

PROCESSO Nº 0000644-64.2024.4.02.8000

ESTUDO TÉCNICO PRELIMINAR

Contratação de serviço profissional de corretagem de serviços em nuvem e de disponibilização e operação da plataforma *multinuvem*, visando gerenciar o uso, o desempenho e a entrega, assim como os relacionamentos entre provedores e consumidores destes serviços, em especial, o serviço de inteligência artificial generativa e suas APIs (Programação de Aplicação ou *Application Programming Interface*) destinado a realizar triagem e organização automatizada de acervos por temas de processos judiciais.

Rio de Janeiro, 10 de outubro de 2024

Histórico de Revisões

Data	Versão	Descrição	Autor
10/10/2024	1.0	Finalização da primeira versão do documento	Integrante requisitante
07/11/2024	3.0	Ajuste de texto e pesquisa preliminar de preços	Integrante técnico

INTRODUÇÃO

O Estudo Técnico Preliminar tem por objetivo identificar e analisar os cenários para o atendimento da demanda que consta no Documento de Oficialização da Demanda, bem como demonstrar a viabilidade técnica e econômica das soluções identificadas, fornecendo as informações necessárias para subsidiar o respectivo processo de contratação.

1. DEFINIÇÃO E ESPECIFICAÇÃO DAS NECESSIDADES E REQUISITOS**1.1 Identificação das necessidades de negócio**

A	Contratação ensejada pelos Ofícios TRF2-OFI-2024/04377 e TRF2-OFI-2024/06068 da Corregedoria do TRF2.
B	Melhorar a eficiência e a qualidade do atendimento aos jurisdicionados.
C	Aprimorar a fiscalização e dar orientação à atividade jurisdicional da Corregedoria.
D	Apoiar a priorização de processos de maior relevância, como os relacionados à Meta 2 do CNJ e outros temas sensíveis.
E	Melhorar a regularização de unidades judiciais a partir da identificação de unidades com processos parados ou com conclusão vencida.
F	Aumentar a eficiência e celeridade na gestão dos acervos com grande volume de processos.
G	Reduzir a carga de trabalho manual mediante a automatização da triagem das peças processuais contribuindo para a organização e eficiência das unidades judiciais.
H	Melhorar a gestão documental dos juízos deste Egrégio Tribunal.

1.2 Identificação das necessidades tecnológicas

A	Implementar melhoria do sistema ApolA que visa dar suporte a magistrados e servidores na gestão juízos na prestação jurisdicional, bem como a integração com sistemas internos do TRF2.
B	Acompanhar a rápida evolução dos modelos de IA nas mais diversas áreas do conhecimento humano, em especial na judicial.
C	Acompanhar a evolução dos recursos de <i>hardware</i> necessários para a execução dos modelos de IA.
D	Otimizar os custos associados a utilização de tecnologias de última geração relacionadas a inteligência artificial.
E	Integrar sistemas internos visando a otimização de processos de trabalho.
F	Oferecer suporte na automatização de tarefas repetitivas e na análise de documentos processuais.
G	Possibilitar o consumo das APIs (Programação de Aplicação ou Application Programming Interface) do serviço de inteligência artificial generativa.

1.3 Demais requisitos necessários e suficientes à escolha da solução de TIC

A	Requisitos legais: Lei 14.133/2021 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.
---	---

