



# ChatTCU: Inteligência Artificial como assistente do auditor

*ChatTCU: Artificial Intelligence as an Auditor's Assistant*

## **Eric Hans Messias da Silva**

Mestre em Computação Aplicada em Ciência de Dados e graduado em Engenharia de Computação. Auditor Federal de Controle Externo, atualmente atuando como coordenador de implantação e líder do Núcleo de Inteligência Artificial do Tribunal de Contas da União.

**Lattes:** <http://lattes.cnpq.br/2146545880541854>

**Orcid:** <https://orcid.org/0000-0003-0743-6627>

E-mail: [erichm@tcu.gov.br](mailto:erichm@tcu.gov.br)

## **Erick Muzart Fonseca dos Santos**

Auditor Federal de Controle Externo do Tribunal de Contas da União com experiência em auditoria de Tecnologia da Informação e em pesquisa e inovação em ciência de dados. Atualmente atua na Secretaria de Tecnologia da Informação e Evolução Digital (Setid), integrando a equipe do Núcleo de Inteligência Artificial, responsável pelo desenvolvimento do ChatTCU. Professor de machine learning, deep learning e processamento de linguagem natural em diversas instituições públicas, universitárias e em cursos preparatórios para concursos públicos. Instrutor certificado pela nVidia em deep learning.

**Lattes:** <http://lattes.cnpq.br/0046446504662016>

**Orcid:** <https://orcid.org/0000-0003-4928-9565>

**E-mail:** [erickmf@tcu.gov.br](mailto:erickmf@tcu.gov.br)

## **Monique Louise de Barros Monteiro**

Doutoranda em Ciência da Computação, na área de Inteligência Artificial, mestra e bacharela em Ciência da Computação. Certificações Project Management Professional (PMP), Big Data Science Professional, SOA Governance Specialist, SOA Architect, Scum Master, SAFe Agilist, Oracle Certified Master Java EE 5 Enterprise Architect, Sun Certified Programmer for the Java 2 Platform, MCP .NET Framework 2.0 Development Foundation, MCTS .NET Framework 2.0 Web Applications e IBM Object-Oriented Analysis and Design with UML. Auditora Federal de Controle Externo do Tribunal de Contas da União, atualmente atuando no Núcleo de Inteligência Artificial, responsável pelo desenvolvimento do ChatTCU.

**Lattes:** <https://lattes.cnpq.br/4760312081508645>

**Orcid:** <https://orcid.org/0009-0008-9620-6455>

**E-mail:** [moniquebm@tcu.gov.br](mailto:moniquebm@tcu.gov.br)



### **Sarah Lima Bezerra**

Especialista em Análise de Dados para o Controle e bacharela em Ciência da Computação. Certificações Sun Certified Java Programmer, Sun Certified Java Associate e IBM Object-Oriented Analysis and Design with UML. Auditora Federal de Controle Externo do Tribunal de Contas da União, integrando atualmente a equipe do Núcleo de Inteligência Artificial do TCU com o foco em pesquisa aplicada, no desenvolvimento e na definição dos requisitos do ChatTCU.

**Lattes:** <http://lattes.cnpq.br/5092459333622148>

**E-mail:** sarahlb@tcu.gov.br

### **Selmison Campelo de Miranda**

Graduado em Engenharia de Redes de Comunicação, auditor Federal de Controle Externo do Tribunal de Contas da União, atualmente atuando no Núcleo de Inteligência Artificial, responsável pelo desenvolvimento do ChatTCU.

**Lattes:** <http://lattes.cnpq.br/0113506227426492>

**E-mail:** campelom@tcu.gov.br

## **RESUMO**

O Tribunal de Contas da União (TCU) tem adotado a inteligência artificial (IA) para otimizar suas operações de auditoria e administrativas, o que resultou na criação do ChatTCU, uma ferramenta avançada baseada no modelo de linguagem ChatGPT da OpenAI. Esse esforço inovador levou ao desenvolvimento de um *chatbot* que inicialmente replicava funcionalidades do ChatGPT e, com evoluções subsequentes, passou a incorporar dados de sistemas internos do TCU. A ferramenta permite aos usuários realizarem uma variedade de funções, como acesso a jurisprudência selecionada, serviços administrativos e sumarização de documentos, destacando o compromisso do TCU com a modernização e a eficiência operacional. O comando dessa IA utiliza engenharia de *prompt* como método para aprimorar a interação com grandes modelos de linguagem (*large language models*, ou LLMs), permitindo respostas mais precisas e contextualizadas. A técnica de *Retrieval-Augmented Generation* (RAG) é utilizada para integrar conhecimento especializado da instituição ou diretamente do usuário. A rápida adoção do ChatTCU pelas unidades do TCU reflete seu valor imediato e a eficácia em atender às necessidades operacionais, ao mesmo tempo em que demonstra um alinhamento estratégico com as metas institucionais de promover a inovação e a melhoria contínua nos processos de auditoria e administrativos.

**Palavras-chave:** ChatTCU; Inteligência Artificial; auditoria; Tribunal de Contas da União; engenharia de *prompt*; Retrieval-Augmented Generation; Modelos de Linguagem; inovação tecnológica em órgãos públicos

## **ABSTRACT**

*The Brazilian Federal Court of Accounts (TCU) has adopted artificial intelligence (AI) to optimize its auditing and administrative operations, resulting in the creation of ChatTCU, an advanced tool based on OpenAI's ChatGPT language model. This innovative effort led to the development of a chatbot that initially replicated ChatGPT functionalities and, with subsequent evolutions, began*



to incorporate data from TCU's internal systems. The tool allows users to perform a variety of functions, such as accessing selected jurisprudence, administrative services, and summarizing documents, highlighting TCU's commitment to modernization and operational efficiency. The command of this AI uses prompt engineering as a method to enhance interaction with large language models (LLMs), enabling more accurate and contextualized responses. The Retrieval-Augmented Generation (RAG) technique is used to integrate specialized knowledge from the institution or directly from the user. The rapid adoption of ChatTCU by TCU units reflects its immediate value and effectiveness in meeting operational needs, while demonstrating a strategic alignment with the institutional goals of promoting innovation and continuous improvement in audit and administrative processes.

**Keywords:** ChatTCU; Artificial Intelligence; Auditing; Federal Court of Accounts; prompt engineering; Retrieval-Augmented Generation; Large Language Models; technological innovation in public agencies.

## 1. INTRODUÇÃO

O advento do ChatGPT<sup>1</sup> no final de 2022 marcou um avanço notável no domínio da Inteligência Artificial (IA). Desenvolvido pela OpenAI<sup>2</sup>, este modelo de linguagem tem a capacidade de gerar textos coerentes e contextualmente relevantes a partir de instruções fornecidas pelos usuários, os chamados *prompts* de usuário.

Em resposta aos avanços tecnológicos do ChatGPT, o Tribunal de Contas da União (TCU) adotou a iniciativa de avaliar suas potencialidades. Em fevereiro de 2023, o Comitê de Gestão de Tecnologia da Informação (CGTI) do TCU autorizou a formação de um grupo de trabalho especializado chamado GT ChatGPT e Assistentes de Redação por IA. Composto por especialistas das áreas de Tecnologia da Informação, Comunicação, Segurança da Informação, Área Educacional (ISC) e Controle Externo, o grupo tem como missão explorar usos práticos do ChatGPT, promover o entendimento sobre inteligência artificial, produzir e divulgar conteúdos relacionados, organizar *workshops* temáticos, desenvolver diretrizes para a implementação eficaz dessas tecnologias e abordar os riscos associados ao seu uso.

Um dos resultados desse esforço colaborativo foi o lançamento do ChatTCU, um *chatbot* integrado ao Microsoft Teams<sup>3</sup>. A primeira versão dessa ferramenta foi projetada para operar exclusivamente com dados públicos, devido à natureza do acesso direto ao ChatGPT da OpenAI, o que levantava preocupações quanto à confidencialidade das informações. Essa limitação era necessária porque os dados dos usuários poderiam ser coletados e utilizados pela OpenAI para aprimorar o modelo de IA.

---

1 Disponível em: <https://chat.openai.com>. Acesso em: 15 maio 2024.

2 Disponível em: <https://openai.com>. Acesso em: 15 maio 2024.

3 Disponível em: <https://www.microsoft.com/pt-br/microsoft-teams/>. Acesso em: 15 maio 2024.



Em abril de 2023, o TCU deu um passo adiante com a criação do Núcleo de Inteligência Artificial (NIA) – uma equipe dedicada exclusivamente ao avanço e à pesquisa em IA, com o objetivo de aprimorar o ChatTCU e promover o conhecimento sobre o tema.

A segunda versão do ChatTCU foi lançada em junho de 2023. A solução passou a ser uma aplicação web, acessível não apenas via Teams, mas também por navegadores e dispositivos móveis. A migração para a nuvem da Microsoft (Azure) garantiu a proteção das informações dos usuários.

Em outubro de 2023, houve o lançamento da terceira versão do ChatTCU, um marco significativo que expandiu as funcionalidades da ferramenta por incluir dados de sistemas internos do TCU. Essa versão ofereceu aos usuários a possibilidade de consultarem jurisprudência selecionada, acessar informações sobre serviços administrativos e obter resumos de documentos do e-TCU – tudo isso potencializado pelo modelo GPT-4<sup>4</sup> 32K.

A versão mais recente, lançada no final de abril deste ano, introduziu a funcionalidade de upload de documentos, permitindo aos usuários fazerem consultas baseadas em seus próprios conjuntos de dados. Além disso, a integração de normativos do TCU, incluindo instruções normativas, portarias, resoluções e decisões normativas, trouxe uma adição valiosa à solução. A transição para o modelo GPT-4 Turbo ampliou significativamente a capacidade de processamento de texto, enriquecendo ainda mais a experiência do usuário.

A equipe responsável pelo ChatTCU está constantemente em busca de inovações, com o objetivo de expandir o leque de funcionalidades disponíveis para os servidores e colaboradores do TCU. Esses esforços visam não apenas simplificar, mas também otimizar os processos de trabalho, promovendo maior eficiência e eficácia na execução das tarefas cotidianas.

