



ESTADO DO CEARÁ
PODER JUDICIÁRIO
TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

ESTUDO TÉCNICO PRELIMINAR

INFORMAÇÕES BÁSICAS

Número do processo: 8512691-65.2025.8.06.0000

Código contratação: AQSETIN2024026

Área da Demanda: SETIN/Gerência de Robotização e Inteligência Artificial

O Estudo Técnico Preliminar tem por objetivo identificar e analisar os cenários para o atendimento da demanda que consta no DOD/DFD, bem como demonstrar a viabilidade técnica e econômica das soluções identificadas, fornecendo as informações necessárias para subsidiar a decisão de atendimento.

1. DESCRIÇÃO DA NECESSIDADE:

1.1. DESCRIÇÃO DA NECESSIDADE DE NEGÓCIO

1.1.1. Diante da política de planejamento, asseguradas no Plano de Contratações de STIC, no Plano Diretor de Tecnologia da Informação e Comunicação (PDTIC), além do Planejamento Estratégico Institucional, alinhada com a Estratégia Nacional do Poder Judiciário e Estratégia Nacional de TIC (ENTIC-JUD), com o objetivo de prover soluções de Inteligência Artificial (IA) inovadoras e disruptivas sem comprometer o abastecimento das unidades administrativas e judiciárias, nem mesmo deixá-las obsoletas, faz-se necessário avaliar a necessidade da disponibilização de serviço de IA generativa relacionado ao DOD/DFD que provocou estes estudos preliminares, a fim de evitar falta ou excesso e desperdício de bens ou serviços.

1.1.2. Neste sentido, primeiramente foram avaliadas as efetivas necessidades que justificam o serviço de IA generativa pretendido, conforme indicado no DOD/DFD a demanda de disponibilização de serviço de IA generativa para as seguintes necessidades de negócio atualmente existentes:



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

1.1.2.1. Sumarização de Processos:

O judiciário brasileiro enfrenta um sério problema quanto à sobrecarga de processos que diariamente chegam aos tribunais de justiça. Com essa problemática, os juízes não conseguem atender todas as demandas e os processos acabam demorando mais para serem julgados e concluídos. Dentro desse contexto, pesquisadores da Universidade de Fortaleza (UNIFOR), instituição da Fundação Edson Queiroz, criaram a Inteligência Artificial (IA) SARA com o objetivo de aumentar a agilidade através da produção automática de resumos dos processos judiciais, facilitando o trabalho dos magistrados. Um resumo de um processo judicial é uma versão simplificada e concisa que destaca os principais elementos e informações do caso, como os pedidos das partes, fundamentos legais, decisões anteriores e a situação atual. A IA começou a ser inicialmente desenvolvida para o Supremo Tribunal de Federal (STF), que buscava uma forma de resumir os processos da Corte. Logo depois, a SARA passou a ser adaptada para fazer o mesmo com os processos que tramitam no Tribunal de Justiça do Ceará (TJCE). A IA é capaz de gerar resumos de qualidade destes processos, aumentando a eficiência de magistrados e servidores, dando maior celeridade aos serviços oferecidos pelo TJCE. Ao gerar resumos automáticos de qualidade, a IA SARA permite que magistrados e servidores acessem de forma rápida e clara os pontos principais de cada processo, reduzindo o tempo necessário para análise e tomada de decisão. Isso é especialmente relevante no contexto de sobrecarga de trabalho enfrentada pelo TJCE, onde o volume de processos é muito maior do que a capacidade de julgamento e movimentação atual. No ANEXO I – RELATO SARA, é apresentada a experiência positiva de magistrados do TJCE com uma prova de conceito da solução SARA. Deste modo, surge a necessidade de implantação de serviço de IA generativa no ambiente do TJCE para a absorção e implantação da solução em ambiente de produção do TJCE;

1.1.2.2. Classificação de Documentos:

O banco de dados do PJE 1º Grau (PJE1G) conta com uma grande quantidade de documentos de processos judiciais de diferentes tipos,



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

tais como petições iniciais, contestações, decisões etc. A classificação do tipo de um documento é de extrema relevância para que este seja entregue aos corretos interessados e processado pelas aplicações corretas. Contudo, atualmente o tipo de um documento é definido pelo usuário durante o processo de inclusão deste documento, sendo tal definição passível de falhas ou inexistente. Deste modo, o TJCE necessita de uma solução de IA capaz de classificar com precisão os diferentes tipos de documentos já armazenados e futuros, com a IA generativa auxiliando a classificação em casos ambíguos;

1.1.2.3. Busca Ativa de Jurisprudência:

Atualmente, o sistema de Busca de Jurisprudência (SJURIS) do TJCE realiza buscas com base em palavras-chave para encontrar acórdãos, decisões monocráticas e súmulas com base na existência destas palavras-chave nos corpos dos documentos. Esta busca permite que magistrados e servidores localizem documentos que possam ser relevantes para novos processos sendo julgados, ajudando a embasar a tomada de decisões. Contudo, o atual sistema de buscas consegue apenas encontrar documentos que possuam exatamente as palavras-chave em seu corpo, não tendo a capacidade de considerar sinônimos, proximidade semântica etc. Com a evolução planejada para integrar IA generativa, será possível identificar com maior acurácia as relações jurídicas presentes nos acórdãos e verificar sua compatibilidade com os casos em análise, oferecendo resultados mais relevantes e coerentes com o processo específico. Essa inovação contribuirá diretamente para a redução do tempo de pesquisa jurídica, aumentando a produtividade dos magistrados e servidores, além de fornecer suporte mais robusto na fundamentação das decisões judiciais;

1.1.2.4. Previsão de Precedentes:

O TJCE visa otimizar o agrupamento e análise de processos por temas de precedentes judiciais, principalmente relacionados a decisões de tribunais superiores, como o STJ. A IA generativa deve permitir a organização eficiente de processos, extraíndo informações relevantes das peças processuais e gerando resumos automáticos de forma estratégica, facilitando a visualização e compreensão dos grupos. Essa abordagem busca aumentar a agilidade, precisão e eficiência no tratamento de grandes volumes de dados jurídicos, contribuindo para



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

uma gestão mais inteligente da informação no TJCE.

- 1.1.3. Atualmente, o TJCE já vem desenvolvendo as soluções de IA supracitadas em parceria com instituições externas através do Programa Cientista Chefe pelo projeto intitulado “Ciência de Dados e Inteligência Artificial em Apoio às Atividades Judiciárias”, em parceria com a Universidade de Fortaleza (UNIFOR); e através do contrato “Modernização do Judiciário” realizado diretamente com pessoa física desenvolvedor de software pelo contrato Nº 105-2022 e o aditivo AD1/CT Nº 105/2022. A UNIFOR implementou as soluções de Sumarização de Processos (Sistema de Análise e Resumos de Ações - SARA), Classificação de Documentos e Busca Ativa de Jurisprudência, todas por meio do Programa Cientista Chefe. Já a “Modernização do Judiciário” desenvolveu a solução de Previsão de Precedentes. Estas soluções em desenvolvimento utilizam IAs generativas, tornando necessária a implantação deste tipo de tecnologia para a absorção e implantação destas soluções no ambiente de produção do TJCE.
- 1.1.4. A absorção urgente da solução SARA no TJCE é essencial para enfrentar a sobrecarga processual e a escassez de recursos humanos, que prejudicam a celeridade na tramitação de processos. A rápida disponibilização do serviço de IA generativa é um requisito prioritário, pois permitirá a automação de tarefas, como a geração de resumos processuais, acelerando o trabalho de magistrados e servidores. Essa solução visa reduzir os tempos de análise e julgamento, melhorando o fluxo de trabalho e a eficiência do atendimento judicial.
- 1.1.5. Portanto, a disponibilização do serviço de IA generativa é necessária para o atendimento da demanda de absorção das aplicações de IA no TJCE, além de ser essencial para o desenvolvimento e implantação eficaz de projetos, salvaguardando a posição de pioneirismo e excelência do TJCE no desenvolvimento tecnológico no Poder Judiciário e assegurando a relevância e impacto positivo das soluções tecnológicas na instituição, vide fonte <https://www.tjce.jus.br/noticias/tjce-atinge-nivel-maximo-de-excelencia-em-indice-de-governanca-de-tecnologia-da-informacao-do-cnj/>.
- 1.1.6. Resta evidenciada a necessidade da conclusão do processo de absorção e implantação das soluções de IA citadas no item 1.1.2, e, para tanto, exige a disponibilização de serviço de IA generativa para o TJCE.