	<p>Resolução Nº 332 de 21/08/2020 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. Portaria Nº 271 de 04/12/2020 que regulamenta o uso de Inteligência Artificial no âmbito do Poder Judiciário.</p> <p>Resolução Nº 332 de 21/08/2020 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.</p> <p>Portaria Nº 271 de 04/12/2020 que regulamenta o uso de Inteligência Artificial no âmbito do Poder Judiciário.</p> <p>Lei 13.709/2018, alterada pela Lei 13.853/2019 - Lei Geral de Proteção de Dados Pessoais (LGPD);</p> <p>Resolução CNJ 396/2021 que estabelece a Estratégia Nacional de Segurança Cibernética do Poder Judiciário (ENSEC-PJ).</p> <p>Portaria SGD/MGI nº 5.950, de 26 de outubro de 2023 que estabelece modelo de contratação de software e de serviços de computação em nuvem, no âmbito dos órgãos e entidades integrantes do Sistema de Administração dos Recursos de Tecnologia da Informação - SISF do Poder Executivo Federal.</p>
B	<p>Requisitos de manutenção:</p> <p>A CONTRATADA deverá atender a solicitações de atendimento (chamado) para suporte técnico ser realizadas no regime 24x7, isto é, 24 (vinte e quatro) horas por dia, 07 (sete) dias da semana, 365 (trezentos e sessenta e cinco) dias por ano, durante a vigência do contrato a ser celebrado com a CONTRATANTE.</p> <p>A CONTRATADA deverá garantir todas as atualizações regulares e melhorias contínuas no serviço sem ônus adicionais à CONTRATANTE.</p> <p>O serviço de suporte técnico compreende a orientação da equipe técnica do TRF2 quanto à solução de problemas causados por erros de funcionamento e atualizações ou melhorias da ferramenta.</p> <p>Deverá ser garantido o atendimento de número ilimitado de chamados, durante a vigência do contrato.</p> <p>Todo atendimento de suporte será realizado em língua portuguesa.</p> <p>Todos os chamados devem ser registrados pela CONTRATADA para acompanhamento e controle da execução do serviço. O número de registro deverá ser disponibilizado para o usuário solicitante.</p> <p>Para cada chamado deverá ser atribuído um grau de prioridade, conforme tabela constante do item 16.7 do Termo de Referência desta contratação.</p>
C	<p>Requisitos temporais:</p> <p>A prestação do serviço especificado no ETP e Termo de Referência deve ser iniciado em até 05 (cinco) dias após a assinatura do Contrato.</p> <p>O contrato deverá ter vigência de 12 meses sendo possível sua prorrogação por igual período.</p> <p>A CONTRATADA deverá atender a solicitações de atendimento (chamado) para suporte técnico ser realizadas no regime 24x7, isto é, 24 (vinte e quatro) horas por dia, 07 (sete) dias da semana, 365</p>

	(trezentos e sessenta e cinco) dias por ano, durante a vigência do contrato a ser celebrado com a CONTRATANTE.
D	Requisitos de capacitação: Deverão ser fornecidos tutoriais, passo a passo e demais documentações com instruções simples e claras necessárias para o pleno entendimento dos usuários e da equipe técnica das etapas a serem seguidas para a utilização do serviço.
E	Requisitos de segurança: Observação rigorosa de todas as normas e procedimentos de segurança adotados no ambiente do CONTRATANTE. São vedadas a divulgação, a reprodução ou a utilização de quaisquer informações, a qualquer título, exceto quando previamente autorizadas. São vedadas a cópia, reprodução, divulgação ou a utilização de quaisquer conteúdos de manuais, documentações ou processos administrativos e judiciais, a qualquer título, exceto quando previamente autorizadas. Estar em conformidade: o serviço deve estar em conformidade com a Lei Geral de Proteção de Dados (LGPD), Marco Civil e outras normas aplicáveis.
F	Requisitos ambientais: Não há impactos ambientais diretos a serem mitigados, tendo em vista que os trâmites da contratação serão realizados de forma eletrônica, sem confecção e transporte de mídias ou papel e, além disso, o acesso aos serviços a serem prestados será feito de forma "on-line".

2 – ESTIMATIVA DA DEMANDA – QUANTIDADE DE BENS E SERVIÇOS

Descrição	Quantidade total a ser adquirida
Contratação de serviço profissional de corretagem de serviços em nuvem e de disponibilização e operação da plataforma <i>multinuvem</i> , visando gerenciar o uso, o desempenho e a entrega, assim como os relacionamentos entre provedores e consumidores destes serviços, em especial, o serviço de inteligência artificial generativa e suas APIs (Programação de Aplicação ou <i>Application Programming Interface</i>) destinado a realizar triagem e organização automatizada de acervos por temas de processos judiciais.	2.802,59
O quantitativo pretendido, estimado em créditos para uso do serviço conforme sistema de corretagem da CONTRATADA, visa atender à demanda oficializada pelo ofício TRF2-OFI-2024/04377 (SIGA-DOC) até o final do exercício de 2024.	