## 2. ENGENHARIA DE *PROMPT* PARA CONTROLE FLEXÍVEL DA IA

A engenharia de *prompt* refere-se ao processo iterativo de elaboração de instruções ou *prompts*, que guiam os grandes modelos de linguagem como o ChatGPT – também conhecidos como LLMs (do inglês, *large language models*) – a produzirem respostas mais úteis e precisas, alinhadas com a intenção do usuário. Os LLMs são extremamente flexíveis e polivalentes, com infinitas tarefas que podem ser realizadas adequadamente pelo mesmo LLM genérico. Assim, o controle fino do que será produzido pelo LLM pode ser realizado, sem necessidade de conhecimento especializado em IA, diretamente pelo usuário final, que redige os comandos em português, em um processo de refinamento progressivo de seus *prompts*, retroalimentado por resultados intermediários de experimentação em diálogo com o LLM, para descobrir abordagens que geram resultados mais satisfatórios.

Técnicas de engenharia de *prompt* foram utilizadas pelos desenvolvedores para otimizar os “*prompts* de sistema”, que controlam como são realizadas as principais funcionalidades genéricas do ChatTCU, como resumo ou perguntas e respostas sobre documentos. Essas mesmas

---

4 Disponível em: <https://platform.openai.com/docs/models/gpt-4-turbo-and-gpt-4>. Acesso em: 15 maio 2024.



técnicas de engenharia de *prompt* podem igualmente ser utilizadas pelo usuário final, para especificar detalhadamente sua intenção para que seja gerado um resultado específico.

A engenharia de *prompt* sistematiza práticas que foram desenvolvidas no uso recente de LLMs, evidenciando a importância de instruções claras e explícitas quanto ao contexto de conhecimento necessário para realizar a tarefa esperada, detalhada nas características do resultado esperado. Isso exige um entendimento tanto do domínio da aplicação quanto das capacidades e limitações do LLM utilizado, requerendo constante experimentação com alternativas de *prompts* e análise dos resultados intermediários obtidos.

Seguem algumas das técnicas mais comuns de engenharia de *prompt*, que requerem experimentação e refino para aplicações específicas:

- Instruir que seja adotada uma “persona” para responder às perguntas como forma indireta de precisar contexto de conhecimento do LLM. Por exemplo, para realizar tarefas típicas de auditoria, pode ser útil mencionar no *prompt* algo como: “Você é um auditor do TCU, com vasta experiência em fiscalização governamental e amplo conhecimento da legislação brasileira sobre a Administração Pública”. Essa simples menção no *prompt* aumenta as chances de essas áreas de conhecimento serem ativadas nas respostas geradas pelo LLM.
- Formatar o *prompt* explicitando seus elementos mais importantes, como instrução ou objetivo, informações de contexto, dados de entrada e formato do dado de saída esperado para a resposta. Exemplo no caso de uso abaixo.
- Fornecer exemplos de resultados esperados diretamente no *prompt*, que servirão tanto de contexto informacional quanto de demonstração das características desejadas para o resultado. Por exemplo, para uma tarefa de extração de nomes de pessoas referenciadas em um texto, o *prompt* poderia conter o seguinte trecho: “Exemplo 1: ‘Maria e João foram passear. Encontraram Pedro no meio do caminho.’ Resultado: ‘Maria, João e Pedro’”. Múltiplos exemplos em diversas circunstâncias da tarefa podem contribuir para a melhora da qualidade dos resultados, o que requer experimentação.
- Induzir ao raciocínio prévio a geração da resposta desejada, técnica conhecida como “*chain-of-thought prompting*” (Wei *et al.*, 2023), pode gerar melhores resultados quando envolve uma sequência de etapas cognitivas. Exemplo: adicionar ao final do *prompt* o seguinte texto “pense passo a passo como produzir o resultado desejado”. Isso leva o LLM a descrever um possível raciocínio para a geração do resultado, o que geralmente resulta em melhor qualidade.

Existem inúmeras outras técnicas de engenharia de *prompt* que podem ser combinadas para melhorar significativamente os resultados gerados por LLM. Essas técnicas estão disponíveis em várias referências gratuitas na internet<sup>5</sup>.

---

5 Destacamos as seguintes referências, disponíveis em:

<https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-the-openai-api>

<https://cookbook.openai.com/>. Acesso em: 15 maio 2024.

<https://learnprompting.org/docs/intro>. Acesso em: 15 maio 2024.

<https://www.promptingguide.ai/>. Acesso em: 15 maio 2024.



Dada a importância da engenharia de *prompt* para obter de LLMs resultados de maior qualidade, a equipe do NIA ofereceu diversas oportunidades de capacitação para servidores do TCU:

- *Workshops* em que servidores apresentavam resultados relevantes que haviam produzido, compartilhando os *prompts* que permitiam gerá-los, assim como as lições aprendidas durante o processo de refinamento das soluções.
- Três turmas de curso de engenharia de *prompt* de 15 horas, para cerca de uma centena de servidores descobrirem o potencial de LLMs e como otimizar seus *prompts*, na plataforma do ChatTCU, para construir funcionalidades úteis para seus trabalhos cotidianos.

### 3. CASOS DE USO

As mais relevantes aplicações do ChatTCU não são imediatamente identificáveis pela equipe de desenvolvedores do NIA, mas requerem considerável conhecimento do negócio, experiência em atuação profissional de auditoria e experimentação em larga escala para identificar as abordagens que geram os melhores resultados. Assim, adotou-se uma busca descentralizada por aplicações relevantes, oferecendo amplo acesso interno ao ChatTCU para envolver servidores de todas as unidades, fomentando a experimentação e o compartilhamento dos melhores resultados.

São descritos, a seguir, alguns casos de uso desenvolvidos por servidores do TCU.

#### 3.1 Geração de relatórios de incidentes de TI pela área técnica

Quando ocorre um novo incidente de TI, como uma falha ou indisponibilidade de sistemas, é formada uma “Sala de Crise”. Esse processo envolve a criação de um grupo no Microsoft Teams, reunindo diversas áreas e empresas relacionadas, direta ou indiretamente, com o incidente. Dentro desse grupo, discussões são realizadas no chat para compreender o problema, identificar suas causas, buscar a solução mais adequada e recuperar o ambiente afetado. Ao concluir essa etapa, é necessário elaborar um relatório que documente o ocorrido. Contudo, vinha sendo um desafio produzir um relatório que fosse conciso, claro e que destacasse os aspectos mais importantes da análise. A introdução do ChatTCU como ferramenta para essa tarefa trouxe melhorias significativas. Agora, os relatórios produzidos são mais bem estruturados, destacam as trocas de mensagens que fundamentam as conclusões e apresentam mais detalhes com uma redação final mais clara. Isso resultou no aumento da qualidade dos relatórios e na diminuição do tempo necessário para completar tal atividade.

Segue o *prompt* utilizado:

“Missão: Gerar um relatório de sala de crise, conforme a metodologia ITIL

Contexto: Você é um especialista na área de TI e está gerenciando uma sala de crise para tratar um incidente ocorrido no ambiente de TI do TCU



Regras: Ser claro, objetivo e completo

Instruções: Descrever o problema (Citar também o nº do SDM); a causa imediata; as ações executadas para tratar a causa imediata e os resultados obtidos; o estudo e a especificação da causa-raiz; o plano de ação traçado para a causa-raiz, sua execução e os resultados obtidos; possíveis melhorias no processo.

Entrada esperada: Chat do Teams onde ocorreram as discussões e o acompanhamento durante a sala de crise (Texto colado abaixo).

Formato de Saída: Texto com até 10000 palavras.

Chat do Teams: ”

Após inserir o texto anterior no *prompt* ainda é necessário copiar todo o texto das trocas de mensagem no Teams, para que sirva de fonte de informação para análise pelo ChatTCU.

### 3.2 Geração de ata de reunião

Há demanda frequente de múltiplas unidades do TCU pela automação de atas de reunião, sejam internas ou externas. Tais atas podem ser redigidas de forma semiautomática, para posterior validação pelo usuário, ativando-se a função de transcrição do Microsoft Teams, que gera um arquivo com a íntegra do que foi falado na reunião e atribuição a cada orador, que pode então ser copiada no ChatTCU e ajustado um *prompt* para geração de ata que traga o conteúdo esperado pelo usuário, seja de simples síntese ou, mais detalhadamente, de lista de participantes, pontos discutidos, decisões tomadas e ações a serem realizadas, identificando responsáveis e prazos que tenham sido mencionados na reunião.

### 3.3 Sugestão de objetivos e resultados-chave (OKR)

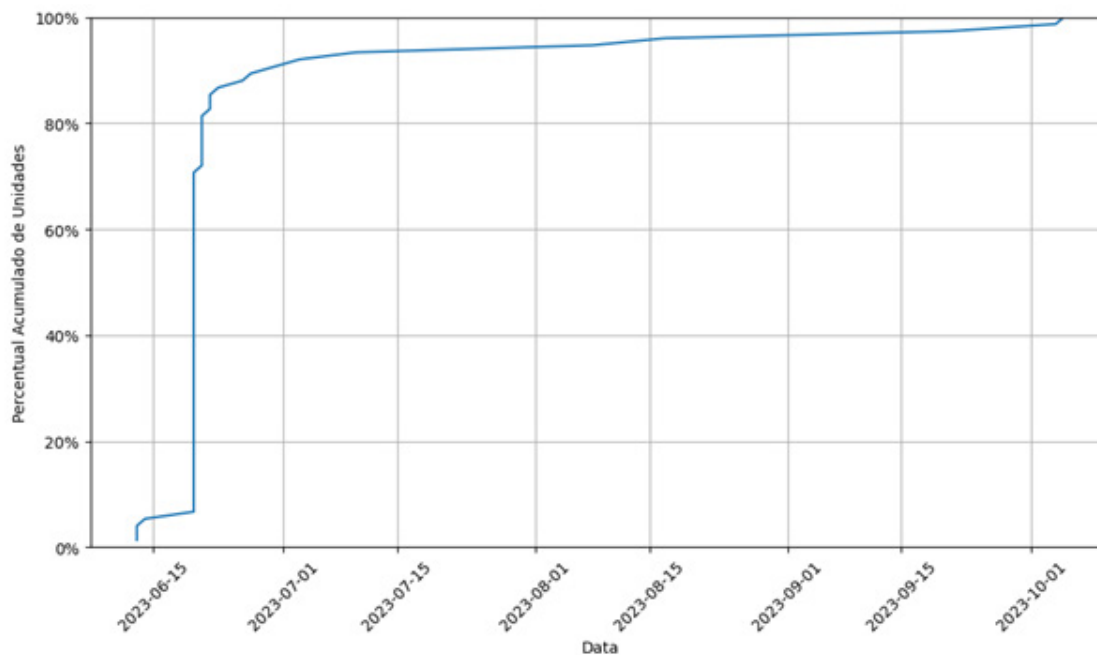
O desafio de explicitar indicadores de desempenho (*key performance indicators*, KPI) e objetivos e resultados-chave (*Objectives and Key Results* – OKR) (Groove, 1983) para estabelecer objetivos e metas para subunidades do TCU pode ser considerável para quem ainda não tem experiência com essa metodologia. Mas, como o GPT-4 deve ter encontrado inúmeros exemplos de OKR durante seu treinamento, houve relatos de aplicações bem-sucedidas do ChatTCU para sugerir possíveis novos OKR. O *prompt* para tal geração pode trazer informações de contexto sobre a unidade, como regimento interno, portarias que descrevam atribuições ou publicações prévias da unidade, assim como exemplos de OKR passados da unidade ou de outras unidades com atribuições similares, solicitando-se que sejam gerados novos OKR plausíveis, maximamente diversos e originais. Cabe então aos responsáveis pela unidade selecionar os mais adequados e ajustá-los com base no próprio contexto.