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

- 1.1.7. Importante, para definir a solução para a necessidade efetiva que sustenta a demanda, essencialmente caracterizada como serviço, que sejam aprofundados os seguintes aspectos:
- 1.1.7.1. Periodicidade da necessidade: 36 meses.
 - 1.1.7.2. Locais da execução: o serviço de IA generativa deverá estar hospedado em uma plataforma de nuvem privada.
 - 1.1.7.3. Unidade de medida de consumo: Créditos de Serviços em Nuvem (CSN).
 - 1.1.7.4. Quantidade requerida aproximada: 909 CSN. O item 8 deste documento apresenta detalhes dos cálculos desta estimativa.
 - 1.1.7.5. Disponibilidade dos serviços: Os serviços devem estar disponíveis 24 horas por dia, 7 dias por semana, durante todo o período do contrato.
 - 1.1.7.6. Objetivando a conformidade com a Lei Geral de Proteção de Dados Pessoais (LGPD) (Redação dada pela Lei nº 13.853, de 2019), que garante a segurança e a privacidade de dados pessoais, bem como a natureza confidencial dos dados tratados no âmbito do TJCE, conforme a Resolução do Órgão Especial nº 40/2022, que estabelece a Política Geral de Privacidade e Proteção de Dados Pessoais (PGPPDP), visando assegurar a proteção e sigilo de dados no uso de tecnologias da informação dentro do Judiciário cearense, apresenta-se a necessidade da contratação de um provedor de serviços de IA generativa que disponibilize serviços privados e com recursos para garantir a segurança dos dados. Desta forma, é vedado o repasse dos dados fornecidos ao serviço para terceiros, bem como sua utilização no treinamento de modelos ou em serviços disponibilizados ao público externo ao TJCE.
- 1.1.8. Havendo o atendimento dessa demanda, o TJCE contará com o melhor serviço de IA generativa que seja compatível com as aplicações já sendo desenvolvidas e que garanta a segurança e privacidade dos dados. Enfatizando que, caso contrário, ocorreria o risco de incapacidade de manter o desenvolvimento e implantação de soluções avançadas de inteligência artificial e ciência de dados. Isso pode resultar em decisões menos embasadas, processos menos eficientes e a perda de oportunidades de inovação, o que poderia afetar negativamente a qualidade e disponibilidade dos serviços relacionados à atividade-fim do TJCE.



ESTADO DO CEARÁ
PODER JUDICIÁRIO
TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

1.2. DESCRIÇÃO DA NECESSIDADE TECNOLÓGICA

- 1.2.1. Atualmente, o desenvolvimento de soluções que utilizam serviço de IA generativa se operacionaliza por meio de parceria com a UNIFOR e através do contrato de “Modernização do Judiciário”, conforme descrito no item 1.1.3, ficando a cargo destes o provimento do serviço de IA generativa necessário. Portanto, é imperioso que a contratação deste serviço seja feita diretamente pelo TJCE para a absorção e implantação destas soluções no ambiente de produção do tribunal.
- 1.2.2. O serviço de IA generativa contratado deve ser compatível com as tecnologias já sendo utilizadas no desenvolvimento das soluções via parcerias, descritas no item 1.1.2, tendo em vista o custo oneroso para a adaptação, experimentação e avaliação das soluções, caso estas utilizem serviços de IA generativa diferentes. Atualmente, as tecnologias empregadas no desenvolvimento utilizam IAs generativas da família GPT, desenvolvidas pela *OpenAI* e fornecidas pela *Microsoft*.
- 1.2.3. É vedado o repasse dos dados fornecidos ao serviço para terceiros, bem como sua utilização no treinamento de modelos ou em serviços disponibilizados ao público externo ao TJCE.
- 1.2.4. Deve ser de responsabilidade do provedor de serviços de IA generativa ou fabricante o cumprimento dos seguintes requisitos:
 - 1.2.4.1. A fabricante deve ser responsável pela instalação e manutenção da infraestrutura necessária para a execução dos serviços, realizando redimensionamento e provisionamento das unidades computacionais e de armazenamento, balanceamento de carga, controle de fluxo de rede etc.
 - 1.2.4.2. A fabricante deve oferecer recursos de segurança que garantam a privacidade dos dados, tais como, *endpoints* de acesso privado, modelo de IA generativa privado, redes virtuais privadas etc.
- 1.2.5. O serviço de IA generativa adquirido deve ter sua qualidade e desempenho amplamente aceitos pela comunidade via *benchmarks*, *rankings*, *leaderboard*, ou quaisquer outros meios de avaliação mantidos por entidades independentes. O modelo oferecido deve alcançar as seguintes pontuações mínimas (ver seção 10.1.4.6) para seus respectivos *benchmarks*:
 - 1.2.5.1. *Massive Multi-task Language Understanding* (MMLU): 88%;



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

1.2.5.2. *HumanEval benchmark - Coding challenges*: 90%;

1.2.5.3. *Multilingual Understanding*: 90%.

2. ANÁLISE DE SOLUÇÕES ANTERIORES

2.1. Não há contratação anterior que seja compatível com a necessidade apresentada, assim não tendo parâmetros de contratações internas para comparação preliminar.

3. FORMAS DE ATENDIMENTO DA NECESSIDADE

3.1. Diante das particularidades da necessidade identificada, além de informações técnicas obtidas, foram consideradas, para a solução da necessidade identificada, os seguintes meios:

3.1.1. Contratação de serviço de IA generativa;

3.2. Analisadas as possíveis formas de solução para o atendimento interno da demanda, foram também promovidas medidas e consideradas outras opções de atendimento da demanda, tais como:

3.2.1. Remanejamento interno: Realocação de recursos ou equipes já existentes no TJCE para absorver a demanda sem a necessidade de contratação ou desenvolvimento de novas soluções, otimizando o uso de pessoal ou infraestrutura disponíveis. No entanto, o TJCE enfrenta uma realidade de sobrecarga processual, onde há um grande número de processos parados aguardando movimentação ou decisões. Essa situação é agravada pela insuficiência de mão de obra em relação à quantidade de processos em tramitação, resultando em longos períodos de espera para que as partes envolvidas obtenham uma resposta judicial. Esse atraso não só compromete o acesso à justiça, mas também eleva o volume de trabalho acumulado, perpetuando o ciclo de morosidade no sistema judicial;

3.2.2. Compartilhamento de outras soluções existentes: Utilização de soluções já implementadas em outras áreas do TJCE ou, se possível, adaptações de ferramentas ou sistemas já em uso que possam ser reutilizados para atender à nova demanda, evitando a criação de novos projetos. Embora já haja a



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

disponibilidade de serviço de IA generativa dentre o catálogo de soluções ofertadas pelos serviços de nuvem adotados pelo TJCE através de outras contratações, seu uso é limitado pelo número de créditos contratados disponíveis. No entanto, estes créditos foram dimensionados considerando apenas as demandas existentes em sua época de contratação. Portanto, hoje, estes são insuficientes para a grande demanda apresentada pelas atuais soluções de IA em desenvolvimento;

3.2.3. Retardamento ou atendimento provisório por solução alternativa: Adoção de uma solução temporária, que atenda parcialmente a demanda enquanto se desenvolve uma solução definitiva. Em alguns casos, pode-se optar por postergar o atendimento da demanda até que seja viável implementar uma solução mais robusta ou ajustada às necessidades específicas. No entanto, a absorção da solução SARA é uma demanda de caráter urgente no TJCE.

3.3. Ao final da análise, identificou-se que a melhor alternativa é a Contratação de serviço de IA generativa. Esta solução possibilita o tratamento imediato das demandas relacionadas sem a necessidade de contratação ou remanejamento de mão obra humana. Além disso, exime o TJCE da necessidade de aquisição, configuração e manutenção da infraestrutura necessária para execução dos serviços. Desta maneira, fica a cargo do provedor o rebalanceamento e a realocação dos recursos necessários para dar suporte à carga de trabalho apresentada pelas aplicações. Além disso, esta solução dá acesso a uma ampla gama de modelos de IA generativa privados, de outro modo inacessíveis por meios *open-source* e implantação local.

4. DESCRIÇÃO DA SOLUÇÃO COMO UM TODO

4.1. A presente demanda refere-se à contratação de serviço de IA generativa para o uso deste no desenvolvimento e implantação de aplicações de IA no TJCE.

4.2. O acesso ao serviço de IA generativa dá-se por meio de plataformas de computação em nuvem, tais como *Microsoft Azure*, *Amazon AWS*, *Google Cloud* etc. As empresas mantenedoras de tais plataformas são denominadas provedor de serviços em nuvem (*cloud provider*). Tais plataformas podem ser



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

assinadas diretamente com os provedores ou através de parceiros locais, denominados *cloud brokers*, responsáveis pela intermediação da distribuição dos serviços de plataformas de nuvem. Um *cloud broker* é um corretor de serviços de nuvem especializado em oferecer serviços em nuvem que se encaixam no perfil de cada cliente. Deste modo, a contratação de serviços em nuvem pelo TJCE será viabilizada pela contratação de uma empresa parceira (*cloud broker*) que intermediará o fornecimento destes serviços.