3 – ANÁLISE DE SOLUÇÕES POSSÍVEIS

3.1 – IDENTIFICAÇÃO DAS SOLUÇÕES

Ao realizar a análise do mercado de TI foram encontradas as seguintes alternativas mais relevantes:

Id	Descrição da solução (ou cenário)
01	Serviço profissional de corretagem de serviços em nuvem e de disponibilização e operação da plataforma <i>multinuvem</i> do SERPRO
02	<i>Claude</i> da <i>Anthropic</i>
03	<i>GEMINI</i> da <i>GOOGLE</i>

3.2 – ANÁLISE COMPARATIVA DAS SOLUÇÕES

A adoção de soluções de inteligência artificial (IA) tem se mostrado cada vez mais estratégica para empresas de todos os portes e setores. No caso de órgãos públicos brasileiros, a implementação racional de modelos de linguagem avançados como GPT, *Claude* e *Gemini* pode trazer inúmeros benefícios para a instituição como a otimização de processos internos ou até a melhoria da interação com seu público-alvo, no caso da Justiça Federal, o jurisdicionado. Este comparativo visa corroborar a opção pela contratação por dispensa da solução ofertada pelo SERPRO apresentando os principais diferenciais e características de cada solução e comparando seus pontos em comum.

Serviço profissional de corretagem de serviços em nuvem e de disponibilização e operação da plataforma *multinuvem* com foco no serviço de inteligência artificial generativa e suas APIs - SERPRO

Pontos fortes:

Modelo oferece a possibilidade de uso de várias plataformas aproveitando o que cada uma pode oferecer de melhor. Atualmente o SEPRO fornece 7 (sete) alternativas de provedores de IA: *Google*, *Azure* (*GPT*), *AWS*, *Oracle*, *IBM*, *Huawei* e Nuvem de Governo.

Capacidade de gerar textos criativos e coerentes em diversos formatos, como artigos, códigos e scripts.

Integração com diversas plataformas e ferramentas.

Possibilidade de contar com soluções de apoio à solução principal como serviço gerencial consultivo, gerenciamento de nuvem e interlocução com as empresas provedoras de serviço.

Mais opções em IA treinadas em língua portuguesa.

A contratação de serviço tipo *Cloud Service Brokerage* (CSB), Serviço de Corretagem para Disponibilização e Operação de Plataforma *Multinuvem*, viabiliza o acesso ao serviço de inteligência artificial generativa por órgão da administração pública a múltiplos provedores concorrentes entre si.

Faturamento e pagamento em Real.

Pontos fracos:

Não possui autonomia sobre a solução de uma forma geral uma vez que não é seu proprietário.

Quaisquer limitações apresentadas pelos provedores prestadores do serviço consumido pela CONTRATANTE.

Modelos fechados, pois, quaisquer mudanças são dependentes de negociação e disponibilidade dos provedores donos das soluções.

Claude (Anthropic)

Pontos fortes:

Foco em segurança e alinhamento com valores humanos buscando evitar a geração de conteúdo prejudicial ou discriminatório.

Capacidade de manter conversas longas e coerentes com um tom mais natural e humano.

Modelo mais recente com avanços em áreas como raciocínio e resolução de problemas.

Pontos fracos:

Menor disponibilidade e documentação em comparação a ferramentas em mais evidência como o *GPT*.

Vendor lock-in: não há possibilidade de selecionar outro provedor de IA.

Menor comunidade de desenvolvedores e menor número de integrações.

Custo potencialmente mais elevado.

Suporte em língua estrangeira.

Embora disponível para uso no Brasil (<https://www.anthropic.com/news/claude-brazil>) não é clara a sua representação comercial neste país.

Pagamento em dólar americano.

Gemini (Google)

Pontos fortes:

Desenvolvido pela *Google* com acesso a vastos recursos computacionais e dados.

Capacidade de processar e gerar diferentes tipos de mídia, como texto, código, áudio e vídeo.

Integração profunda com outros produtos *Google* como o *Google Search*.

Pontos fracos:

Modelo ainda em desenvolvimento com menos informações públicas disponíveis.

Vendor lock-in: não há possibilidade de selecionar outro provedor de IA. Foco inicial em inglês com menor suporte para a língua portuguesa.

Dependência da infraestrutura da *Google* o que pode limitar a flexibilidade.

Valores no site apresentados em dólar americano.