## 4. ADOÇÃO DO ChatTCU

O gráfico a seguir revela uma tendência marcante e promissora na adesão do ChatTCU por parte das unidades do TCU. Notou-se uma aderência acelerada logo após o lançamento da ferramenta, evidenciada pela ascensão quase vertical inicial. Essa aceitação quase imediata é um testemunho eloquente da relevância e da adequação do ChatTCU às necessidades operacionais das unidades.

**Figura 1** – Uso do ChatTCU por unidades ao longo do tempo – percentual de unidades



Fonte: os autores.

A velocidade dessa adesão inicial pode ser atribuída a uma série de fatores positivos. Primeiramente, reflete a percepção imediata do valor agregado pela ferramenta, com sua promessa de otimizar e tornar mais eficientes as atividades do Tribunal. Além disso, esse rápido engajamento das unidades indica que a implementação do ChatTCU foi realizada com êxito, acompanhada de uma comunicação efetiva e, possivelmente, de uma transição suave para a nova tecnologia.

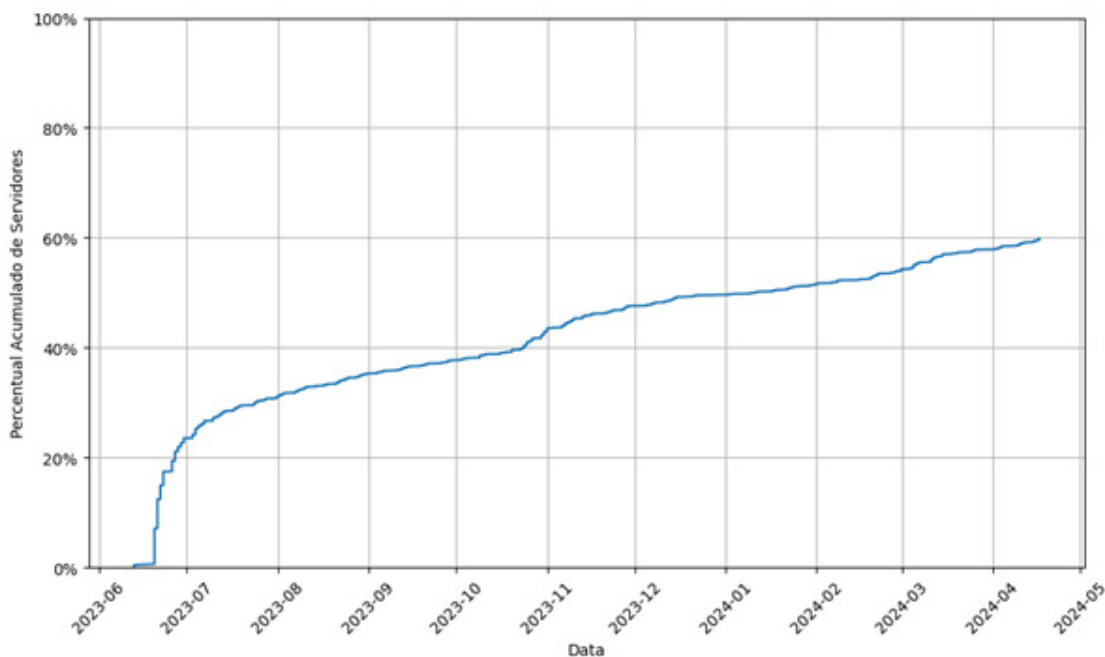
Concomitantemente, a fase de estabilização subsequente ao crescimento acentuado demonstra a capacidade do ChatTCU de alcançar uma cobertura ampla e de se manter como uma ferramenta relevante para as unidades do TCU. Isso não apenas ressalta a eficácia do processo de implementação, mas também sinaliza um alinhamento estratégico com as metas institucionais de inovação e melhoria contínua.

Por fim, é imperativo reconhecer a natureza progressiva dessa adesão como um indicativo da sustentabilidade da ferramenta. A estabilização no patamar elevado de uso reforça a ideia de que o ChatTCU foi integrado com sucesso nas operações cotidianas do tribunal, consolidando-se como um componente essencial no ecossistema de tecnologia da informação do TCU.





**Figura 2** – Uso do ChatTCU por usuários ao longo do tempo – percentual de servidores



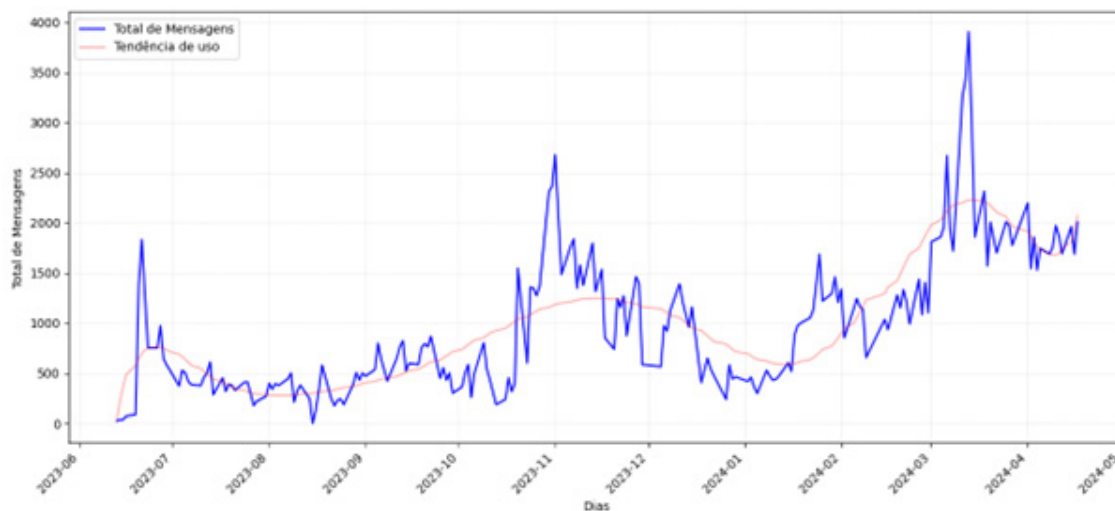
Fonte: os autores.

A análise do segundo gráfico proporciona uma visão esclarecedora do comportamento de adesão dos servidores do TCU ao ChatTCU, revelando uma trajetória de crescimento consistente e sustentável. Diferentemente da curva exponencial observada na adesão por unidades, o perfil de adoção pelos servidores é caracterizado por uma progressão mais suave, o que ressalta um processo de integração consciente e deliberado da ferramenta nas atividades diárias dos servidores.

Esse padrão gradual de crescimento é indicativo de uma estratégia de adoção bem-sucedida que prioriza a compreensão plena da ferramenta e a sua assimilação efetiva no fluxo de trabalho. A ascensão constante do uso sugere que os servidores estão se tornando cada vez mais adeptos e confiantes no uso do ChatTCU, o que implica uma curva de aprendizado positiva e uma valorização crescente da ferramenta como um recurso para aprimoramento das tarefas de auditoria e outros processos internos.

Ademais, a sustentabilidade desse crescimento sugere uma integração bem planejada e executada, na qual os usuários não foram apenas expostos ao ChatTCU, mas também foram adequadamente apoiados por meio de treinamentos e *workshops*. Isso não só demonstra a viabilidade da ferramenta como também fortalece o comprometimento institucional com a inovação e o desenvolvimento profissional contínuo.

Portanto, o gráfico confirma o engajamento positivo dos servidores com o ChatTCU, refletindo a aceitação da ferramenta como um complemento valioso às suas funções. Essa adesão progressiva alinha-se com os objetivos estratégicos do TCU de promover a eficiência, a transparência e a modernização das práticas de controle e auditoria.

**Figura 3** – Mensagens trocadas no ChatTCU por dia

Fonte: os autores.

Ao incorporar os eventos significativos no contexto da análise do gráfico, pode-se compreender melhor o padrão de adoção do ChatTCU pelos servidores do TCU. Os dois primeiros picos expressivos na linha azul correspondem ao lançamento das versões V2 e V3 do ChatTCU.

O primeiro pico marca o lançamento da versão V2 do ChatTCU, que foi a primeira a estar disponível para todo o tribunal, evidenciando um interesse imediato e um pico de curiosidade e experimentação entre os servidores.

O segundo pico saliente na linha do tempo alinha-se com a introdução da versão V3 do ChatTCU, um lançamento significativo por ser a primeira versão a se integrar com aplicações internas do TCU, proporcionando funcionalidades robustas como a geração de resumos de documentos processuais, a interação com quase 300 sistemas administrativos do Sistema CASA, e o acesso facilitado à jurisprudência selecionada. A incorporação do modelo GPT-4 nessa versão representou um salto qualitativo no desempenho da ferramenta, o que pode ter motivado adoção e exploração mais intensas da aplicação.

O terceiro pico reflete o impacto de uma série de apresentações direcionadas às equipes do TCU, com o intuito de promover ainda mais a utilização do ChatTCU. Esse aumento substancial na atividade indica o sucesso dessas iniciativas de engajamento em estimular o uso da ferramenta e em destacar a sua relevância para as atividades do tribunal.

Por outro lado, o vale notado no gráfico entre 17/12 e 16/1 corresponde ao período de recesso no TCU, em que muitos servidores não estão em atividade regular. Essa queda acentuada no uso do ChatTCU nesse intervalo de tempo resalta a correlação direta entre a presença dos servidores e a atividade na plataforma, reiterando que o ChatTCU é uma ferramenta de uso cotidiano que reflete os ciclos operacionais do tribunal.