- 4.3. Serviços de IA generativa utilizam o modelo de pagamento *Pay-As-You-Go* (PAYG), onde o pagamento é feito conforme o uso. Esses serviços são cobrados com base na quantidade de *tokens* (*palavras*) processados. O pagamento dos serviços contratados via um *cloud broker* pode ser feito através de créditos de serviços em nuvem. Cada unidade de crédito equivale a um valor monetário a ser consumido num período específico para serviços em nuvens que sejam utilizados. Este valor é composto de um valor base, estipulado pela provedora de serviços de nuvem, multiplicado por um *valor de markup* para cada fornecedor parceiro, que inclui os custos e margem de lucro da empresa. O valor de *markup* é um índice utilizado para determinar o preço de venda de um produto ou serviço a partir do seu custo. Ele representa a diferença entre o custo de produção (ou aquisição) e o preço final que será cobrado do consumidor, geralmente expressa como uma porcentagem sobre o custo. O valor de *markup* também se aplica aos serviços a serem consumidos, sendo sempre mais vantajosa para o órgão contratante a compra da unidade de crédito com o preço mais barato.
- 4.4. As soluções apresentadas no item 1.1.2 já estão em desenvolvimento no âmbito de parcerias com instituições externas, como esclarecido no ponto 1.1.3, e serão absorvidas para serem implantadas no ambiente do TJCE, sendo estas prioritárias em um primeiro momento.
- 4.5. A seguir estão listados alguns exemplos de propostas de soluções que podem vir a ser desenvolvidas no TJCE e que serão beneficiadas com esta contratação:
- 4.5.1. Assistente virtual para perguntas e respostas contextualizadas: Atua sobre um documento específico selecionado pelo usuário, respondendo perguntas de acordo apenas com o conteúdo do documento, visando o



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

auxílio da compreensão e a rápida localização de informações específicas.

- 4.5.2. Assistente virtual para a escrita semiautomatizada de documentos de contratação: considerando o tipo do documento a ser escrito, o sistema solicita ao usuário as informações necessárias para preencher o padrão conhecido do documento, gerando como resultado a sugestão de texto para o documento com base em documentos similares já previamente criados.
- 4.5.3. Sistema de extração de informações estruturadas de contratos: o sistema minera e extrai informações de contratos, gerando representações estruturadas destes. Tais informações podem ser utilizadas por sistemas já existentes para arquivar e gerenciar novos contratos.
- 4.5.4. Sistema de validação e higienização de dados: o sistema minera e extrai informações estruturadas de documentos não estruturados entrados pelo usuário. Tais informações são então comparadas com as informadas pelo usuário no ato de inserção do documento, indicando possíveis inconsistências e solicitando sua revisão.
- 4.5.5. Assistente virtual para busca de processos similares: indica processos e seus trechos similares já cadastrados. Tal informação pode ser utilizada para confirmar a intenção do autor de incluir um novo processo em litispendência e/ou identificar possíveis casos de perempção.
- 4.5.6. Sistema de recomendação para soluções em processo de mediação/conciliação: sistema para auxiliar o servidor/juiz, apresentando situações prévias semelhantes e recomendando possíveis soluções.
- 4.5.7. *Chatbot* para a Central de Atendimento Judicial (CAJ): solução inteligente utilizando técnicas de *Retrieval-Augmented Generation* (RAG) sobre a base da CAJ para gerar um *Chatbot* humanizado para a central de atendimento, atendendo demandas primárias dos usuários.



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

4.6. Desta forma, o fornecimento identificado para suprir a necessidade objeto deste estudo se relaciona indiretamente com a atividade-fim do TJCE, pois diz respeito ao objetivo do plano estratégico: prover soluções de TIC inovadoras e integradas para a transformação digital, que é essencial para a prestação dos serviços atribuídos ao TJCE.

5. ANÁLISE DE SOLUÇÕES POSSÍVEIS

5.1. Para a contratação em questão, foram analisados processos similares anteriores, feitos pelo TJCE e por outros órgãos e entidades, como também pesquisa de oferta de soluções do mercado, com objetivo de identificar as diversas possibilidades e a existência de novas metodologias, tecnologias ou inovações que melhor supririam as necessidades do TJCE.

5.2. ANÁLISE COMPARATIVA DE SOLUÇÕES

5.2.1. Solução A: Aditivar contratação de fornecimento de créditos do *Microsoft Azure Prepayment* já existente no TJCE.

5.2.2. Descrição da solução A: Atualmente, o TJCE já possui contratos com a plataforma de nuvem *Microsoft Azure* por meio do contrato N° 54/2023 do TJCE, Solução de Infraestrutura para o Balcão Virtual do TJCE. Esta contratação resultou na aquisição de 15 créditos *Microsoft Azure Prepayment (SKU 6QK-00001)*, totalizando o valor de R\$ 359.010,90. Os créditos *Microsoft Azure Prepayment (SKU 6QK-00001)* permitem acesso aos mais de 200 serviços disponíveis no catálogo da plataforma *Microsoft Azure*, localizados no link <https://azure.microsoft.com/pt-br/products>. Dentre estes serviços, encontra-se o *Azure OpenAI Service*, que disponibiliza os modelos de IA generativa da família GPT de maneira privada para clientes da plataforma. O serviço *Azure OpenAI* fornece acesso à API REST para modelos de linguagem avançados do *OpenAI*, incluindo GPT-4o, GPT-4 Turbo com Visão, GPT-4, GPT-3.5-Turbo e as séries de modelos de incorporações. Esses modelos podem ser facilmente adaptados a tarefas específicas, incluindo, entre outros, geração de conteúdo, sumarização, reconhecimento de imagem, pesquisa semântica e tradução de linguagem natural para código. Os usuários podem acessar o serviço por meio de APIs REST, SDK do Python ou a partir da interface



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

baseada na *Web* no Estúdio do *OpenAI* do *Azure*. Portanto, a solução mais simples para o atendimento da atual demanda seria o compartilhamento dos créditos já adquiridos para a solução do Balcão Virtual. Contudo, estes créditos foram dimensionados considerando apenas as demandas existentes em sua época de contratação para a aplicação do Balcão Virtual. Por consequência, hoje, estes são insuficientes para a grande demanda apresentada pelas atuais soluções de IA em desenvolvimento. Por sua vez, a aditivação da contratação em até 25% do valor do contrato, para adicionar mais créditos, obteria no máximo um total de 3,75 créditos *Microsoft Azure Prepayment*, que seria insuficiente para suprir a atual demanda das soluções de IA, tendo em vista o quantitativo necessário exposto no item 8. Diante disso, a opção por esta solução não atenderia nossa demanda e resultaria na necessidade de futuras contratações.

- 5.2.3. Solução B: Contratação de empresa certificada pela *Microsoft* como *government partner* (GP) para fornecimento de serviço de IA generativa através da plataforma *Microsoft Azure*.
- 5.2.4. Descrição da solução B: A *Microsoft* tem uma política de comercialização para a Administração Pública, na qual os parceiros, *Government Partners* – GP, são habilitados para atuar no setor público com o objetivo de assinar contratos nos modelos dos clientes, de forma que estes tenham total capacidade para participar de processos licitatórios junto ao poder público, conforme se verifica no Portal da *Microsoft*, na página de Parceiros LSP, disponível em <https://partner.microsoft.com/pt-br/licensing/Parceiros%20LSP>. Onde verifica-se que a empresa possui uma política interna de não conceder registro de oportunidade para parceiros comerciais. Como já explicado na solução A, o serviço *Azure OpenAI* disponibiliza os modelos de IA generativa já utilizados durante o desenvolvimento das soluções apresentadas no item 1.1.2. Além disso, com o *Azure OpenAI*, os clientes obtêm os recursos de segurança do *Microsoft Azure* enquanto executam os mesmos modelos que os da *OpenAI*. O *Azure OpenAI* oferece rede privada (*VNETS*), *Endpoints* de acesso privados, disponibilidade regional e filtragem de conteúdo de IA responsável. Além disso, o serviço conta com uma fácil integração a todos os outros serviços providos pela *Microsoft* e uma ampla rede de suporte. O serviço está



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

disponível no link <https://azure.microsoft.com/pt-br/products/ai-services/openai-service>. Portanto, uma solução viável para a atual demanda é a contratação de um GP local da *Microsoft* para fornecer a totalidade do serviço de IA generativa *Azure OpenAI Service* necessário para a atual demanda, incluindo os recursos da plataforma *Microsoft Azure* que permitirão desenvolver, testar, implantar e monitorar as aplicações que fazem uso da IA generativa.