Quadro comparativo das características mais relevantes de cada solução

Característica	<i>Multinuvem SERPRO</i>	<i>Claude</i>	<i>Gemini</i>
Popularidade	Média	Média	Média
Segurança	Alta	Alta	Alta
Capacidades	Gerar textos, traduzir, escrever diferentes tipos de conteúdo criativo	Manter conversas longas, raciocínio, resolução de problemas	Processar e gerar diversos tipos de mídia, integração com <i>Google</i>
Idiomas	Múltiplos, incluindo Português	Múltiplos, com foco em inglês	Múltiplos, com foco em inglês
Personalização	Média	Média	Média
Custo	Médio	Médio	Alto
Uso de dados para melhoria da própria ferramenta	Sim	Sim	Sim
Respeita propriedade intelectual	Depende do provedor	Sim	Sim

É importante ressaltar que a segurança em modelos de linguagem é um campo em constante evolução e as informações podem variar com o tempo. No entanto, analisando os variados cenários propostos por cada IA, *Claude* e *Gemini* são frequentemente associados a uma maior segurança em comparação ao GPT porque:

1. Foco em alinhamento:

Mitigação de vieses: Tanto *Claude* quanto *Gemini* são treinados com um foco concentrado em alinhamento, o que significa que são projetados para evitar gerar respostas tendenciosas, prejudiciais ou que violem princípios éticos.

Prevenção de informações falsas: Esses modelos são mais cautelosos ao gerar informações que possam ser falsas ou enganosas, buscando sempre fontes confiáveis.

2. Arquiteturas e treinamentos mais robustos:

Melhorias contínuas: As empresas por trás do *Claude* e do *Gemini* investem mais em pesquisas e desenvolvimentos contínuos para aprimorar a segurança de seus modelos. Isso inclui o uso de técnicas mais avançadas de aprendizado de máquina e a implementação de salvaguardas adicionais.

Menos propensos a "alucinações", que são respostas falsas ou inventadas pelos modelos de linguagem. *Claude* e *Gemini*, por suas arquiteturas e treinamentos, tendem a ser menos propensos a esse tipo de comportamento, pelo menos em comparação com a versão livre o GPT.

3. Controles e filtros mais rigorosos:

Moderação de conteúdo: Esses modelos geralmente possuem sistemas de moderação de conteúdo mais sofisticados, que ajudam a filtrar e remover respostas que possam ser ofensivas, prejudiciais ou inadequadas.

Proteção contra uso indevido: As empresas desenvolvedoras desses produtos implementam medidas para evitar que seus modelos sejam utilizados para fins maliciosos, como a geração de *deepfakes* (geração eletrônica e artificial de conteúdo audiovisual sobre pessoas ou fatos) ou a disseminação de informações falsas.

É importante lembrar que a segurança não é uma característica absoluta, mas sim relativa. Todos os modelos de linguagem possuem limitações e podem gerar respostas imprevisíveis em determinadas situações. No entanto, *Claude* e *Gemini* são frequentemente mencionados como modelos mais seguros devido aos esforços das empresas desenvolvedoras em construir sistemas mais robustos e alinhados com os valores humanos.

Ademais, as estimativas de custos são generalizações e podem variar significativamente dependendo de diversos fatores:

Modelo específico: Cada modelo possui diferentes versões e capacidades, o que impacta diretamente no preço. Por exemplo, uma versão mais avançada de um modelo será mais cara do que uma versão básica.

Uso: O custo pode variar de acordo com a quantidade de *tokens* (unidades de texto) processados, a frequência de uso e o tipo de aplicação.

Integração com outras ferramentas: A integração com outros serviços ou plataformas pode gerar custos adicionais.

Disponibilidade comercial: Nem todos os modelos estão disponíveis comercialmente da mesma forma. Alguns podem ser acessados através de APIs, enquanto outros podem exigir contratos mais complexos.

Contudo, de uma forma geral, ponderando todas as variáveis supracitadas a solução ofertada pelo SERPRO apresenta o melhor custo-benefício pois, inclusive, possibilita que estejam à disposição da CONTRATANTE mais de uma IA generativa, até mesmo as examinadas como alternativas nesse estudo, em seu pacote de serviços. Isso confere à CONTRATANTE uma grande liberdade de escolha e, portanto, abre um leque múltiplo de possibilidades tecnológicas a serem aferidas com o seu ambiente, softwares e demandas.

Não obstante as inteligências artificiais *Claude* e *Gemini* possam se destacar um pouco mais quanto à segurança a solução *multinuvem* do SERPRO reúne características difíceis de serem superadas como: conformidade à língua portuguesa, custo-benefício, opção de escolha entre 7 IA e suas versões, possibilidade de oferecer um pacote de recursos, faturamento em moeda corrente nacional, além de podermos contar com a expertise da maior empresa pública de tecnologia do mundo confere.