Essa análise dinâmica do uso do ChatTCU fornece *insights* valiosos sobre a interação dos servidores com a plataforma, ressaltando os aspectos positivos de cada atualização e iniciativa de engajamento, e reconhecendo os padrões naturais de atividade institucional que influenciam a adoção da tecnologia.

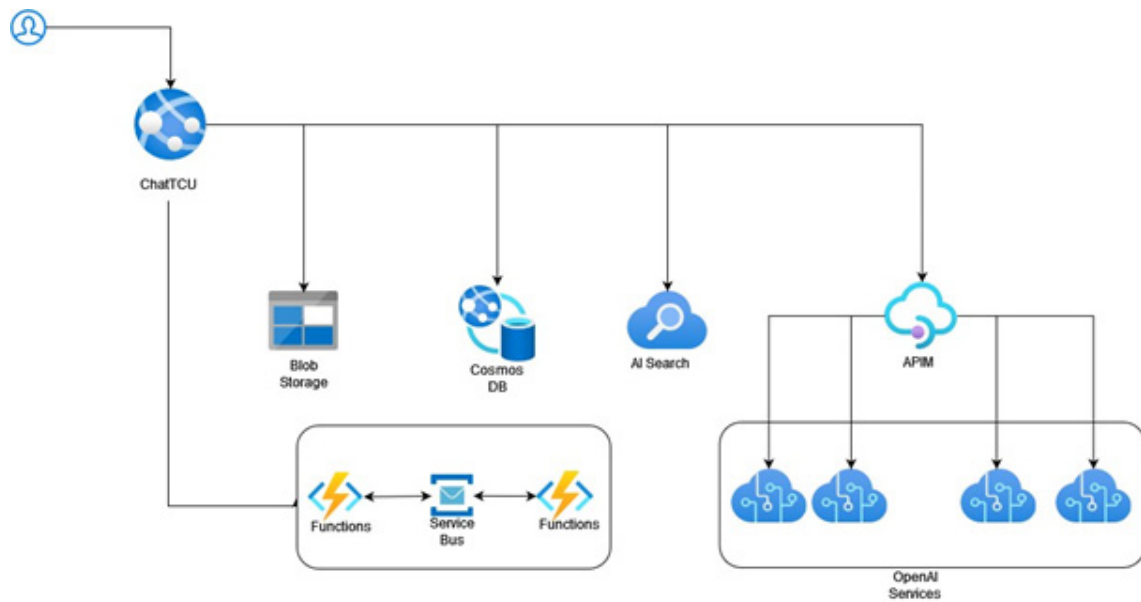
## 5. INFRAESTRUTURA TECNOLÓGICA

Os grandes modelos de linguagem natural (LLMs) costumam exigir uma quantidade significativa de recursos computacionais, sendo bastante natural o uso de uma infraestrutura computacional em nuvem que proporcione maior flexibilidade, escalabilidade e efetividade de custos.

Dadas essas circunstâncias, a infraestrutura tecnológica do ChatTCU está pautada desde o início por uma forte adoção de serviços computacionais disponíveis em provedores de nuvem.

O diagrama a seguir é uma representação simplificada dos principais serviços constantes na Azure atualmente em adoção pelo ChatTCU:

**Figura 4** – Infraestrutura Tecnológica do ChatTCU



Fonte: os autores.



Breve descrição dos componentes:

- 1. Cosmos DB:** armazena os dados estruturados do sistema.
- 2. Blob Storage:** armazena dados não estruturados, como documentos pdf, .docx, .xlsx etc.
- 3. AI Search:** provê pesquisa avançada de dados textuais com uso de tecnologias especializadas em buscas textuais e base de dados vetoriais.
- 4. Service Bus:** serviço de mensageria e comunicação entre serviços, responsável por coordenar o processo de indexação de documentos.
- 5. Functions:** executa a indexação de documentos no AI Search sob a coordenação do *Service Bus*.
- 6. OpenAI Service:** provê acesso aos diversos modelos de Inteligência Artificial desenvolvidos pela OpenAI, tais como os modelos GPT, em suas várias versões, e *Text Embeddings*.
- 7. API Management (APIM):** gerencia, balanceia e controla os acessos aos modelos da OpenAI e de outras *Application Programming Interfaces* (API).

Vale destacar que a descrição acima é uma fotografia do momento. Toda a solução tecnológica do ChatTCU está sujeita a constantes evoluções e melhorias.

## 6. EXTENSÃO DO CONHECIMENTO

Como seria de se esperar de toda tecnologia disruptiva de propósito geral, os modelos de linguagem ou LLMs não possuem nativamente conhecimento de domínios especializados. A solução tradicionalmente adotada por desenvolvedores de sistemas baseados em inteligência artificial consiste em prosseguir com o treinamento de tais sistemas por meio de dados provenientes do domínio em específico. Por outro lado, no contexto de sistemas baseados em IA generativa – em geral baseados em modelos de IA complexos, que envolvem altos custos de treinamento –, o treinamento continuado nem sempre é uma opção viável.

Empresas provedoras de modelos de linguagem, a exemplo da OpenAI, muitas vezes proveem soluções de treinamento continuado (comumente referenciado pelo termo *fine-tuning* ou “ajuste fino”, em tradução livre), nas quais o usuário envia seus dados para a nuvem do provedor e aguarda pelo treinamento de uma instância customizada do modelo de IA. Ainda assim, apesar da facilidade, delegar o treinamento de um modelo customizado para um provedor externo envolve, além do custo financeiro, questões como propriedade e sigilo de dados, o que geralmente inviabiliza esse tipo de solução.



## 6.1 Geração aumentada por recuperação

Como alternativa à inviabilidade de realizar treinamento customizado, seja por razões financeiras ou por aspectos de confidencialidade dos dados, surge a solução denominada *Retrieval-Augmented Generation* (RAG) (Lewis *et al.*, 2020), ou “geração aumentada por recuperação”, em tradução livre. Em linhas gerais, uma solução de RAG baseia-se na utilização de dois componentes na arquitetura do sistema:

- Recuperador de informações: comumente qualquer solução corporativa de busca textual.
- Gerador de respostas: aqui faz-se referência tão somente ao modelo de linguagem subjacente (ex.: GPT-4) (Lewis *et al.*, 2020).

O fluxo de funcionamento da solução RAG consiste nos seguintes passos:

1. A pergunta digitada pelo usuário na interface do *chatbot* é encaminhada ao módulo recuperador de informações, que retorna o conjunto de documentos ou trechos de documentos mais relevantes de acordo com a pergunta/consulta.
2. O conjunto de documentos retornado no passo anterior é repassado para o módulo gerador de respostas (modelo de linguagem), juntamente com a pergunta do usuário. A partir daí, o modelo será capaz de “raciocinar” sobre o conteúdo recebido e de, então, gerar uma resposta para a pergunta do usuário.
3. Caso o modelo não tenha sido capaz de gerar uma resposta a partir dos documentos recuperados, ele deverá responder ao usuário que não tem condições de responder, em vez de fornecer respostas aleatórias ou fictícias (alucinações). Com isso, reduz-se a probabilidade de alucinações, o que contribui, ao mesmo tempo, para a confiabilidade e a adaptabilidade da solução para o domínio desejado.

Nas próximas seções, será descrito como a solução acima foi empregada em alguns dos principais domínios específicos do ChatTCU. Antes disso, é importante ressaltar que o ChatTCU também tem a capacidade intrínseca de identificar o assunto ou domínio específico com o qual a pergunta realizada pelo usuário tem mais afinidade, processo denominado detecção de intenção. A maneira como a detecção de intenção é implementada baseia-se, na versão atual, no uso do próprio modelo de linguagem subjacente para classificar o domínio ao qual a pergunta se refere. Embora esse processo tenha uma margem de erro, ainda assim o usuário pode “corrigir” a decisão do modelo especificando, no texto da pergunta, a qual base de conhecimento ou domínio ele se refere, conforme será descrito mais adiante.

### 6.1.1 Jurisprudência e normativos do TCU

A primeira base incorporada ao ChatTCU foi a chamada **jurisprudência selecionada**, subconjunto da jurisprudência do TCU, classificada de acordo com temas recorrentes.



Os dados do referido subconjunto da jurisprudência do TCU foram indexados no motor de busca Azure AI Search<sup>6</sup>. Excertos e enunciados da jurisprudência foram quebrados em trechos menores, para tornar o resultado das buscas mais preciso do ponto de vista do usuário, e a recuperação em tempo de consulta é feita por meio de um algoritmo padrão de busca textual (BM25) (OpenAI, 2023).

Posteriormente, foi incorporada a base de normas internas do TCU, incluindo portarias, decisões normativas, instruções normativas e resoluções, todas relacionadas ao tema controle externo. Aqui, verificou-se que os melhores resultados para a busca eram obtidos por meio de busca textual com rearranqueamento, tecnologia automaticamente provida pela plataforma Azure AI Search.

A seguir, apresentam-se exemplos de perguntas e respostas relativas a cada uma das bases mencionadas. As **Figuras 5 e 7** mostram respostas a perguntas do usuário, relativas, respectivamente, à jurisprudência e a normas do TCU, e as **Figuras 6 e 8** mostram, respectivamente, janelas modais exibidas quando o usuário clica em alguma das referências citadas na resposta anteriormente formulada pelo sistema.

**Figura 5** – Pergunta relacionada à jurisprudência do TCU

MM O que há na jurisprudência sobre cesta de preços?

Base de informações: **Jurisprudência Selecionada**

A jurisprudência do Tribunal de Contas da União (TCU) aborda a "cesta de preços" principalmente no contexto de aquisição de bens e contratação de serviços pela Administração Pública. A "cesta de preços" é um conjunto de preços aceitáveis que serve como referência para a estimativa de custos em licitações e contratações.

O Acórdão 4958/2022-Primeira Câmara e o Acórdão 1875/2021-Plenário destacam que as pesquisas de preços devem ser baseadas em uma "cesta de preços", dando preferência para preços praticados no âmbito da Administração Pública, oriundos de outros certames. A pesquisa de preços feita exclusivamente junto a fornecedores deve ser utilizada em último caso, na ausência de preços obtidos em contratações públicas anteriores ou cestas de preços referenciais [1](#) [2](#).

O Acórdão 2637/2015-Plenário reforça que as estimativas de preços prévias às licitações devem estar baseadas em uma "cesta de preços aceitáveis", que pode ser oriunda de pesquisas junto a fornecedores, valores adjudicados em licitações de órgãos públicos, entre outras fontes [3](#).