- 5.2.5. Solução C: Implantação de serviço de IAs generativas *open-source*.
- 5.2.6. Descrição da solução C: IAs generativas *open-source* oferecem uma alternativa poderosa e flexível às soluções proprietárias. Com acesso total ao código-fonte, as empresas podem adaptar e personalizar esses modelos conforme suas necessidades específicas, promovendo inovação e independência tecnológica. Pode ser mais econômico do que soluções proprietárias, especialmente para empresas que possuem recursos internos para implementar e manter esses modelos. Além disso, a ausência de taxas de licenciamento oferece uma economia significativa a longo prazo. Com o controle total sobre o código-fonte, as empresas podem adaptar as IAs generativas para atender aos requisitos de segurança e conformidade específicos do seu setor, garantindo que os dados sejam tratados de acordo com os mais altos padrões. O modelo mais popular de IA generativa *open-source* da atualidade é o Llama, desenvolvido pela empresa Meta (antiga Facebook), disponível no link <https://llama.meta.com/>.
- 5.2.7. Solução D: Contratação de um serviço de IA generativa de um fabricante diferente da *Microsoft*.
- 5.2.8. Descrição da solução D: Além da *Microsoft*, empresas como *Google*, *Amazon* e *OpenAI* oferecem serviços de nuvem com acesso a serviços de IA generativa. O *Vertex AI* é uma plataforma integrante do *Google Cloud* que oferece acesso aos modelos do Google, tais como, o *Gemini*, *Imagen* e *Codey*; bem como modelos de terceiros (família de modelos Claude da *Anthropic*) e de código aberto (*Gemma* e *Llama 3.1*). O serviço está disponível no link <https://cloud.google.com/vertex-ai>. O *Amazon Bedrock* é um serviço totalmente gerenciável que oferece várias opções de modelos de base de alta performance das principais empresas de IA, como *AI21 Labs*, *Anthropic*, *Cohere*, *Meta*, *Mistral AI*, *Stability AI* e *Amazon*, por meio de uma única API, além de um amplo conjunto de recursos necessários para criar aplicações de IA generativa com segurança, privacidade e IA



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

responsável. O serviço está disponível no link <https://aws.amazon.com/pt/bedrock/>. O *OpenAI Enterprise* é a versão corporativa das soluções de inteligência artificial da OpenAI, voltada para grandes empresas que buscam incorporar IA de ponta em suas operações. Essa plataforma oferece acesso a modelos de IA avançados, como o GPT-4, GPT-4o e GPT-4o mini, com maior capacidade de processamento, segurança de nível empresarial, conformidade com regulamentações e personalização para necessidades específicas de negócios. O serviço está disponível no link <https://openai.com/chatgpt/enterprise/>.

5.3. REGISTRO DE SOLUÇÕES CONSIDERADAS INVIÁVEIS

5.3.1. Dentre as opções consideradas, foram identificadas inviáveis as seguintes:

5.3.1.1. Implantação de serviço de IAs generativas *open-source* (Solução C)

5.3.1.1.1. Esta solução é inviável porque não possibilita o acesso aos modelos proprietários considerados os mais avançados atualmente, além da imposição da exigência de o TJCE instalar e manter infraestrutura do serviço internamente. A seguir, encontra-se uma justificativa detalhada para a inviabilidade da adoção de soluções *open-source*:

5.3.1.1.1.1. IAs generativas exigem recursos computacionais avançados, como servidores potentes, unidades de processamento gráfico (GPUs) para acelerar a computação e sistemas de armazenamento de alta capacidade. Além disso, é necessário manter uma infraestrutura de software robusta para suportar o processamento em larga escala.

5.3.1.1.1.2. A implementação de IA generativa demanda uma equipe com conhecimento especializado em linguagem natural, aprendizado de máquina e processamento de texto. Isso pode exigir o treinamento da equipe atual ou a contratação de especialistas externos para ajustar e otimizar os modelos.

5.3.1.1.1.3. Os modelos de IA generativa precisam de grandes conjuntos de dados para gerar textos de qualidade. Obter e preparar esses dados pode ser difícil, especialmente em áreas específicas, exigindo esforço significativo para coleta,



ESTADO DO CEARÁ
PODER JUDICIÁRIO
TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

limpeza e preparação.

5.3.1.1.1.4. O treinamento de uma IA generativa é demorado e intensivo em recursos, podendo levar semanas ou meses, dependendo da complexidade do modelo e da infraestrutura. O processo demanda monitoramento contínuo e ajustes ao longo do treinamento.

5.3.1.1.1.5. IAs generativas precisam ser constantemente atualizadas devido aos avanços regulares na área. Isso exige um esforço contínuo para garantir que os modelos estejam atualizados e alinhados com as melhores práticas.

5.3.1.2. Contratação de um serviço de IA generativa de um fabricante diferente da *Microsoft* (Solução D)

5.3.1.2.1. Durante o processo de desenvolvimento das soluções apresentadas no item 1.1.2, foi adotada a IA generativa da família GPT fornecida pelos serviços disponibilizados na plataforma *Microsoft Azure*. Os custos envolvidos no desenvolvimento das soluções em parceria com a UNIFOR foram de R\$ 1.824.150,00 durante 24 meses, havendo também a necessidade da participação de juízes do TJCE na etapa de avaliação qualitativa dos resultados. Por sua vez, os custos envolvidos no desenvolvimento da solução através do contrato “Modernização do Judiciário” foram de R\$ 246.952,59 durante 13 meses. Portanto, a somatória dos investimentos para o desenvolvimento das atuais soluções é totalizada em R\$ 2.071.102,59 por 37 meses.

5.3.1.2.2. Segundo análise da equipe técnica de desenvolvimento, a mudança do serviço de IA generativa utilizado exigiria adaptações nas soluções já desenvolvidas. Tais adaptações não são limitadas a simples alteração de linhas de código, incidindo também a necessidade de realização de novos experimentos e reavaliação qualitativa dos resultados, gerando assim custos monetários e temporais para os desenvolvedores e juízes envolvidos.

5.3.1.2.3. Segundo análise da equipe técnica de desenvolvimento, os custos necessários para adaptação e validação das soluções, caso ocorra a mudança de modelo de IA generativa, é prevista em



ESTADO DO CEARÁ
PODER JUDICIÁRIO
TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

50% dos atuais custos de desenvolvimento, tendo em vista a alta demanda de tempo nas tarefas de experimentação e validação dos resultados, vide ANEXO II - CUSTO TROCA LLM.

5.3.1.2.4. Esta solução foi considerada inviável, pois os custos de adaptação são estimados em R\$ 1.035.551,29 por 19 meses. Custos esses que ainda serão acrescidos aos custos de manutenção do serviço de IA generativa.

5.3.1.2.5. A média de preço dos serviços de IA generativa concorrentes é USD 0,0027 e USD 0,012 para 1.000 *tokens* de entrada e saída respectivamente, como apresentado na Tabela 1. Considerando este preço médio, resultará num custo total de R\$ 7.127.580,00.

5.3.1.2.6. O custo médio estimado para manutenção do serviço de IA generativa de outros fabricantes (R\$ 7.127.580,00) somado aos custos para adaptação (R\$ 1.035.551,29) resultaria num custo esperado de R\$ 8.163.131,00, ultrapassando os valores estimados para o uso da solução *Microsoft Azure* apresentados no item 8.

Serviços de IA Generativa	Preço por 1.000 Tokens de entrada (USD)	Preço por 1.000 Tokens de saída (USD)
Azure OpenAI Service GPT 4-o	0,0025	0,010
OpenAI GPT 4-o	0,0025	0,010
Google Gemini 1.5 Pro	0,0025	0,010
Anthropic Claude 3.5 Sonnet	0,003	0,015
Média excluindo Azure OpenAI Service	0,0027	0,012



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

Tabela 1: Comparação de preços entre serviços de IA Generativa. Fonte: Preços extraídos dos sites oficiais dos serviços no dia 06/02/2025.

5.3.1.2.7. Além disso, o custo adicional de tempo de 19 meses para adaptação e experimentação das soluções de IA em desenvolvimento é um ponto digno de atenção ao ser optado por esta solução, tendo em vista a urgência da necessidade para absorção da solução SARA.

6. PREVISÃO DA CONTRATAÇÃO NO PLANO DE CONTRATAÇÃO ANUAL

6.1. A contratação ora pretendida está em consonância com os objetivos estratégicos deste TJCE (conforme Planejamento Estratégico 2025), visto que prevê prover soluções de TIC inovadoras e integradas para a transformação digital, o que é imprescindível para o funcionamento do TJCE no desempenho de suas atividades institucionais.

6.2. O objeto da contratação está previsto no Plano de Contratações Anual 2025, especificamente no Código da Contratação TJCESETIN_2025_0026.

7. REQUISITOS DA CONTRATAÇÃO

7.1. A empresa deve possuir estrutura e experiência em fornecimentos compatíveis com objeto demandado;

7.2. Nos casos de fornecimentos, ou parte deles, controlados ou de exercício mediante autorização prévia, caberá à empresa a regularização e obtenção de respectiva(s) licença(s) ou registro(s);

7.3. No caso de produtos de mercado restrito, a empresa deverá certificar-se, ainda antes de eventual participação em licitação ou contratação, de que possui fabricantes ou fornecedores aptos ao tipo de objeto requerido nesta demanda.