A tabela abaixo mostra uma análise comparativa entre as soluções identificadas quanto a alguns requisitos. Ressaltando que a solução 1 tem foco no GPT:

Requisito	Solução 01	Solução 02	Solução 03
Permite acesso fácil?	X	X	X
Compatível a língua portuguesa e a linguagem jurídica?	X		
Tem representação oficial no Brasil?	X		X
Fatura em moeda nacional?	X		X
Oferece várias opções em caso de indisponibilidade?	X		
Possui contrato com órgão público brasileiro?	X		X

O caráter inédito desta contratação, tanto no que tange a tecnologia da solução quanto a contratação propriamente dita, requer uma exegese precisa sobre o cálculo dos valores cobrados pelos serviços de IA de cada uma das alternativas analisadas neste estudo.

A base de tudo é a unidade de mensuração dos serviços prestados. As IAs podem ser utilizadas como apoio para uma grande variedade de atividades humanas. Já é realidade a utilização de IA no diagnóstico de doenças, na criação de estratégias de marketing, em cálculos matemáticos com ampla aplicabilidade, na definição de estratégias logísticas, na criação de conteúdo audiovisual e nas várias áreas da linguística, a que aqui mais interessa. Especificamente no contexto linguístico, as IAs realizam o chamado Processamento de Linguagem Natural (PLN) que, em uma visão generalizada, é a análise da frequência, similaridade e redundância de palavras e semântica textual e todas suas dimensões.

A IA faz o Processamento de Linguagem Natural (PLN) convertendo texto em dados que pode interpretar, analisando padrões e relacionamentos entre palavras e frases. Usa modelos treinados com grandes quantidades de texto de determinada área de expertise para entender o contexto, o significado e a estrutura das sentenças. Com isso, concretiza tarefas como tradução, resumo, geração de respostas e análise de sentimentos, simulando a compreensão humana da linguagem.

Em última instância, nesse cenário, a IA recebe textos através de sua API (programação de aplicação), os analisa e os devolve classificados e organizados dentro de critérios pré-determinados. Em jargão técnico, a IA recebe uma massa de dados (*input*), divide esses dados em unidades menores (*tokens*), faz seu processamento e, ao término dessa etapa, devolve a conclusão da análise (*output*).

Mas o que é um token? Um *token*, no universo da inteligência artificial, é uma unidade básica de texto usada para processar e calcular informações. Ele representa palavras, subpalavras, ou caracteres, dependendo do tipo de *tokenização* aplicada, e é fundamental no cálculo dos custos de uso em modelos de IA.

Isto posto, praticamente já se tem os elementos necessários para se calcular o valor a pagar pela prestação do serviço de PLN. Contudo, juntamente, a quantidade de tokens processados pela IA também vai ser uma das variáveis desta equação. Então, vamos convencionar, para fins de apuração de valores absolutos, o processamento da quantidade de 72 milhões de tokens de input e a entrega de 20 milhões de tokens de output por mês.

Assim, aplicando-se a fórmula a seguir obtém-se o valor bruto a serem cobrados.

- **Quantidade de *tokens* de entrada x Valor cobrado por *input* de 1 milhão de *tokens* + Quantidade de *tokens* de saída x Valor cobrado por *output* de 1 milhão de *tokens* = Valor total a ser cobrando**

A tabela abaixo mostra a comparação de custos em moeda estrangeira (**dólar americano**) entre as soluções selecionadas e suas respectivas versões identificadas como tecnicamente equivalentes que pudessem tornar exequível a comparação didática. Faz-se ainda necessário informar que o cálculo do valor preliminar do SERPRO foi baseado no GPT 4o Global Deployment 128k, uma vez que, em se tratando de serviço multinuvel, o valor pode variar dependendo da IA escolhida pelo cliente e de sua respectiva versão.

	Valor mensal do input	Valor mensal do output	Valor mensal total	Desconto	Valor final mensal
GPT 4o Global Deployment 128K (SERPRO)	\$360,00	\$300,00	\$660,00	\$165,00	\$495,00
CLAUDE 3.5 Sonnet 200k	\$216,00	\$300,00	\$516,00	\$0,00	\$516,00
GEMINI 1.5 Pro >128K	\$180,00	\$200,00	\$380,00	\$0,00	\$380,00

*O desconto aplicado ao valor cobrado pelo SERPRO é derivado dos acordos firmados diretamente com os provedores.

**Os valores de input e output variam de acordo com a versão do produto, com o seu limite máximo e com a região onde está localizado o cliente que consome o serviço.

O quadro a seguir compara os valores totais em moeda estrangeira (**dólar americano**) e a média apurada.