O Acórdão 2399/2022-Segunda Câmara menciona que a jurisprudência evoluiu para a recomendação de adoção de cesta de preços aceitáveis, adotando-se preços públicos como referência sempre que possível [4](#).

O Acórdão 1875/2021-Plenário também recomenda ao Ministério da Economia que as pesquisas de preços para estimativa de valor de objetos a serem licitados sejam baseadas em uma "cesta de preços", dando preferência para preços públicos, oriundos de outros certames [5](#).

O Acórdão 2637/2015-Plenário e o Acórdão 2352/2016-Plenário também mencionam que as estimativas de preços devem considerar uma cesta de preços aceitáveis, que podem ser oriundos de pesquisas diretas com fornecedores ou em seus catálogos, valores adjudicados em licitações de órgãos públicos, sistemas de compras (Comprasnet), valores registrados em atas de SRP, avaliação de contratos recentes ou vigentes, compras e contratações realizadas por corporações privadas em condições idênticas ou semelhantes [6](#) [7](#).

Em resumo, a "cesta de preços" é uma ferramenta importante para a estimativa de custos em licitações e contratações pela Administração Pública, e deve ser baseada em preços aceitáveis oriundos de diversas fontes, dando preferência para preços públicos. A pesquisa de preços junto a fornecedores deve ser utilizada como último recurso, na ausência de outras fontes de preços.

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#)

É necessário revisar as respostas, pois podem conter dados incorretos.

Fonte: ChatTCU.

6 x Azure AI Search: <https://azure.microsoft.com/en-us/products/ai-services/ai-search/>.



**Figura 6** – Detalhamento de referência citada pela resposta à pergunta formulada

**MM** O que há na jurisprudência sobre cesta de preços?

Base de informações: **Jurisprudência Selecionada**

A jurisprudência do Tribunal de Contas da União (TCU) aborda a "cesta de preços" principalmente no contexto da Administração Pública. A "cesta de preços" é um conjunto de preços aceitáveis que serve como referência para contratações.

O Acórdão 4958/2022-Primeira Câmara e o Acórdão 1875/2021-Plenário destacam que as pesquisas de preços devem dar preferência para preços praticados no âmbito da Administração Pública, oriundos de outros certames.

**Acórdão 4958/2022-Primeira Câmara\_1**

As pesquisas de preços para aquisição de bens e contratação de serviços em geral devem ser baseadas em uma "cesta de preços", devendo-se dar preferência para preços praticados no âmbito da Administração Pública, oriundos de outros certames. A pesquisa de preços feita exclusivamente junto a fornecedores deve ser utilizada em último caso, na ausência de preços obtidos em contratações públicas anteriores ou cestas de preços referenciais (Instrução Normativa Seges-ME 73/2020) .

Link: <https://pesquisa.apps.tcu.gov.br/redireciona/jurisprudencia-selecionada/JURISPRUDENCIA-SELECIONADA-139562>

**Fechar**




Fonte: ChatTCU.

**Figura 7** – Pergunta relacionada a normativos do TCU

**MM** Qual é o objetivo da Portaria SEGECEX n° 12/2016?

Base de informações: **Normativos do TCU**

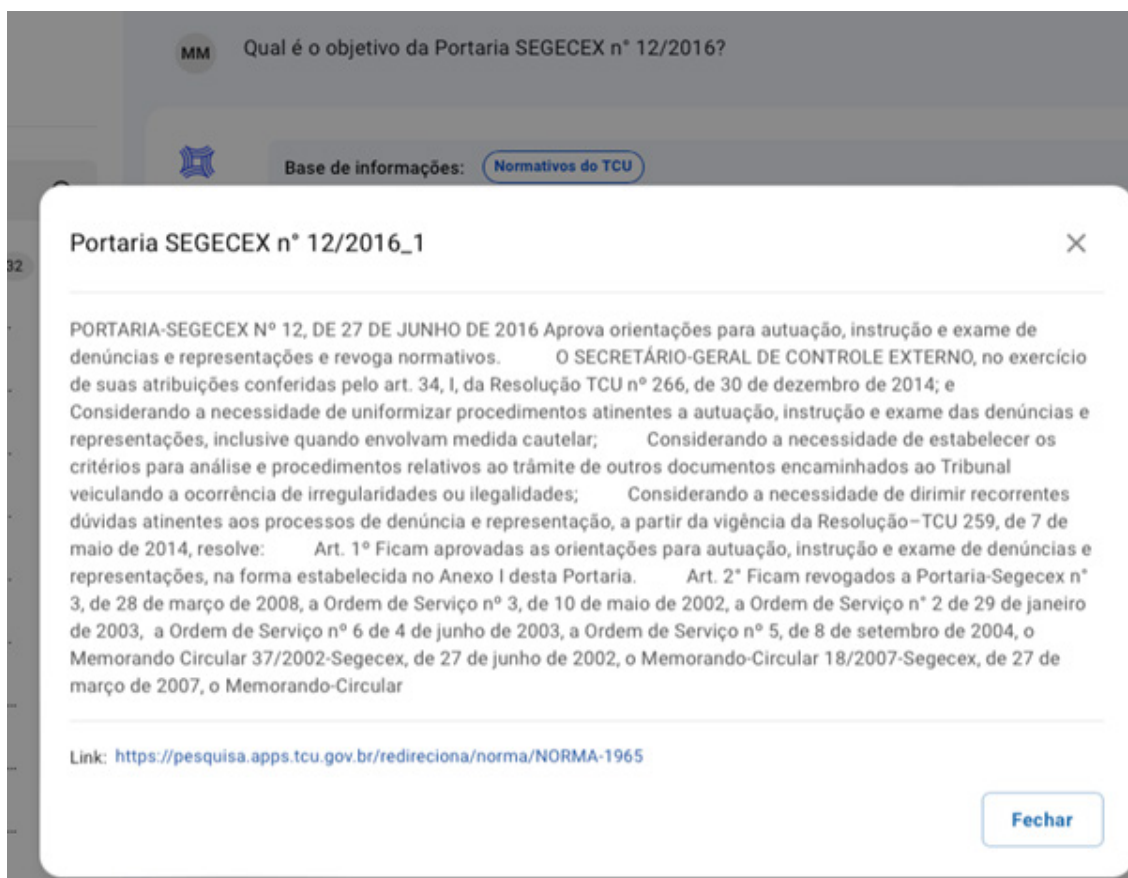
A Portaria SEGECEX n° 12/2016 tem como objetivo aprovar orientações para autuação, instrução e exame de denúncias e representações, além de revogar normativos anteriores. Ela foi estabelecida considerando a necessidade de uniformizar procedimentos relacionados à autuação, instrução e exame das denúncias e representações, inclusive quando envolvam medida cautelar. Além disso, a portaria também considera a necessidade de estabelecer os critérios para análise e procedimentos relativos ao trâmite de outros documentos encaminhados ao Tribunal que veiculem a ocorrência de irregularidades ou ilegalidades. A portaria também busca dirimir recorrentes dúvidas atinentes aos processos de denúncia e representação, a partir da vigência da Resolução – TCU 259, de 7 de maio de 2014. A Portaria SEGECEX n° 12/2016 também revogou a Portaria-Segecex n° 3, de 28 de março de 2008, a Ordem de Serviço n° 3, de 10 de maio de 2002, a Ordem de Serviço n° 2 de 29 de janeiro de 2003, a Ordem de Serviço n° 6 de 4 de junho de 2003, a Ordem de Serviço n° 5, de 8 de setembro de 2004, o Memorando Circular 37/2002-Segecex, de 27 de junho de 2002, o Memorando-Circular 18/2007-Segecex, de 27 de março de 2007, o Memorando-Circular 35/2011-Segecex, de 22 de julho de 2011, Memorando-Circular 6/2012-Segecex, de 8 de março de 2012, o Memorando-Circular 25-Segecex, 12 de junho de 2013, o Memorando-Circular 27/2014 – Segecex, de 23 de julho de 2014 e o Memorando-Circular 38/2015, de 25 de novembro de 2015 <sup>1</sup> <sup>2</sup> .

É necessário revisar as respostas, pois podem conter dados incorretos.

Fonte: ChatTCU.



**Figura 8** – Detalhamento de referência citada pela resposta à pergunta formulada

Fonte: ChatTCU.

### 6.1.2 Sistemas administrativos

Outra base incorporada ao ChatTCU foi a do Portal CASA<sup>7</sup>, sistema que possui informações de quase 400 serviços administrativos oferecidos ao público interno do TCU. Isso possibilitou ao usuário do ChatTCU obter informações que vão desde como pedir um cafezinho para uma reunião até descobrir quais etapas e prazos são necessários para solicitar uma licença específica.

Para essa funcionalidade, os dados dos serviços disponíveis no Portal CASA foram indexados no motor de busca Azure AI Search. Para recuperar os serviços considerados mais relevantes para responder às perguntas dos usuários, a busca vetorial<sup>8</sup> foi a que apresentou os melhores resultados.

7 Disponível em: <https://casa.apps.tcu.gov.br/>. Acesso em: 15 maio 2024.

8 Para vetorização, foi utilizado o modelo text-embedding-ada-002. Mais informações disponíveis em: <https://platform.openai.com/docs/guides/embeddings>. Acesso em: 15 maio 2024.






A **Figura 9** mostra alguns exemplos de perguntas relacionadas aos serviços administrativos. As respostas geralmente mencionam os requisitos, o público-alvo e os passos necessários para o usuário solicitar o serviço, e sempre possuem um link para o Portal CASA, onde o usuário poderá obter mais informações sobre o serviço em questão.

**Figura 9** – Perguntas relacionadas aos serviços administrativos

**SB** Como faço para marcar férias?




Base de informações: [Sistema CASA](#)

 Para marcar férias, você deve seguir os seguintes passos:

1. Acesse o **Sistema GRH**. Para isso, clique no **botão Acessar** disponível no link fornecido.
2. No menu **Frequência**, selecione a opção **(Férias) Solicitação / Gestão** para consultar seu saldo, solicitar ou cancelar períodos de férias.
3. Após a solicitação, sua chefia receberá um e-mail solicitando a homologação desse período de férias, com um link direto para a tela de **Gestão de Férias**.
4. A chefia deve então homologar a solicitação no **GRH**, acessando o menu **Frequência** e selecionando **(Férias) Homologar Férias**.


Este serviço é destinado a servidores ativos. Para mais informações e para acessar o sistema, você pode visitar o [link fornecido](#).




Para mais detalhes sobre o processo de solicitação e gestão de férias, você pode consultar a página de informações disponível [aqui](#).