7.4. Comprovar, como condição prévia à assinatura do contrato e para a manutenção contratual, o atendimento das seguintes condições:



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

- 7.5. Não possuir inscrição no cadastro de empregadores flagrados explorando trabalhadores em condições análogas às de escravo, instituído pela Portaria Interministerial MTPS/MMIRDH Nº 4 DE 11/05/2016;
- 7.6. Não ter sido condenada, a EMPRESA ou seus dirigentes, por infringir as leis de combate à discriminação de raça ou de gênero, ao trabalho infantil e ao trabalho escravo, em afronta a previsão aos artigos 1º e 170 da Constituição Federal de 1988; do artigo 149 do Código Penal Brasileiro; do Decreto nº 5.017, de 12 de março de 2004 (promulga o Protocolo de Palermo) e das Convenções da OIT nos 29 e 105;

8. ESTIMATIVAS DE QUANTIDADE

- 8.1. Na observância do volume da necessidade e seu detalhamento, foram considerados:
- 8.1.1. Projetos em desenvolvimento pela unidade demandante e implantação de soluções;
 - 8.1.2. Históricos de demandas internas;
 - 8.1.3. Perspectiva de novos projetos;
- 8.2. Atualmente, os provedores de serviços de IAs generativas adotam o modelo de pagamento *Pay-As-You-Go* (PAYG), ou pagamento conforme o uso. Neste modelo o pagamento dos serviços é realizado conforme a quantidade de recursos utilizados. O preço do serviço é calculado de acordo com a quantidade de *tokens* dados como entrada e gerados como saída. Nesse contexto, os *tokens* são unidades fundamentais de texto utilizadas como entrada pelos modelos de *Large Language Models* (LLMs). Deste modo, a quantidade necessária do serviço deve ser medida em função do número de *tokens* a serem tratados nas aplicações apresentadas no item 4 deste documento.
- 8.3. A análise detalhada das estimativas de volumetria e custos está apresentada no ANEXO III - VOLUMETRIA. Neste ETP, utilizamos o custo máximo esperado para a execução unitária das aplicações multiplicado pela quantidade de execuções necessárias tanto para o processamento da carga de trabalho pendente (dados históricos), quanto para a nova carga de trabalho projetada



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

para os próximos 36 meses, referentes às aplicações analisadas. O objetivo desta metodologia é evitar a falta de recursos necessários para fornecer o serviço de IA generativa. A análise de custos considera a tabela de preços fornecida pelo *Azure OpenAI Service* da Microsoft, disponível em <https://azure.microsoft.com/pt-br/pricing/details/cognitive-services/openai-service/>, e obtida no dia 06/02/2025.

8.4. Portanto, considerando uma análise técnica das estatísticas de uso das aplicações em estágio de implementação e de outras soluções similares implementadas no TJCE, a melhor estimativa de *tokens* da carga de trabalho final é estimada em 411.130.285.867 *tokens* de entrada e 13.928.354.702 *tokens* de saída, equivalentes à R\$ 6.849.764,20, considerando a calculadora de preços fornecida pelo *Azure OpenAI Service* da Microsoft, disponível no link <https://azure.microsoft.com/en-au/pricing/calculator/>, e obtida no dia 06/02/2025, pois:

Aplicação	<i>Tokens</i> de entrada	<i>Tokens</i> de saída	Custo (R\$)
SARA	361.307.410.191	13.742.450.163	R\$ 6.107.827,36
Classificador de documentos	39.412.450.696	5.579.339	R\$ 578.606,60
Previsão de precedentes	5.476.012.980	39.342.000	R\$ 82.655,77
Busca ativa por jurisprudência	4.934.412.000	140.983.200	R\$ 80.674,45
Total	411.130.285.867	13.928.354.702	R\$ 6.849.764,20

8.5. Esta demanda pode ser convertida em créditos *Microsoft Azure* utilizando o preço médio anual praticado pelas parceiras locais de R\$ 8.670,58, conforme aferido no MAPA DE PREÇOS.

8.6. Por fim, é possível concluir que a estimativa de quantidade demandada é de aproximadamente 790 unidades de créditos *Microsoft Azure*, equivalentes a R\$ 6.849.764,20. No entanto, é importante esclarecer que esta quantidade foi estimada com base em estatísticas de desenvolvimento e uso preliminar das



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

soluções de IA. Diante disso, não é possível prever com precisão a taxa de adoção e uso futuro destas e de outras soluções que usarão o serviço de IA generativa. Portanto, os quantitativos indicados são meramente estimativos, não gerando para o TJCE a obrigação de pagamento por quantidades não solicitadas.

9. ANÁLISE COMPARATIVA DE CUSTOS (TCO)

9.1. Cálculo dos Custos Totais de Propriedade

9.1.1. Solução A - Aditivar contratação de fornecimento de créditos do *Microsoft Azure Prepayment* já existente no TJCE.

9.1.1.1. Para o cálculo do *Total Cost of Ownership* (TCO), consideramos os seguintes aspectos:

9.1.1.2. Ao optar pela solução A, os custos envolvidos inicialmente seriam apenas os custos dos novos créditos *Microsoft Azure Prepayment*.

9.1.1.3. Contudo, tendo em vista a impossibilidade do suprimento da quantidade total demandada pelas soluções de IA em desenvolvimento, a opção por esta solução representaria a necessidade de futuras novas contratações.

9.1.1.4. Foi considerado o valor médio de créditos obtidos atualmente no mercado presente no Mapa de preços de R\$ 8.670,58.

9.1.2. Solução B - Contratação de empresa certificada pela *Microsoft* como *Government Partner* (GP) para fornecimento de serviço de IA generativa através da *plataforma Microsoft Azure*.

9.1.2.1. Para o cálculo do TCO, consideramos os seguintes aspectos:

9.1.2.1.1. Ao optar pela solução B, é necessário computar apenas os custos envolvidos inicialmente seriam apenas os custos dos novos créditos *Microsoft Azure Prepayment*.

9.1.2.1.2. Foi considerado o valor médio de créditos obtidos atualmente no mercado presente no Mapa de preços de R\$ 8.670,58.



ESTADO DO CEARÁ
PODER JUDICIÁRIO
TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

9.1.3. Solução C - Implantação de serviço de IAs generativas *open-source*.

9.1.3.1. Para o cálculo do *Total Cost of Ownership* (TCO), consideramos os seguintes aspectos:

9.1.3.2. Ao optar pela solução C, os custos envolvidos inicialmente são os custos de obtenção dos equipamentos, tais como Graphics Processing Unit (GPU) ou Unidades de Processamento Gráfico, Unidades de Armazenamento, etc. Contudo, também é necessário considerar os custos da manutenção do equipamento, tais como, eletricidade, equipe de suporte técnico, etc. Por conta da complexidade do cálculo envolvido no custo da manutenção, neste documento será avaliado apenas o custo dos equipamentos.

9.1.3.3. Foram consideradas, apenas as GPUs, sendo estas os equipamentos mais críticos necessários. O modelo considerado é conforme a análise do Blog: <https://www.eddieoz.com/quais-gpus-podem-rodar-llama-3-1-405b/>

9.1.3.3.1. 8 x GPUs NVIDIA H100 80GB no modo de 8 bits. Com o valor unitário médio de cada GPU sendo, conforme pesquisa no Google R\$ 253.666,00 realizada em 16/12/2024 apresentada na Figura 1.

Produto	Preço	Entrega
Placa de vídeo NVIDIA H100 (80GB, PCIe 5.0) - PNY NVH100TCGPU-KIT PCI Express - PNY	R\$ 259.999,99	Entrega gratuita
NVIDIA H100 Placa Gráfica, 80GB HBM2e Memória, Deep Learning GPU HDMI - HBM2 - Placa gráfica	R\$ 593.984,00	Entrega de R\$ 50,00
Placa de vídeo NVIDIA H100 NVL (94GB, PCIe 5.0) - NVH100NLTCGPU-KIT PCI Express - PNY	R\$ 259.999,99	Entrega gratuita
Nvh100tcgpu-kit Pny Placa de Vídeo H100 Tensor Core 80gb Hbm3 PCI Express - PNY - HBM	R\$ 134.680,50	Frete não incluído



ESTADO DO CEARÁ
PODER JUDICIÁRIO
TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

Figura 1. Preços da GPU NVIDIA H100 80GB encontrados no Google em 16/12/2024.

9.1.4. Solução D - Contratação de um serviço de IA generativa de um fabricante diferente da Microsoft.

9.1.4.1. Para o cálculo do *Total Cost of Ownership* (TCO), consideramos os seguintes aspectos:

9.1.4.2. Ao optar pela solução D, são envolvidos os custos para adaptação das atuais soluções, além dos custos necessários para manutenção do serviço de IA generativa.

