção de Encomenda Tecnológica para a Instrução Assistida de Representações e Denúncias por Inteligência Artificial**, constituída pela ORDEM DE SERVIÇO-CCG Nº 02/2022 no âmbito do CHAMAMENTO PÚBLICO DO TRIBUNAL DE CONTAS DA UNIÃO Nº 1/2022, **DECLARA COMO PROPONENTE PRIMEIRO COLOCADO** para firmar Contrato de Encomenda Tecnológica com o TCU:

O **Consórcio NEURALMIND TERRANOVA**, formado pela empresa NEURALMIND INTELIGÊNCIA ARTIFICIAL S.A., CNPJ 10.448.583/0001-71, e pela empresa TERRANOVA CONSULTORIA ESTATÍSTICA LTDA., CNPJ 23.374.326/0001-84.

Fica o proponente acima CONVOCADO para a etapa de acolhimento das considerações da Consultoria Jurídica do TCU sobre a minuta de contrato negociada. Após o que o resultado da seleção deverá ser homologado, para que o contrato possa ser celebrado.

Motivação da decisão

Dois concorrentes foram destacados pelo Comitê Técnico de Especialistas como tendo *"nível equivalente de competência técnica, sendo plenamente capazes de executar as tarefas da ETEC"*. São eles, em ordem alfabética: Projeto CoreJur (consórcio formado por RedeEmpresarial LTDA e CEIA/UFG) e Projeto IN² (consórcio formado por NeuralMind SA e Terranova Consultoria LTDA). Os projetos adotam rotas tecnológicas diferentes em relação ao uso de modelos de linguagem baseados em Transformers. O projeto CoreJur propõe, preferencialmente, a aplicação de vários modelos especialistas com classes, como os originados da família Bert. O projeto IN² prefere um único modelo few shot com prompts, como o GPT-3.

Conforme parecer dos negociadores do TCU: *"A negociação foi concluída com sucesso em relação a ambos os candidatos. Foi possível obter consenso quanto aos pontos que os negociadores do TCU consideraram relevantes. Os negociadores consideram que as posições defendidas pelo TCU foram atendidas em todos os casos. Destaque-se que ambos os candidatos aceitam seguir pela rota tecnológica que não lhes é preferencial em caso de insucesso da rota preferencial."*

Em resumo, ambos os concorrentes se equivalem quanto a:

- a) Qualidade do Projeto e sua aderência ao Edital.
- b) Competência técnica, avaliada pelo CTE em termos de portfólio e equipe apresentados.
- c) Custos, com uma diferença de apenas 5% nas estimativas dos concorrentes.
- d) Prazo, com uma estimativa de 2 meses de diferença entre eles, para um projeto de 2 ½ anos.
- e) Concordância com as exigências de Propriedade Intelectual do TCU. E com o posicionamento técnico do TCU sobre transferência de conhecimento e aferição dos resultados de P&D.

A decisão recai principalmente, portanto, sobre qual proposta melhor atende aos aspectos de estratégia de negócio sob o ponto de vista do TCU.

Esta comissão entende que o Projeto do consórcio NeuralMind Terranova carrega grau de incerteza maior, em função do estágio ainda incipiente da rota tecnológica proposta. De fato, há apenas 6 meses ainda não existiam técnicas de *tunning* que provavelmente serão usadas na ETEC, por exemplo, para evitar que o modelo se perca nas datas e valores expressos nas peças processuais. Embora recente, a rota é a aposta de grandes players do mercado, como Nvidia, Google, Microsoft e OpenAI, que investem vultosas quantias na criação de modelos generalistas o suficiente para servirem aos mais diversos clientes e aplicações.