**SB** Quem tem direito ao auxílio-moradia?

Base de informações: [Sistema CASA](#)

 O auxílio-moradia é um ressarcimento de despesas com aluguel ou hospedagem em nova cidade, por deslocamento para assumir cargo ou função comissionada. Para solicitar o auxílio-moradia, você deve preencher o formulário Auxílio-moradia — Requerimento inicial, autuar processo no e-TCU, juntar a documentação e encaminhar para SCV/SecPessoas. No caso de comprovação anual, é necessário preencher o formulário Auxílio-moradia — Comprovação anual, assinar eletronicamente pelo e-TCU, juntar a documentação necessária e encaminhar para SPA/SecPessoas. O público-alvo deste serviço são autoridades ativas e servidores ativos. Para mais informações, você pode visitar o [link fornecido](#).

Fonte: ChatTCU.

### 6.1.3 Correção do caminho escolhido pelo modelo

Em alguns momentos, particularmente na presença de domínios com forte intersecção temática, é possível que o mecanismo de identificação de intenção classifique erroneamente uma pergunta. A título de exemplo, como há termos e assuntos presentes tanto na jurisprudência selecionada como na base de normativos do TCU, uma pergunta que o usuário tinha originalmente a intenção de dirigir à base de normativos pode ser direcionada à base



de jurisprudência, ou vice-versa. Em tais situações, é possível incluir nas perguntas prefixos como “Considerando a jurisprudência do TCU [...]” ou “Considerando a base de normativos do TCU, [...]”, ou ainda, corrigir o caminho a ser seguido *a posteriori*, de acordo com a resposta fornecida pelo sistema, conforme mostrado na **Figura 10**.

**Figura 10** – Correção de fluxo de respostas fornecidas pelo ChatTCU

**MM** Quais as hipóteses de prescrição de processos do Tribunal?

**Base de informações:** [Jurisprudência Selecionada](#)

As hipóteses de prescrição de processos no Tribunal de Contas da União (TCU) são regidas por diferentes normas e jurisprudências. A prescrição pode ocorrer tanto na esfera principal quanto na intercorrente.

Na esfera principal, a prescrição ocorre quando o processo fica paralisado por mais de cinco anos sem a prática de qualquer ato que seja considerado causa interruptiva da prescrição principal. O termo inicial da prescrição é a data do fato ou a data do vencimento da apresentação da prestação de contas ou a sua própria apresentação, o que ocorrer primeiro [1](#) [4](#).

Já a prescrição intercorrente ocorre quando o processo fica paralisado por mais de três anos, pendente de julgamento ou despacho, sem a prática de qualquer ato que evidencie o andamento regular do processo. A prescrição intercorrente pode ser interrompida por qualquer ato que evidencie o andamento regular do processo, excetuando-se pedido e concessão de vista dos autos, emissão de certidões, prestação de informações, juntada de procuração ou substabelecimento e outros atos que não interfiram de modo relevante no curso das apurações [2](#) [3](#).

Vale ressaltar que as causas impeditivas, suspensivas e interruptivas da prescrição principal também impedem, suspendem ou interrompem a prescrição intercorrente [3](#).

Além disso, atos processuais praticados pelas partes não interrompem a contagem do prazo prescricional, podendo, no máximo e a depender das circunstâncias de cada processo, suspender o prazo prescricional [3](#).

Por fim, é importante destacar que a prescrição é consequência sancionadora da inércia de quem tem o dever de agir, no caso, o dever de dar o impulso oficial ao processo [3](#).

[1](#) [2](#) [3](#)

É necessário revisar as respostas, pois podem conter dados incorretos.

**MM** Por favor, considere os normativos do TCU para responder à pergunta acima

**Base de informações:** [Normativos do TCU](#)

As hipóteses de prescrição de processos no Tribunal de Contas da União (TCU) são regidas pela Resolução TCU 344/2022 e pela Lei 9.873/1999, conforme estabelecido na Portaria SEGEXEX n° 10/2022.

A prescrição é o fenômeno jurídico que extingue o direito de ação em razão do decurso do tempo, e no caso do TCU, “prescrevem em cinco anos as pretensões punitiva e de ressarcimento” nos processos de controle externo. O termo inicial da contagem do prazo prescricional é definido da seguinte forma:

Utilizado: 2503 de 20000 palavras totais nessa conversa

Fonte: ChatTCU.

#### 6.1.4 Avaliação semiautomática de resultados

Durante o processo de indexação das bases citadas, com os objetivos tanto de checar a qualidade das soluções de busca adotadas, como também de comparar as diferentes



alternativas tecnológicas (ex.: busca textual tradicional *versus* busca vetorial), idealmente seria necessário ter uma base de perguntas/consultas. Adicionalmente, seria preciso checar a acurácia do mecanismo de detecção de intenção (classificação temática das perguntas digitadas pelo usuário).

Diante da escassez de uma base rotulada de perguntas/respostas por tema (jurisprudência, normativos ou sistemas administrativos), adotou-se a opção de gerar bases sintéticas, o que significa que, a partir de subconjuntos dos respectivos documentos de cada índice, solicitou-se ao GPT-4 que gerasse de uma a três perguntas relativas a cada um dos referidos documentos, do que resultou uma base com alguns milhares de perguntas sobre cada um dos três temas. Assim, houve a possibilidade de realizar avaliações quantitativas a partir de tal base, que se torna cada vez mais útil à medida que são adicionados novos documentos e novas funcionalidades de consulta ao sistema.

### 6.1.5 Upload de arquivos

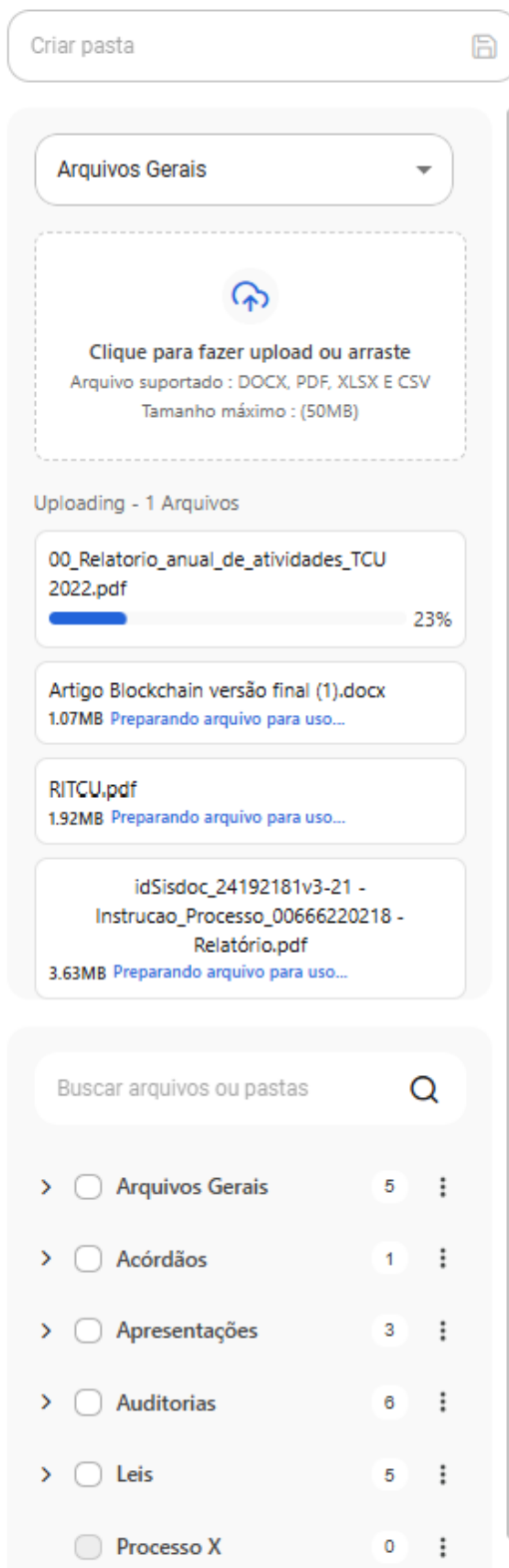
O recurso de *upload* de arquivos foi adicionado na última versão do ChatTCU, ampliando significativamente as possibilidades de interação com a aplicação. Ele permite ao usuário obter informações detalhadas e personalizadas a partir de seus próprios documentos. Assim, o usuário pode solicitar um resumo ou elaborar perguntas com base no conteúdo de tais documentos.

Cada arquivo que é carregado passa por um processo de preparação antes de ficar acessível para uso. O conteúdo do arquivo é segmentado em trechos que são indexados no motor de busca Azure AI Search. Para a recuperação dos dados, é usada uma pesquisa que combina a busca textual e a busca vetorial.

Atualmente, o ChatTCU permite o *upload* de documentos em formato PDF, planilhas (.xlsx), arquivos de texto (.docx) e arquivos de dados (.csv). O usuário pode fazer o carregamento de um ou mais arquivos simultaneamente. A **Figura 11** mostra um menu de *uploads* após o usuário ter subido quatro arquivos para a pasta Arquivos Gerais. Nesse exemplo, o conteúdo de um deles está sendo carregado para a aplicação, e os outros três estão na fase de preparação descrita acima.



**Figura 11** – Exemplo de menu de uploads de um usuário



Fonte: ChatTCU.



Para adicionar arquivos no contexto de uma conversa do ChatTCU, o usuário tem a opção de fazer o *upload* desses documentos no momento em que deseja utilizá-los ou escolher entre os documentos previamente carregados no seu menu de *uploads*.

O conteúdo dos arquivos selecionados é a base de conhecimento passada para o modelo de linguagem, que será utilizada para responder ao *prompt* do usuário. O modelo se baseará apenas nessas informações para gerar a resposta, evitando, assim, possíveis alucinações.

A **Figura 12** mostra um exemplo de conversa que utiliza arquivos como base de conhecimento. A **Figura 13** apresenta o detalhamento de uma das referências mencionadas na resposta, incluindo o nome do arquivo, o número da página, o número do trecho e o conteúdo do trecho em que o GPT-4 se baseou para gerar aquela resposta.

**Figura 12** – Conversa com arquivo como base de conhecimento

SB
Quais as vantagens na adoção de blockchain?

Base de informações: Arquivos

A adoção da tecnologia blockchain traz várias vantagens significativas para diversos setores, incluindo a indústria, a sociedade e até mesmo entre países. Uma das principais vantagens é a descentralização, que permite a transferência de controle e tomada de decisão de uma entidade centralizada para uma rede distribuída. Isso cria oportunidades para transações de interesse público serem realizadas de forma colaborativa, aumentando a transparência e a rastreabilidade. Essa característica permite que diversos setores da sociedade participem ativamente dos processos de auditoria ou de interesse público, possibilitando o surgimento de novos arranjos institucionais <sup>1</sup>.