9.1.4.3. Os custos de adaptação são estimados em R\$ 1.035.551,29 por 19 meses, como apresentado no item 5.3.1.2.4.

9.1.4.4. A Tabela 1 no item 5.3.1.2.6 apresenta a comparação dos custos para a manutenção dos serviços de IA generativa de diversos fabricantes. Nestes cálculos serão utilizados os custos médios no caso da opção de fabricantes diferentes da *Microsoft*.

9.1.4.4.1. Custo por *tokens* de entrada do modelo, onde \$ 0,0027 USD são cobrados por 1000 *tokens*.

9.1.4.4.2. Custo por *tokens* gerados na saída do modelo, onde \$ 0,012 USD são cobrados por 1000 *tokens*.

9.1.4.4.3. A cotação do Dólar adotada nestes cálculos foi 1 USD = R\$ 5,56.

9.2. Mapa Comparativo dos Cálculos Totais de Propriedade (TCO)

9.2.1. Considerando as diversas formas para atender a necessidade descrita neste documento, foram considerados os respectivos valores aproximados para o fornecimento, abaixo listados, que indicam como razoável a estimativa em torno de R\$ 7.881.557,22, pois:

9.2.1.1. A solução A, que envolve o aditivo de 25% da contratação do Balcão Virtual, resultaria em 3,75 *Microsoft Azure Prepayment* – 6QK-00001, sendo considerados 4 unidades, tendo em vista que os créditos são indivisíveis. Esta quantidade de créditos resultaria em média um custo estimado de R\$ 34.682,32. Contudo, esta opção não



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

seria capaz de suprir toda demanda apresentada no item 8.

- 9.2.1.2. A solução B, que envolve a Contratação de empresa certificada pela *Microsoft* como *government partner* (GP) para o fornecimento de IA generativa no volume máximo de, aproximadamente, 790 créditos *Microsoft Azure*, resultaria numa estimativa de custo, aproximada, de R\$ 6.849.764,20.
- 9.2.1.3. A solução C, que envolve a Implantação de serviço de IAs generativas *open-source*. Considerando a quantidade mínima de equipamentos necessários, o custo da aquisição é estimado em aproximadamente R\$ 2.029.328,00 para as 8 unidades de GPUs NVIDIA H100 80GB. Além disso, seria necessário também considerar os preços de outros equipamentos, tais como fontes de alimentação, unidades de armazenamento, etc., além dos custos de manutenção, tais como, eletricidade, equipe de suporte técnico, etc. No entanto, tal solução não possibilita o acesso às IAs da família GPT, indicando também a necessidade da adaptação do código, incidindo no valor de R\$ 1.035.551,29 por 19 meses. Portanto, o custo mínimo desta solução pode ser avaliado em R\$ 3.064.879,29, embora seja impossível avaliar no momento a infraestrutura necessárias com exatidão, além do seu preço de manutenção.
- 9.2.1.4. A solução D, que envolve a Contratação de um serviço de IA generativa de um fabricante diferente da *Microsoft* no volume máximo de, aproximadamente, 790 créditos *Microsoft Azure*, resultaria numa estimativa de custo, aproximada, de R\$ 8.163.131,00, ao considerar o custo médio estimado para manutenção do serviço de IA generativa de outros fabricantes (R\$ 7.127.580) somado aos custos para adaptação (R\$ 1.035.551,29).
- 9.2.1.5. Assim, a estimativa que melhor se alinha aos requisitos desta contratação, em conformidade com os princípios de economicidade, é apresentada pela Solução B, no valor de R\$ 6.849.764,20 para a aquisição de 790 unidades de créditos *Microsoft Azure*. Entretanto, considerando o aumento contínuo no volume de dados e a crescente demanda por soluções de inteligência artificial no Tribunal de Justiça do Estado do Ceará (TJCE), recomenda-se a inclusão de um percentual de crescimento de 15%. A aplicação deste percentual resulta em uma necessidade de 908,5 créditos *Microsoft Azure*, tendo



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

em vista a indivisibilidade do objeto, é necessário proceder ao arredondamento para 909 unidades. Conseqüentemente, uma estimativa razoável de custo totaliza R\$ 7.881.557,22 para 909 créditos Azure.

10. SOLUÇÃO ESCOLHIDA

10.1. Após as análises das particularidades da necessidade e das possibilidades de atendimento, identificou-se como a melhor opção para solução da necessidade a “Contratação de empresa certificada pela *Microsoft* como *government partner* (GP) para fornecimento de serviço de IA generativa através da plataforma *Microsoft Azure*” apresentado como “solução B” no item 5.2.3. Com o intento de confrontar não apenas vantagens pecuniárias, mas, de forma primordial, o atendimento aos requisitos descritos no item “Requisitos de Negócio da Área Requisitante – Necessidade de Negócio e Necessidade Tecnológica” e na já existência de contratações ao serviço dentro do TCJE, de fato, o aporte para a contratação de serviço de IA generativa disponibilizada através da plataforma *Microsoft Azure* apresenta significativos ganhos tecnológicos e administrativos, pois esta solução apresenta as seguintes vantagens que embasaram sua escolha:

10.1.1. Embora a Solução A seja inicialmente viável, a opção por esta solução não seria capaz de suprir toda a demanda apresentada no item 8, efetuando a contratação de apenas 4 créditos dos 790 necessários. Logo, tornando necessárias futuras contratações. Portanto, a opção pela solução B gera economia ao dispensar a necessidade de elaboração de futuras contratações, simplificando o processo de gerenciamento de múltiplas contratações, além de fixar os preços atuais de mercado que estão propensos a sofrerem correções, ocasionado seu aumento de preço;

10.1.2. As Soluções C e D não forneceria o acesso aos modelos de IA generativa requisitados pelas soluções de IA já desenvolvidas, acarretando nos de adaptação.

10.1.3. Além disso, a manutenção do uso dos serviços da fabricante *Microsoft*, em contraste as Soluções C e D, oferece as seguintes vantagens:



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

- 10.1.3.1. Evitaria o desperdício de recursos, tendo em vista que já houve investimentos de R\$ 2.071.102,59 por 37 meses utilizando os serviços de IA generativa desta fabricante;
- 10.1.3.2. Além disso, o custo médio estimado para manutenção dos serviços de IA generativa de outros fabricantes (R\$ 7.127.580,00) somado aos custos para adaptação (R\$ 1.035.551,29) resultaria em um custo esperado de R\$ 8.163.131,00, ultrapassando os valores estimados para o uso da solução *Azure OpenAI Service* apresentados no item 8.
- 10.1.3.3. Além disso, o custo adicional de tempo de 19 meses para adaptação e experimentação das soluções de IA em desenvolvimento é um ponto digno de atenção, tendo em vista a urgência da necessidade para absorção da solução SARA.
- 10.1.4. Como vantagens técnicas da IA generativa da fabricante *Microsoft*, temos:
- 10.1.4.1. É compatível com as soluções das demandas já sendo desenvolvidas, evitando o custo oneroso para a adaptação, experimentação e avaliação das soluções, caso estas utilizem serviços de IA generativa diferentes;
- 10.1.4.2. Rápida implantação das soluções já desenvolvidas, tendo em vista que haveria apenas a necessidade da alteração da chave de acesso da API do serviço de IA generativa;
- 10.1.4.3. Já há contratação ao serviço de nuvem *Microsoft Azure* no TJCE;
- 10.1.4.4. Fácil integração com os serviços da *Microsoft* já utilizados no TJCE;
- 10.1.4.5. Funcionalidades da *Microsoft Azure* que permitem desenvolver, testar, implantar e monitorar as aplicações que fazem uso da IA generativa;
- 10.1.4.6. Proporciona avançados recursos de segurança e privacidade, tais como, *Endpoints* privados e *VNETs* para o acesso aos serviços de maneira segura, garantindo a privacidade dos dados. O *Microsoft*



ESTADO DO CEARÁ PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

Azure se revela uma plataforma de computação em nuvem que cumpre rigorosamente os requisitos de segurança do governo brasileiro, abrangendo inclusive a Lei Geral de Proteção de Dados (LGPD). Sua robustez na área de segurança é atestada por certificações emitidas pelo Instituto Nacional de Tecnologia da Informação (ITI) e do Conselho Nacional de Segurança da Informação (CNSI).