A rota proposta pelo Projeto CoreJur, embora também recente e em evolução, pode ser vista como a abordagem mais utilizada no mercado atualmente. Requer um conjunto de operações com classes e parágrafos hoje considerado como o padrão de mercado para se lidar com tarefas de NLP. Ou seja, é a rota mais consolidada para o desenvolvimento de produtos de inteligência artificial envolvendo textos. Tais produtos existentes no mercado estão aquém do que é pedido nesta ETEC. Conforme justificado nos Estudos (peça 02) que embasaram a presente contratação, o objeto da ETEC requer a resolução cumulativa de vários problemas de extração de significado textual (como as irregularidades alegadas nas petições iniciais), classificação (quanto aos critérios de admissibilidade, critérios para atendimento aos pedidos de medidas cautelares, critérios de semelhança entre causas etc.) e geração de textos (sumarização da causa, geração de encaminhamentos, dentre outros). Cada um dos problemas pode ser considerado difícil, exigindo um modelo especialista próprio. A solução completa é a soma de todas as tarefas a serem resolvidas, com o erro na resolução de cada tarefa se propagando para as próximas, cumulativamente. De modo que a rota tecnológica proposta leva a uma solução composta de muitas engrenagens a serem mantidas. Outro risco que se vislumbra nesta abordagem está no inevitável erro humano introduzido pela necessidade de se criar um dataset anotado para viabilizar o treinamento dos modelos. Embora mitigado pelo fato de requerer menor volume de dados do que abordagens ora ultrapassadas, ainda assim seriam necessários muitas centenas de processos, com milhares de peças e dezenas de milhares de parágrafos anotados. O entendimento desta Comissão, portanto, é o de que a rota proposta pela CoreJur pode ser

difícil de se escalar para o tamanho da complexidade do objeto da ETEC. E que a solução gerada pode exigir manutenção laboriosa e especializada.

Ao revés, a rota proposta pelo Projeto NeuralMind TerraNova não requer treinamento de vários modelos, mas o refino de um único megamodelo através de procedimentos como os de engenharia de prompts, cujo conhecimento se aparenta mais fácil de ser transferido ao corpo técnico do TCU. Ademais, requer bem menos documentos anotados, permitindo a revisão de todos os casos para eliminar o erro humano na anotação.

Em termos da operação da solução e seus custos, ambas as propostas permitem o uso em nuvem ou *on premisses*. A operação da solução CoreJur exige menor capacidade de processamento, possivelmente atendida pelas máquinas hoje existentes no CPD do TCU. Mas também é possível usar computação em nuvem. A solução NeuralMind requer a assinatura de um megamodelo de linguagem como serviço. Ou a aquisição, pelo TCU, de hardware específico, cujo custo já foi previsto na proposta do consórcio.

Quanto às perspectivas de como a solução poderá induzir novas transformações digitais no TCU, entende-se que a evolução da solução proposta pela NeuralMind é significativamente menos custosa. Sendo possível, por exemplo, transpor a solução para um novo megamodelo de linguagem futuro, operação que dispensa retreinamentos. Ainda neste diapasão, a solução pronta poderia ser adaptada por especialistas do TCU para outros cenários de instrução assistida. Não apenas futuras temáticas ou inovações legislativas no campo das Representações e Denúncias. Mas sim expandindo-a para outros tipos processuais.

Sobre os resultados esperáveis, o Parecer do CTE apresenta as duas propostas como tendo relevante chance de sucesso. E esclarece que os resultados atualmente considerados como estado da arte no mercado são obtidos com modelos especialistas altamente treinados (rota proposta pela CoreJur). Megamodelos generalistas (rota NeuralMind) são promissores, pois apresentam bons resultados sem ou praticamente sem qualquer refino (zero ou few shot). Talvez possam vir a superar o atual estado da arte com as técnicas de refinamento que vêm sendo desenvolvidas. Simplesmente não há certeza técnica hoje sobre qual caminho levará ao melhor resultado no caso concreto da ETEC. Tal incerteza, aliás, é

um dos motivadores da ETEC, com a intenção de fomentar a inovação tecnológica brasileira nesta área. Em tal cenário, entende-se que a proposta do consórcio NeuralMind e TerraNova está mais alinhada com os objetivos de inovação da ETEC e que tem chance de levar a resultados bons e até aos melhores resultados possíveis. Contudo, dado seu maior grau de risco tecnológico, se faz indispensável exigir uma comprovação, ainda no início da execução da ETEC, que a solução proposta apresenta um desempenho mínimo em todos os problemas a serem resolvidos. Como a rota dispensa um longo período de saneamento de dados, foi negociado que a fase de saneamento prevista no Edital seja utilizada para a construção de um Mínimo Produto Viável que estabeleça uma linha de base do desempenho que se pode esperar da solução. Com base no MVP o TCU decidirá pela continuidade ou não da ETEC na rota tecnológica.

Conclusão e encaminhamento

Fica selecionada a proposta do Consórcio NeuralMind TerraNova. Finda a fase de saneamento (6 meses de duração), esta Comissão voltará a se reunir para decidir sobre a continuidade da ETEC na mesma rota tecnológica.

Brasília, 28 de junho de 2022

*Assinam eletronicamente este documento membros da
Comissão de Seleção da ETEC
designados pela OS-CCG 02/2022.*