Outra vantagem importante é a hiper transparência e auditabilidade. O livro-razão da blockchain é acessível e público para todos que fazem parte da rede, permitindo que os participantes vejam todo o histórico das transações em tempo real. Isso é especialmente relevante para aplicações governamentais, onde muitas informações devem ser públicas, permitindo a auditoria direta pela sociedade civil <sup>2</sup>.

A blockchain também oferece a vantagem da desintermediação e automação de transações, reduzindo a necessidade de processos complexos de reconciliação entre as partes e diminuindo custos. Isso é possível graças ao uso de contratos inteligentes, que são executados automaticamente de acordo com regras pré-definidas <sup>3</sup>.

Além disso, a tecnologia blockchain é conhecida por sua imutabilidade e integridade, utilizando técnicas criptográficas para proteger seus registros. Qualquer tentativa de adulteração é facilmente percebida, garantindo que os registros sejam imutáveis e seguros <sup>3</sup>.

Por fim, a blockchain pode ser utilizada no combate à fraude e corrupção, tanto como controle preventivo quanto detectivo. A utilização de tecnologias distribuídas permite a criação de trilhas de auditoria para rastrear operações de governo, favorecendo a abertura de dados e aumentando a confiabilidade, ao mesmo tempo que reduz as oportunidades de fraude e dificulta a ocorrência de delitos e comportamentos antiéticos <sup>1</sup>.

Palavras totais: 366 / 60000

↑
Digite sua mensagem

Arquivos Gerais

Artigo Blockchain versão final ...
×

➤

Atenção: O ChatTCU pode apresentar informações imprecisas, inclusive dados incorretos. Por isso, cheque as respostas

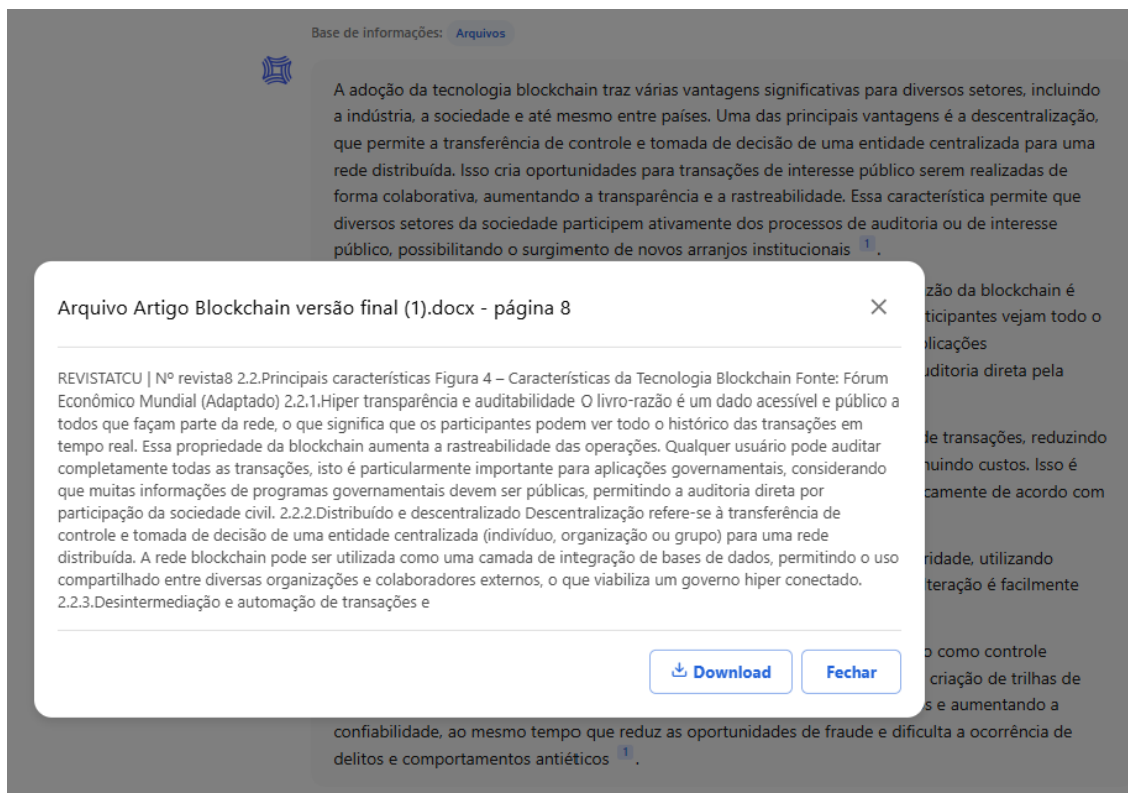
Fonte: ChatTCU.

Janeiro - Junho | 2024

39



Figura 13 – Detalhamento da referência de um arquivo



Fonte: ChatTCU.


O ChatTCU também possibilita que as perguntas sobre os documentos sejam feitas especificando um intervalo de páginas, usando expressões no *prompt* como “até a página X”, “da página X até a página Y” ou “a partir da página X”. A Figura 14 mostra um exemplo de uma pergunta feita pelo usuário em que ele delimita as páginas do arquivo.



**Figura 14** – Pergunta com intervalo de página

**SB** Quais as competências do plenário descritas até a página 23?


Base de informações: [Arquivos](#)

 As competências do Plenário do Tribunal de Contas da União (TCU), conforme descritas até a página 23 do documento, incluem uma variedade de responsabilidades importantes. Primeiramente, o Plenário, sob a direção do Presidente do Tribunal, tem a competência exclusiva para deliberar originariamente sobre várias matérias, como o parecer prévio relativo às Contas do Presidente da República, pedidos de informação ou solicitações sobre matéria da competência do Tribunal endereçados pelo Congresso Nacional, suas casas ou comissões, solicitações de pronunciamento formuladas pela comissão mista permanente de senadores e deputados, incidentes de uniformização de jurisprudência, conflitos de lei ou de ato normativo do poder público com a Constituição Federal em matéria da competência do Tribunal, e a fixação dos coeficientes destinados ao cálculo das parcelas a serem entregues aos estados, Distrito Federal e municípios, referentes aos recursos do Fundo de Participação dos Estados e do Distrito Federal (FPE) e do Fundo de Participação dos Municípios (FPM) <sup>1</sup>.


Além disso, o Plenário tem competências adicionais, como deliberar sobre recursos de revisão, decisões adotadas pelo Presidente sobre matéria administrativa, aprovar propostas de acordo de cooperação e instrumentos congêneres envolvendo transferência de recursos financeiros, aprovar o plano de controle externo, e aprovar os enunciados da Súmula da Jurisprudência do Tribunal <sup>2</sup>. Também é responsável por aprovar propostas relativas a projetos de lei que o Tribunal deva encaminhar ao Poder Legislativo, deliberar sobre a lista tríplice dos ministros-substitutos e dos membros do Ministério Público junto ao Tribunal para preenchimento de cargo de ministro, constituir comissões temporárias, apreciar questões administrativas de caráter relevante, deliberar sobre processos por ele avocados em razão de sua relevância, deliberar sobre processos remetidos pelo relator ou pelas câmaras, e deliberar sobre propostas de fixação de entendimento de especial relevância para a Administração Pública, sobre questão de direito, que somente poderão ser aprovadas por 2/3 dos ministros, incluindo ministros-substitutos convocados <sup>3</sup>.


Essas competências refletem a amplitude e a importância do papel do Plenário do TCU no controle e na fiscalização da administração pública, garantindo a legalidade, a legitimidade e a economicidade da gestão dos recursos públicos federais.

Palavras totais: 380 / 60000

 Digite sua mensagem

Arquivos Gerais

RITCU.pdf 



Fonte: ChatTCU.

## 6.2 Sumarização de peças e documentos processuais

Por fim, um recurso bastante útil do ChatTCU é a sumarização, não apenas de documentos ou conteúdos quaisquer, como especificamente de peças e documentos processuais custodiados pelo TCU, por meio da ferramenta de gestão processual e-TCU. Dessa forma, o usuário não precisa *a priori* ter o documento ao seu dispor para solicitar resumos, bastando, para isso, apenas conhecer o número de peça e do processo ou o número do documento. Entretanto, é importante ressaltar que o usuário precisa ter perfil de acesso ao referido conteúdo, que pode conter até mesmo dados sigilosos ou sensíveis que só podem ser acessados por determinados perfis de usuário.




Em situações nas quais o documento possui conteúdo de tamanho superior ao tamanho máximo suportado pelo modelo de IA subjacente, o sistema o quebra em subdocumentos menores. A partir daí, o modelo de IA é instruído a gerar um resumo incremental a partir de cada um dos subdocumentos, em um mecanismo de refinamentos sucessivos.




Como exemplo das interações, temos as Figuras 15 e 16.

**Figura 15** – Exemplo de sumarização de peça processual

**SB** Resuma a peça 3 do processo 031.710/2022-0

Base de informações: **Processos e-TCU**

 A peça 3 do processo TC 031.710/2022-0 corresponde ao Aviso nº [REDACTED]. Este documento comunica o recebimento da Proposta de Fiscalização e Controle (PFC) Nº [REDACTED], encaminhada pela [REDACTED], sob a autoria do [REDACTED] e outros. A proposta solicita uma fiscalização sobre os procedimentos de elaboração, aplicação e gestão do [REDACTED]. O aviso destaca que o documento está sendo tratado com urgência e tramitação preferencial, conforme estabelecido pela Resolução-TCU nº 215/2008, e foi encaminhado à Secretaria-Geral de Controle Externo (Segecex) do TCU para as providências necessárias. O documento é assinado eletronicamente pelo [REDACTED], que estava exercendo a Vice-Presidência do TCU na ocasião, e é endereçado ao [REDACTED].


Legenda: as tarjas mostradas na figura são apenas para fins de publicação nesta Revista, não tendo sido gerada pelo ChatTCU.




Fonte: ChatTCU.