10.1.4.7. Proporciona acesso a uma variedade de modelos e aplicações de IA generativa, tais como, aos modelos da família GPT. Os modelos da família GPT são considerados atualmente os mais avançados conforme o *benchmark* independente “*Artificial Analysis*”, disponível no link <https://artificialanalysis.ai/leaderboards/models>. O *leaderboard* que agrega e compara os relatórios técnicos dos modelos disponíveis, gerenciado pela instituição *Vocify Inc.*, disponível no link <https://www.vellum.ai/llm-leaderboard>, também atesta os modelos da família GPT como os melhores em cargas de trabalho multitarefas e matemáticas, sendo o segundo melhor em tarefas de codificação. Abaixo, são apresentadas comparações do mais atual modelo da família GPT, o GPT-4o, disponibilizadas na atual data (11/09/2024) que fundamentam estas conclusões:

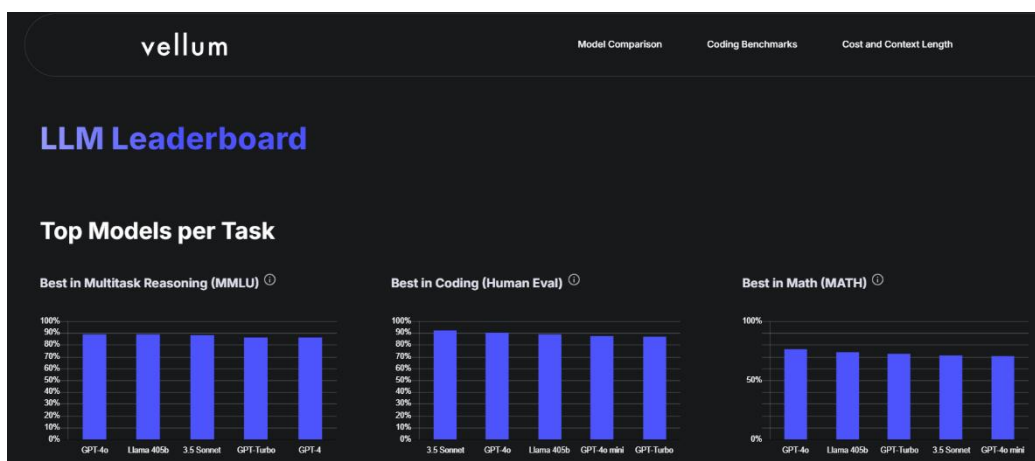


Figura 3: Comparação geral dos principais modelos de IA generativa. Fonte: <https://www.vellum.ai/llm-leaderboard>



ESTADO DO CEARÁ
PODER JUDICIÁRIO
TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

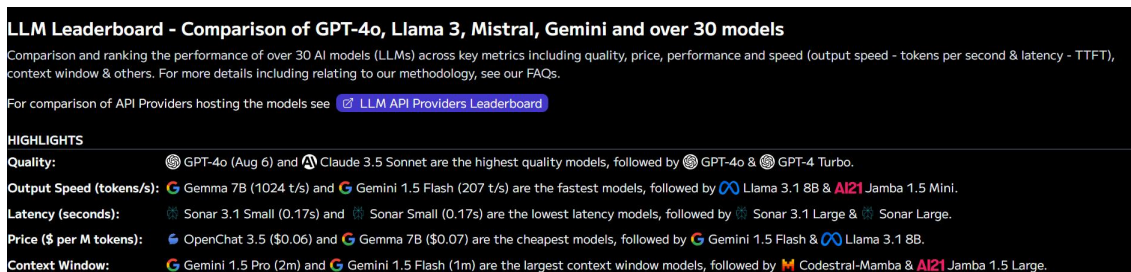


Figura 4: Comparação geral dos principais modelos de IA generativa. Fonte: <https://artificialanalysis.ai/leaderboards/models>

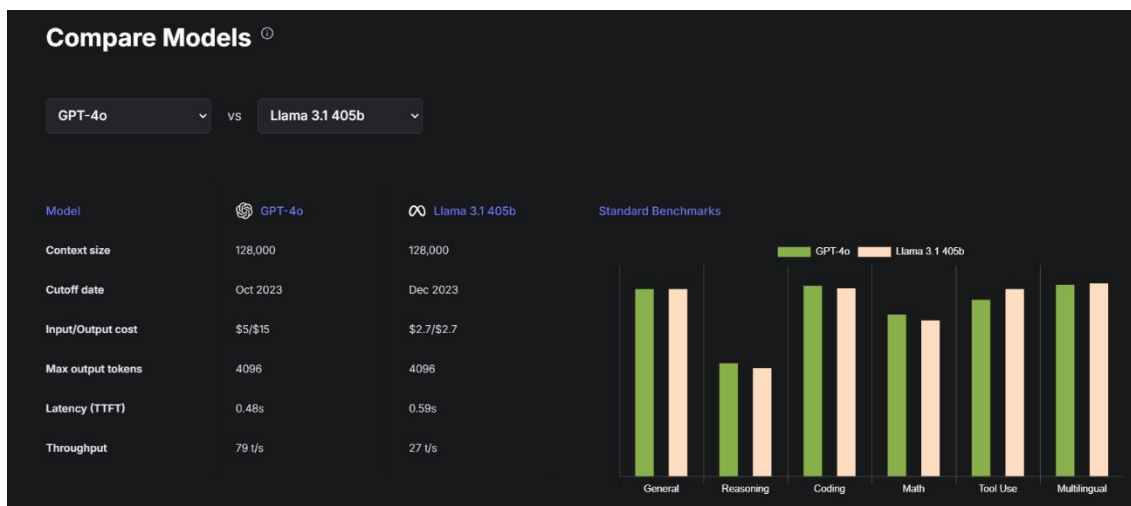


Figura 5: Comparação GPT-4o e LaMa 3.1 405b. Fonte: <https://www.vellum.ai/llm-leaderboard>



ESTADO DO CEARÁ
PODER JUDICIÁRIO
TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

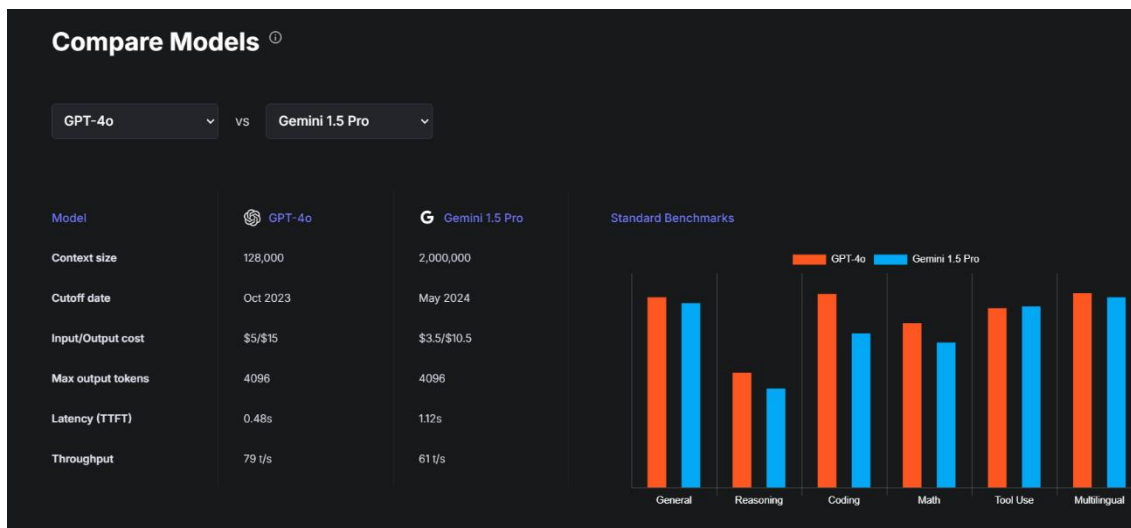


Figura 6: Comparação GPT-4o e Gemini 1.5 Pro. Fonte: <https://www.vellum.ai/llm-leaderboard>

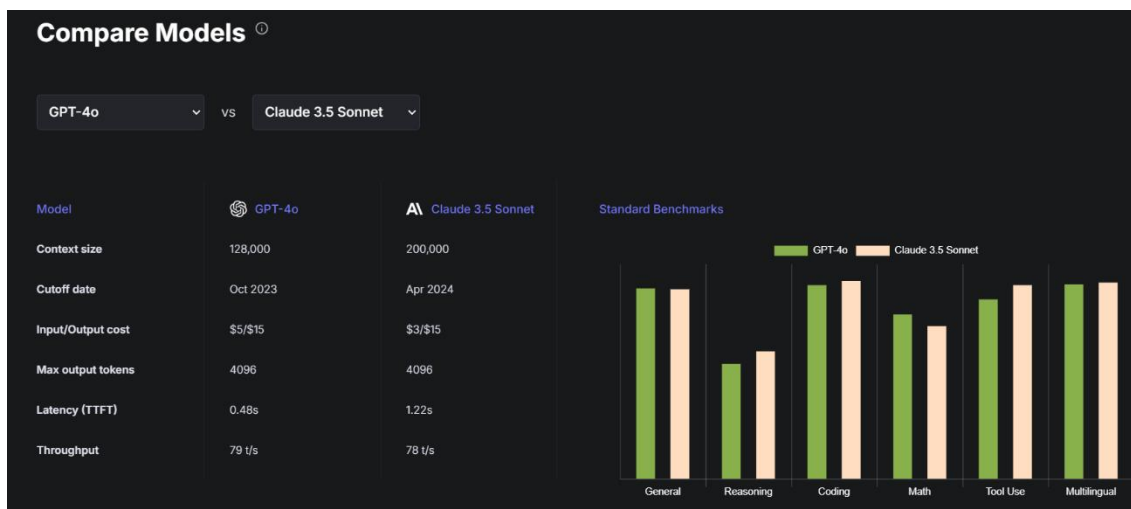


Figura 7: Comparação GPT-4o e Claude 3.5 Sonnet Pro. Fonte: <https://www.vellum.ai/llm-leaderboard>