Id	Descrição	Preço 1	Preço 2	Preço 3	Média	Observação
01	Corretagem de serviços em nuvem e de disponibilização e operação da	\$ 495,00	\$ 516,00	\$ 380,00	\$ 464,00	Preço 1: https://www.serpro.gov.br/ Preço 2: https://Claude.ai/ Preço3: https://Gemini.Google.com/

plataforma <i>multinuvem</i> , em especial, o serviço de inteligência artificial generativa e suas APIs					
<p>Outrossim, em se tratando de contratações por órgãos da União, é sabido ser obrigatório que o faturamento e o pagamento sejam feitos em moeda nacional. Todavia as empresas Anthropic e Google não apresentam valores em Reais com impostos ou encargos inclusos em suas páginas na web. Mesmo assim é possível assinar os serviços de suas IAs no Brasil, mas geralmente utilizando cartão de crédito internacional e, conseqüentemente, arcando com todos os ônus financeiros e administrativos que essa forma de pagamento acarreta (impostos, câmbio flutuante etc). A clareza quanto a cobrança e faturamento são mais algumas das razões que definem a escolha da equipe técnica pelo produto oferecido pelo SERPRO.</p> <p>Mas como o SERPRO tornou possível a concretização de compra de serviços em moeda estrangeira por órgãos públicos? A resposta é a unidade batizada de <i>Cloud Service Brokerage</i> (CSB) criada como um serviço profissional de corretagem de serviços em nuvem e de disponibilização e operação de sua plataforma multinuvem, visando gerenciar o uso, o desempenho e a entrega, assim como os relacionamentos entre os provedores e consumidores desses serviços em nuvem.</p> <p>O Cloud Services Brokerage segue a seguinte fórmula:</p> <ul style="list-style-type: none">• Valor a ser faturado por projeto (R\$) = Volume de CSB x Preço. <p>Onde:</p> <ul style="list-style-type: none">• Volume de CSB = Volume de CSB = $(vcc * fc * (1-fa)) + (vmp * fc * (1+fm))$ <p>Sendo:</p> <ul style="list-style-type: none">• vcc (Volume de Consumo no Console): É o valor total apurado no console de cálculo do provedor, geralmente disponível via <i>web</i>, naquele mês de referência.• Fator de Câmbio (fc): corresponde à importância numérica utilizada para equalização do volume de consumo, de acordo com a forma de apresentação no console de cálculo.• fa (Fator de Ajuste): corresponde ao percentual excedente àquele aplicado diretamente pelo provedor no console a partir dos recursos computacionais instanciados nos provedores parceiros.• vmp (Volume de consumo no Marketplace): que corresponde ao volume de consumo de produtos de <i>software</i> no <i>marketplace</i> do provedor. É dado em dólares se o volume do provedor for em dólares, e em reais se o volume do provedor for em reais. O Volume de Consumo no Marketplace se refere exclusivamente à conta de produtos de softwares obtidos pelo CLIENTE por meio do Marketplace dos provedores.• fm (Fator de Marketplace): trata-se de um percentual de 0% que corresponde ao montante adicional para realização de transação financeira pelo uso das lojas dos provedores. <p>Destarte, o valor em reais a ser cobrado pelo SERPRO levando em conta a apuração supracitada com valor do dólar americano cotado em 08/11/2024 está descrito na tabela a seguir e em anexo.</p>					

SERPRO						desconto que os provedores dão ao serpro			
TRF2									
Service category	Service type	Custom Region name	Description	Part number(s)	Estimated monthly	Com Fator de Ajuste	Valor do CSB Mensal	Valor do CSB Anual	
IA + Machine Learning	Azure OpenAI Service	Brazil South	Modelos de Linguagem, Standard (On-Demand), GPT-4o-Global-Deployment-128K, 72.000 x 1.000 input tokens, 20.000 x 1.000 output tokens		\$660,00	\$495,00	R\$ 4.960,59	R\$ 59.527,03	
			Licensing Program	Enterprise Agreement (EA)					
			Enrollment Number	4732259					
Total					\$660,00	\$495,00	\$4.960,59	\$59.527,03	

***Nesse cálculo o valor do CSB Mensal foi de 1,77.**

O quadro abaixo apresenta a utilização e a aderência das soluções quanto a determinadas políticas, modelos e padrões de governo existentes:

Requisito	Solução			Não se aplica
	01	Sim	Não	
A Solução encontra-se implantada em outro órgão ou entidade da Administração Pública?	01	X		
	02		X	
	03		X	
A Solução está disponível no Portal do Software Público Brasileiro? (quando se tratar de software)	01		X	
	02		X	
	03		X	
A Solução é composta por software livre ou software público? (quando se tratar de software)	01		X	
	02		X	
	03		X	
A Solução é aderente às políticas, premissas e especificações técnicas definidas pelos Padrões de governo ePing, eMag, ePWG?	01			X
	02			X
	03			X
A Solução é aderente às orientações, premissas e especificações técnicas e funcionais do e-ARQ Brasil? (quando o objetivo da solução abranger documentos arquivísticos).	01			X
	02			X
	03			X
A Solução é aderente à Resolução Nº 332 de 21/08/2020 que dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário?	01	X		
	02		X	
	03		X	

4 – REGISTRO DE SOLUÇÕES CONSIDERADAS INVIÁVEIS

Todas as soluções identificadas apresentam aspectos tecnológicos positivos. Entretanto, ao analisar as condições de aquisição comercial e de observância das leis brasileiras das soluções, percebeu-se maior factibilidade e conveniência na solução 01. Além disso, essa solução é a única que permite:

1. Clareza da forma de faturamento em moeda nacional e sob demanda;
2. Contrato a ser celebrado com o Serviço Federal de Processamento de Dados – SERPRO, empresa pública federal e maior empresa pública de tecnologia da informação do mundo;
3. Possibilidade de escolha entre várias opções de IA generativa disponíveis no marketplace da CONTRATADA não limitando a CONTRATANTE somente a um prestador do serviço tencionado.

Portanto, considerando a economicidade, a eficiência e praticidade das propostas examinadas, a equipe de planejamento considera a solução 01 como sendo a mais exequível e vantajosa dentre as supracitadas.

5 – ANÁLISE COMPARATIVA DE CUSTOS (TCO) DAS SOLUÇÕES TÉCNICA E FUNCIONALMENTE VIÁVEIS

5.1 – CÁLCULOS DOS CUSTOS TOTAIS DE PROPRIEDADE

O cálculo dos custos totais foi realizado considerando todas as soluções técnicas funcionalmente viáveis, usando-se como base de cálculo o preço médio obtido no item 3.2 para cada uma das alternativas multiplicado pela quantidade total informada no item 2. O período de 03 (três) anos foi vislumbrado com base em uma projeção do período inicial de vigência estabelecido do contrato a ser celebrado em decorrência da contratação bem-sucedida.

Valores Estimados e Tendências

Característica	Multinuvem SERPRO	Claude (Anthropic)	Gemini (Google)
Unidade de mensuração	<i>Tokens</i> (equivalente a palavras)	<i>Tokens</i>	<i>Tokens</i> ou consultas
Modelo de preços	Por demanda	Por demanda (<i>Pay-as-you-go</i>), planos personalizados	Incluído na <i>Google Cloud Platform</i> , pagamento por uso
Fatores que influenciam o preço	Número de <i>tokens</i> , modelo utilizado, complexidade da tarefa, personalização	Número de <i>tokens</i> , complexidade da tarefa, personalização	Quantidade de dados processados, tipo de instância, uso de GPU
Moeda	Aquisição de créditos de uso em Real	Dólar Americano	Dólar Americano (integração com a <i>Google Cloud Platform</i>)
Transparência de preços	Documentação detalhada sobre os preços e sistema de aquisição de créditos	Documentação disponível, mas com menos detalhes que o <i>GPT</i>	Preços integrados à <i>Google Cloud Platform</i> , com diversas opções de configuração
Outros Custos	Suporte técnico, treinamento, customização	Suporte técnico, treinamento, personalização	Custos de infraestrutura, armazenamento, customização

5.2 – MAPA COMPARATIVO DOS CÁLCULOS DE PROPRIEDADE (TCO) DAS SOLUÇÕES VIÁVEIS				
Id	Descrição	TCO (1º ano)	TCO (2º ano)	TCO (3º ano)
01	Aquisição 1	R\$ 54.950,43	R\$ 109.900,86	R\$ 164.851,29
02	Aquisição 2	Valores em Reais indisponíveis considerando impostos/encargos	Valores em Reais indisponíveis considerando impostos/encargos	Valores em Reais indisponíveis considerando impostos/encargos
03	Aquisição 3	Valores em Reais indisponíveis considerando impostos/encargos	Valores em Reais indisponíveis considerando impostos/encargos	Valores em Reais indisponíveis considerando impostos/encargos