**Figura 16** – Exemplo de sumarização de documento eletrônico

**MM** Resuma o conteúdo do documento 73.087.887-0.

Base de informações: **Processos e-TCU**

 O documento número 73.087.887-0 é um processo de execução fiscal que envolve a União Federal e outros interessados, relacionado a um dano ao erário. A Advocacia Geral da União (AGU) analisou o caso e constatou que o recolhimento de um valor, que foi solicitado para restituição, acabou sendo realizado na conta do Tribunal de Contas da União (TCU), o que impede a AGU de proceder com a devolução desse valor. Portanto, a recomendação é que a solicitação de restituição seja encaminhada diretamente ao TCU. O documento também menciona detalhes de um comprovante de levantamento judicial e informações sobre uma Guia de Recolhimento da União (GRU) associada ao processo, que indica o pagamento de um valor específico à conta do TCU.

Fonte: ChatTCU.





## 7. PRÓXIMOS PASSOS

O ChatTCU está em constante evolução, buscando sintonia com as necessidades dos usuários finais. Dentre as possibilidades vislumbradas para evolução da ferramenta, pode-se citar:

- Integração com o Microsoft Bing<sup>9</sup> e Microsoft Copilot<sup>10</sup>.
- Geração de consultas a bases de dados relacionais custodiadas pelo TCU, permitindo que o usuário especifique sua consulta em linguagem natural, sem exigência de conhecimentos em SQL ou da estrutura do banco de dados a ser consultado (o esquema do banco de dados pode ser utilizado internamente no *prompt* para que o LLM gere o SQL correspondente à consulta descrita pelo usuário).
- Inserção da ferramenta como um assistente embarcado acessível a partir de outras aplicações, sejam aplicações do Microsoft Office<sup>11</sup> ou mesmo sistemas internos do Tribunal (gestão processual, pesquisa interna, instrução de recursos etc.).
- Descentralização do processo de criação de assistentes especialistas com o ChatTCU, permitindo que equipes do Tribunal instanciem versões da ferramenta que sejam especialistas em assuntos específicos (determinadas auditorias/fiscalizações, legislação recente pertinente a determinado tema, documentação ou suporte a um sistema etc.), evitando, com isso, que o ChatTCU geral tenha que decidir por si só qual base de conhecimento utilizar para produzir sua resposta, além de permitir que o usuário/desenvolvedor já determine as ferramentas e o conhecimento que serão exclusivamente utilizados por esse especialista. Exemplo: em vez de perguntar para o ChatTCU geral informações sobre determinada jurisprudência e confiar que ele entenda perfeitamente a intenção do usuário, já seria possível escolher o especialista que só trata de jurisprudência selecionada.
- Suporte a outros modelos concorrentes e/ou complementares aos modelos da Open AI, a exemplo de modelos fechados de outros provedores (Robertson; Zaragoza, 2009; Gemini Team Google *et al.*, 2024) ou modelos abertos (Anthropic, 2024) que venham a ser especializados para tarefas específicas, por meio de treinamento adicional.
- Suporte a arquitetura multiagente que permita combinar diferentes modelos subjacentes, de forma a quebrar tarefas complexas em subtarefas e permitir que diferentes modelos de linguagem, atuando como agentes inteligentes, cooperem entre si para resolução do problema (Meta, 2024). Há estudos que apontam que tal abordagem tende a gerar resultados superiores ao uso de um único modelo para resolução de uma tarefa complexa. É possível ainda que um agente seja capaz de, autonomamente, dividir a tarefa complexa em subtarefas, processo conhecido como planejamento (Wei *et al.*, 2023; Wu *et al.*, 2023; Shen *et al.*, 2023).

---

9 Disponível em: <https://www.bing.com>. Acesso em: 15 maio 2024.

10 Disponível em: <https://www.microsoft.com/pt-br/microsoft-copilot>. Acesso em: 15 maio 2024.

11 Disponível em: <https://www.microsoft.com/pt-br/microsoft-365/microsoft-office>. Acesso em: 15 maio 2024.



Por fim, na esteira do uso bem-sucedido da IA generativa no TCU, concretizado na aplicação ChatTCU, vislumbra-se ainda outras possibilidades de aplicação desse novo paradigma. Um exemplo clássico é o desenvolvimento de classificadores explicáveis, que não apenas fornecem uma resposta, mas também explicam seu raciocínio, favorecendo maior transparência diante do desafio imposto pelos sistemas de aprendizagem de máquina nos últimos anos, muitas vezes implementados como verdadeiras “caixas-pretas”. Adicionalmente, o uso da IA generativa tem-se mostrado cada vez mais promissor no campo do desenvolvimento de *software*, a exemplo de correção automática de defeitos e geração de casos de teste. Tais aplicações podem auxiliar times não apenas do TCU, mas, obviamente, são exemplos “universais” de aplicações da tecnologia para maior produtividade, eficácia e eficiência, podendo ser replicados para demais entes da Administração Pública.

## 8. CONSIDERAÇÕES FINAIS

O ChatTCU emerge como uma plataforma transformadora no âmbito do Tribunal de Contas da União (TCU), destacando-se pela sua capacidade polivalente de servir a todos os auditores, independentemente de seu nível de expertise em Tecnologia da Informação (TI) ou Inteligência Artificial (IA). Essa ferramenta democratiza o acesso à tecnologia de ponta, permitindo que os auditores, por meio de comandos simples em português, configurem e adaptem a ferramenta para atender às suas necessidades específicas. Essa característica evidencia o potencial do ChatTCU de se tornar um instrumento valioso no arsenal de cada auditor, facilitando uma variedade de tarefas, desde a geração de relatórios até a análise de jurisprudência, sem a necessidade de conhecimento prévio em programação ou IA.

A base tecnológica sobre a qual o ChatTCU é construído, centrada nos modelos de linguagem de grande escala (LLMs), está em constante evolução. A rápida progressão nesse campo da IA assegura que o ChatTCU não apenas mantenha sua relevância, mas também amplie suas capacidades, oferecendo aos auditores do TCU funcionalidades cada vez mais sofisticadas, úteis e de alta qualidade. Esse compromisso com a inovação contínua garante que a ferramenta esteja sempre alinhada com o estado da arte em tecnologia, possibilitando que o TCU antecipe e se adapte às mudanças no cenário tecnológico, mantendo-se à frente em eficiência e eficácia de auditoria.

Além disso, reconhece-se que os melhores usos dos LLMs para o controle externo ainda estão por ser plenamente explorados. A chave para desbloquear o pleno potencial dessa tecnologia reside na experimentação descentralizada conduzida pelos auditores. Equipados com vasto conhecimento e experiência em auditoria, esses profissionais estão em uma posição única para investigar e identificar as aplicações mais promissoras dos LLMs nos respectivos processos de trabalho. Essa abordagem colaborativa e exploratória não apenas fomenta a inovação, mas também assegura que o desenvolvimento do ChatTCU seja guiado pelas necessidades reais e práticas do tribunal.

Portanto, as considerações finais destacam o ChatTCU como um exemplo emblemático de como a aplicação de IA no setor público pode impactar a maneira como o trabalho de auditoria é realizado, promovendo maior eficiência, precisão e inovação. A ferramenta está posicionada



para evoluir e se adaptar às rápidas mudanças no campo da IA, garantindo que os auditores do TCU estejam sempre equipados com as melhores e mais atualizadas tecnologias. À medida que se explora o recém-descoberto universo de aplicações de LLMs para o controle externo, o ChatTCU se destaca como uma plataforma que não apenas promete transformar os processos de auditoria, mas também empodera cada auditor para ser um agente ativo nessa transformação.

## REFERÊNCIAS

- ANTHROPIC. **The claude 3 model family**: Opus, Sonnet, Haiku. Anthropic, 2024. Disponível em: [https://www-cdn.anthropic.com/de8ba9b01c9ab7cbabf5c33b80b7bbc618857627/Model\\_Card\\_Claude\\_3.pdf](https://www-cdn.anthropic.com/de8ba9b01c9ab7cbabf5c33b80b7bbc618857627/Model_Card_Claude_3.pdf). Acesso em: 15 maio 2024.
- GEMINI Team Google *et al.* Gemini: a family of highly capable multimodal models. **Arxiv**, [s. l.], abr. 2024. Disponível em: <https://arxiv.org/abs/2312.11805>. Acesso em: 15 maio 2024.
- GROOVE, Andrew S. **High Output Management**. University of Michigan: Random House, 1983.
- LEWIS, Patrick; PEREZ, Ethan; PIKTUS, Aleksandra; PETRONI, Fabio; KARPUKHIN, Vladimir; GOYAL, Naman; KÜTTLER, Heinrich; LEWIS, Mike; YIH, Wen-tau; ROCKTÄSCHEL, Tim; RIEDEL, Sebastian; KIELA, Douwe. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. **Proceedings of the 34th International Conference on Neural Information Processing Systems**, 2020.
- META. Build the future of AI with Meta Llama 3. **Meta**, 18 abril 2024. Llama 3. Disponível em: <https://llama.meta.com/llama3/>. Acesso em: 15 maio 2024.
- OpenAI *et al.* GPT-4 Technical Report. **Arxiv**, [s. l.], mar. 2023. Disponível em: <https://arxiv.org/abs/2303.08774>. Acesso em: 15 maio 2024.
- ROBERTSON, Stephen E.; ZARAGOZA, Hugo. The probabilistic relevance framework: BM25 and beyond. **Foundations and Trends in Information Retrieval**, [s. l.], v. 3, n. 4, p. 333-389, abr. 2009.
- SHEN, Yongliang; SONG, Kaitao; TAN, Xu; LI, Dongsheng; LU, Weiming; ZHUANG, Yueting. HuggingGPT: solving ai tasks with ChatGPT and its friends in hugging face. **Arxiv**, [s. l.], dez. 2023. Disponível em: <https://arxiv.org/abs/2303.17580>. Acesso em: 15 maio 2024.
- WEI, Jason; WANG, Xuezhi; SCHUURMANS, Dale; BOSMA, Maarten; ICHTER, Brian; XIA, Fei; CHI, Ed; LE, Quoc; ZHOU, Denny. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. **Arxiv**, [s. l.], jan. 2023. Disponível em: <https://arxiv.org/abs/2201.11903>. Acesso em: 15 maio 2024.
- WU, Qingyun; BANSAL, Gangan; ZHANG, Jieyu; WU, Yiran; LI, Beibin; ZHU, Erkang; JIANG, Li; ZHANG, Xiaoyun; ZHANG, Shaokun; LIU, Jiale; AWADALLAH, Ahmed Hassan; WHITE, Ryan W; BURGER, Doug; WANG, Chi. AutoGen: enabling next-gen LLM applications via multi-agent conversation. **Arxiv**, [s. l.], out. 2023. Disponível em: <https://arxiv.org/abs/2308.08155>. Acesso em: 15 maio 2024.