11. JUSTIFICATIVA DO PARCELAMENTO OU NÃO



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

11.1. Avaliando a possibilidade e a pertinência do parcelamento do objeto para atendimento da necessidade, considerou-se o tipo de objeto pretendido, assim como os aspectos técnicos, operacionais e econômicos, de modo que resultou na identificação de melhor opção em licitar lote único, pois importa em:

11.1.1. serviço único e com características especiais, heterogeneidade ou complexidade;

11.1.2. menor preço do objeto;

11.1.3. pertinência de concentração de responsabilidade técnica pela solução;

11.1.4. dificuldade e oneração excessiva para administrar mais de um contrato;

12. DEMONSTRATIVO DE RESULTADOS PRETENDIDOS

12.1. A solução indicada permitirá o suprimento das necessidades de implementação de tecnologias de IA generativa nas aplicações de IA do TJCE. Isso garantirá, ao menos em relação a este insumo, a não interrupção e a continuidade eficiente das seguintes áreas:

12.1.1. Implantação no TJCE das aplicações desenvolvidas em parcerias externas: A disponibilização destes serviços para o setor de IA do TJCE permitirá a implantação em ambiente de produção do TJCE das aplicações desenvolvidas em ambientes mantidos pelas instituições parceiras. Isto acarretará uma maior eficiência e produtividade nas atividades-fim do tribunal, contribuindo com aplicações de IA que automatizem processos e forneçam sugestões aos usuários.

12.1.2. Desenvolvimento de novas soluções de IA internamente: A disponibilização destes serviços para o setor de IA do TJCE permitirá o



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

desenvolvimento de novas soluções de IA pela equipe interna de servidores do TJCE, aumentando a autonomia do setor.

12.1.3. Maior sigilo dos dados: A contratação direta de um serviço privado fornecerá ao tribunal maior capacidade de gerência e monitoramento dos serviços. Quando acompanhados pelos recursos de segurança e privacidade garantidos pelos provedores privados, garantirá a conformidade com os requisitos de sigilo inerentes da natureza dos dados tratados no poder judiciário.

12.2. A solução escolhida é viável técnica e economicamente, sendo a mais eficiente para o TJCE, uma vez que aproveita as soluções já desenvolvidas no âmbito de parcerias externas, além de integrar-se às soluções da *Microsoft* já adotadas no TJCE.

13. PROVIDÊNCIAS A SEREM ADOTADAS PELA ADMINISTRAÇÃO

13.1. Para a execução e viabilidade da solução, não será necessária a adequação dos ambientes de trabalho do órgão, exigindo-se apenas que os servidores possuam acesso à internet e equipamentos compatíveis (computadores, *tablets* ou *smartphones*).

14. CONTRATAÇÕES CORRELATAS E/OU INTERDEPENDENTES

14.1. No TJCE já houve a contratação do serviço de nuvem *Microsoft Azure* para atender a solução específica do Balcão Virtual (contrato N° 54/2023 do TJCE, Solução de Infraestrutura para o Balcão Virtual do TJCE). A *Microsoft Azure* oferece em seu amplo catálogo, o serviço de IAs generativas, garantindo critérios de qualidade e segurança. Contudo, os créditos na plataforma foram dimensionados especificamente para a aplicação do Balcão Virtual, impossibilitando seu compartilhamento. A análise da contratação havida fez com que a Equipe de Planejamento formasse convicção de que não há inconsistências relevantes e merece implementação do mesmo tipo de solução outrora implementada.

15. DESCRIÇÕES DE POSSÍVEIS IMPACTOS AMBIENTAIS

15.1. Seguindo o Plano de Logística Sustentável do Poder Judiciário do Estado do Ceará – PLS-TJCE 2021-2026 – que é um normativo de planejamento que



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

permite a institucionalização de práticas de sustentabilidade, visando, dentre outros objetivos, a racionalização de gastos e de consumo por meio da construção e análise de indicadores e metas.

- 15.2. A empresa deverá possuir a licenças ambientais condizentes com a sua atividade produtiva e estar em dia com as respectivas licenças;
- 15.3. Os produtos devem observar os critérios de sustentabilidade ambiental decorrentes de sua fabricação, nos termos da legislação de regência e suas eventuais alterações;
- 15.4. As empresas poderão comprovar (por outros meios de prova válidos e regulares admitidos pelo direito) que seus produtos atendem aos requisitos de sustentabilidade ambiental (Acórdão no. 508/2013 – TCU Plenário; Acórdão no. 2.403/2012 – TCU – Plenário e Acórdão no. 1.929/2013 – TCU – Plenário).
- 15.5. Os resíduos decorrentes dos produtos cotados deverão ter destinação ambiental adequada, como coleta seletiva nas unidades do TJCE.

16. CLASSIFICAÇÃO DOS ESTUDOS TÉCNICOS PRELIMINARES

- 16.1. Não há necessidade de classificar estes Estudos Preliminares como sigilosos, nos termos da Lei nº 12. 527, de 2011 (Lei de Acesso à Informação).

17. LEGISLAÇÃO APLICÁVEL AO OBJETO

- 17.1. O tipo de solução identificada como mais acertada para atendimento da necessidade atrai a disciplina específica das seguintes normas, que merecem atenção na implementação da solução:
 - 17.1.1. Lei nº 14.133, de 1º de abril de 2021 que estabelece normas gerais de licitação e contratação para as Administrações Públicas diretas, autárquicas e fundacionais da União, dos Estados, do Distrito Federal e dos Municípios;
 - 17.1.2. Resolução CNJ Nº 468 de 15/07/2022, que dispõe sobre diretrizes para as contratações de Solução de Tecnologia da Informação e Comunicação



ESTADO DO CEARÁ

PODER JUDICIÁRIO

TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

pelos órgãos submetidos ao controle administrativo e financeiro do Conselho Nacional de Justiça (CNJ);

17.1.3. Resolução N° 332 de 21/08/2020 do CNJ que dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências;

17.1.4. Resolução do Órgão Especial n° 40/2022 do TJCE que dispõe sobre a Política Geral de Privacidade e Proteção de Dados Pessoais (PGPPDP) do Poder Judiciário do Estado do Ceará.

18. POSICIONAMENTO CONCLUSIVO

18.1. Com base nas informações levantadas ao longo deste Estudo Técnico Preliminar, foi identificada solução viável de prosseguir e ser concretizada para atendimento da necessidade, na medida em que:

18.1.1. A necessidade apontada é clara e adequadamente justificada;

18.1.2. O atendimento está alinhado com os objetivos estratégicos do órgão e com os programas/atividades inerentes ao TJCE;

18.1.3. As quantidades estão coerentes com os requisitos quantitativos e qualitativos que precisam ser atendidos para resolução da necessidade identificada;

18.1.4. A análise de opções demonstra haver forma de atender ao demandado.

18.2. Os resultados pretendidos com a solução escolhida atendem aos requisitos apresentados e agregam ganhos de eficiência administrativa;

18.3. Foram realizadas estimativas expeditas de preços de mercado, a fim de que se permita avaliar, aprovar e programar o provimento dos recursos necessários ao longo de todo o período de implantação da solução e os valores estimados mostram-se razoáveis e coerentes ao que a solução abrange;

18.4. Diante do exposto, indica-se como viável e recomendado promover-se a contratação de empresa certificada pela *Microsoft* como *government partner*



ESTADO DO CEARÁ
PODER JUDICIÁRIO
TRIBUNAL DE JUSTIÇA DO ESTADO DO CEARÁ

(GP) para fornecimento de serviços de IA generativa através da plataforma *Microsoft Azure*.

Fortaleza, 04 de junho de 2025

Equipe de Planejamento:

Documento assinado digitalmente
gov.br CAIO VIKTOR DA SILVA AVILA
Data: 04/06/2025 11:17:29-0300
Verifique em <https://validar.iti.gov.br>

Caio Viktor da Silva Avila - 52463

Integrante Técnico

JOSEANE PEREIRA
RODRIGUES:011456
16569

Assinado de forma digital por
JOSEANE PEREIRA
RODRIGUES:01145616569
Dados: 2025.06.06 12:47:31 -03'00'

Joseane Pereira Rodrigues - 9347

Integrante Requisitante

FRANCISCO JOSE
PESSOA
FURTADO:32211
473334

Assinado de forma digital
por FRANCISCO JOSE
PESSOA
FURTADO:32211473334
Dados: 2025.06.06
17:15:23 -03'00'

Francisco José Pessoa Furtado - 8284

Integrante Administrativo