6 – DESCRIÇÃO DA SOLUÇÃO DE TIC A SER CONTRATADA
<p>A solução a ser contratada tem a finalidade de garantir todos os requisitos e etapas necessárias para, em suma, implementar melhoria do sistema ApolA que visa dar suporte a magistrados e servidores na gestão documental de juízos na prestação jurisdicional, bem como a integração com sistemas internos do TRF2. Portanto, a solução deve contemplar serviços de garantia e suporte durante todo o período de validade do contrato a ser celebrado de forma a solucionar problemas e corrigir erros, evitando a interrupção dos processos de negócio que dependem da solução em si. A execução de todos os serviços deverá ser prestada com pagamento mensal sob demanda (de acordo com a necessidade).</p> <p>Características gerais da solução:</p> <ul style="list-style-type: none">● Possibilitar a implementação de melhoria do sistema ApolA.● Viabilizar o acesso a serviço de inteligência artificial generativa;● Possibilitar o consumo das APIs (Programação de Aplicação ou <i>Application Programming Interface</i>) do serviço de inteligência artificial generativa;● Agilizar a análise de textos jurídicos e respectivos processamentos;● Integrar sistemas internos do TRF2.

7 – ESTIMATIVA DE CUSTO TOTAL DA CONTRATAÇÃO DA SOLUÇÃO ESCOLHIDA
Conforme planilha de pesquisa de preços de mercado

8 – DECLARAÇÃO DE VIABILIDADE DA CONTRATAÇÃO
<p>O presente Estudo Técnico Preliminar está de acordo com as necessidades técnicas e operacionais do Órgão. Também está consoante com o objetivo estratégico "Aperfeiçoar e Assegurar efetividade dos serviços de TI para a Justiça Federal" do Plano Estratégico de TI da Justiça Federal (PETI-JF) 2021-2026, bem como o objetivo "Promover Serviços de Infraestrutura e Soluções Corporativas" que consta no ENTIC-JUD conforme Resolução nº 370/2021 alterada pela Resolução CJF nº 851/2023 do CNJ.</p>

Durante a elaboração do presente estudo foi levada em consideração a melhor forma de atender demanda oficializada pelos ofícios TRF2-OFI-2024/04377 e TRF2-OFI-2024/06068 da Corregedoria do TRF2 e esmerar a eficiência e a qualidade do atendimento ao jurisdicionado.

A alternativa escolhida na fase de Análise de Soluções Possíveis foi a que se mostrou viável e exequível do ponto de vista técnico e que melhor atende às necessidades de negócio, enquanto a Pesquisa de Preços de Mercado preliminar demonstrou que a alternativa possui custos adequados à disponibilidade orçamentária e factível quanto a execução financeira e orçamentária em consonância com as leis brasileiras.

Os quantitativos levantados na estimativa da demanda foram calculados de forma a atender a todos os usuários que necessitarão dos serviços especificados, respeitando os requisitos funcionais existentes e prevenindo a interrupção dos processos de negócio que dependem dessa contratação.

Assim sendo, a equipe de planejamento da contratação entende que o presente estudo está de acordo com as necessidades do órgão, que é justificadamente viável quanto aos requisitos de negócios, administrativos e técnicos a serem alcançados, declarando realizável a aquisição proposta.

9 – DA APROVAÇÃO DO ETP E ASSINATURA

9.1. A Equipe de Planejamento da Contratação foi instituída pela Portaria SEI DG/TRF2 Nº 36, de 26 de setembro de 2024. Conforme o § 2º do Art. 11 da IN SGD/ME nº 01 de 2019, o Estudo Técnico Preliminar deverá ser aprovado e assinado pelos Integrantes Técnicos e Demandantes e pela autoridade máxima da área de TIC.

PAPEL	NOME	MATR.	SETOR
Integrante Requisitante (titular):	MARKENSON PAULO FRANÇA	12074	COIMEL
Integrante Requisitante (suplente):	EDSON SALES DA ROCHA	12068	SEINTA
Integrante Técnico (titular):	RODRIGO MORGADO COELHO	12353	SEINTA
Integrante Técnico (suplente):	RUBEN EDWARD ROSE JUNIOR	12032	SEINTA
Integrante Administrativo (titular):	MARCO AURELIO CORREIA DE BRITO	11021	SCON
Integrante Administrativo (suplente):	ADRIANA BRAGA DA SILVA	12154	SECTAP