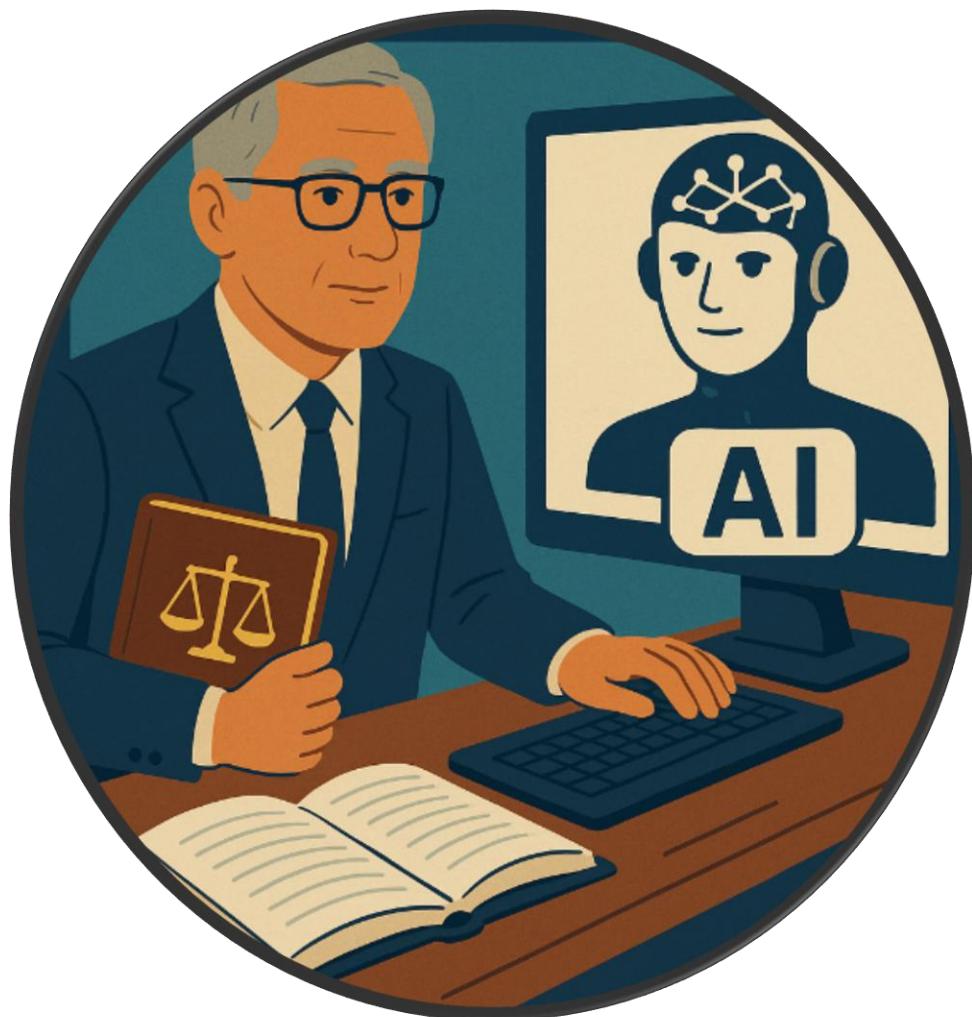


José Eduardo de Souza Pimentel

## A IA GENERATIVA NA PROMOTORIA: Aprendizado, Limitações e o que Funciona Agora



## **José Eduardo de Souza Pimentel**

Promotor de Justiça em Piracicaba, com atuação criminal, mestre em Direito pela PUC/SP, especialista em Inteligência Artificial e Computacional pela UFV e em Gestão e Governança de Tecnologia da Informação pelo SENAC/SP. Possui graduação em Tecnologia da Segurança da Informação pela FATEC de Americana/SP.

## **Versionamento**

- 30/04/2025: 1<sup>a</sup>. edição

Procure pelas novas versões deste trabalho em [https://jespimentel.github.io/ia\\_gen\\_na\\_promotoria/](https://jespimentel.github.io/ia_gen_na_promotoria/)

## **Licença**

[A IA Generativa na Promotoria: aprendizado, limitações e o que funciona agora](#) © 2025 by [José Eduardo de Souza Pimentel](#) is licensed under [CC BY-NC-SA 4.0](#)

Esta licença permite a distribuição, adaptação e criação a partir deste material, em qualquer meio ou formato, apenas para fins não comerciais, dando-se crédito ao autor. Os trabalhos modificados ou adaptados devem ser licenciados sob os mesmos termos.

## CONTEÚDO

<b>1. INTRODUÇÃO.....</b>	<b>5</b>
<b>2. SOBRE O NOSSO CURSO.....</b>	<b>6</b>
<b>3. O QUE É IA GENERATIVA? .....</b>	<b>8</b>
<b>4. POSSO USAR A IA GENERATIVA NA PROMOTORIA? .....</b>	<b>12</b>
<b>5. COMO UMA LLM PODE ME AJUDAR?.....</b>	<b>16</b>
<b>6. PRIVACIDADE E “ALUCINAÇÕES” .....</b>	<b>18</b>
<b>7. PRINCIPAIS MODELOS .....</b>	<b>22</b>
<b>8. ENGENHARIA DE PROMPT .....</b>	<b>24</b>
8.1. PROMPTS: INSTRUÇÃO OU PERGUNTA .....	24
8.2. ELEMENTOS ESSENCIAIS DE UM PROMPT EFICAZ .....	25
8.3. DICAS GERAIS PARA ELABORAÇÃO DE BONS PROMPTS.....	26
8.4. TÉCNICAS AVANÇADAS DE PROMPTING.....	29
8.5. ENCADEAMENTO DE PROMPTS.....	30
8.6. OBSERVAÇÕES IMPORTANTES.....	31
<b>9. OUTROS CONCEITOS RELACIONADOS A PROMPTS.....</b>	<b>33</b>
<b>10. A IA GENERATIVA NO FLUXO DE TRABALHO .....</b>	<b>37</b>
10.1. VISÃO GERAL DO CAPÍTULO .....	37
10.2. TODO O MARKDOWN DE QUE EU PRECISO .....	38
10.3. O FRAMEWORK BÁSICO .....	39
10.4. CONFIGURAÇÕES BÁSICAS DOS PRINCIPAIS MODELOS .....	41
10.5. EXEMPLOS DE USO E RESPECTIVOS PROMPTS .....	56
10.5.1. <i>Pesquisa doutrinária e jurisprudencial</i> .....	56
10.5.2. <i>Pesquisa em fontes específicas</i> .....	59
10.5.3. <i>Resumos de depoimentos escritos a partir de captura de tela</i> .....	61
10.5.4. <i>Resumos de APFD para audiência de custódia</i> .....	63
10.5.5. <i>Resumos estruturados de processos (com mapas mentais e linha do tempo)</i> .....	64
10.5.6. <i>Análise aprofundada de inquérito policial</i> .....	66
10.5.7. <i>Identificação de contradições entre depoimentos</i> .....	69
10.5.8. <i>Produção de peças processuais: denúncia</i> .....	70
10.5.9. <i>Produção de peças processuais: alegações finais</i> .....	73
10.5.10. <i>Criação de GEM para a elaboração de contrarrazões</i> .....	79
10.5.11. <i>Criação de “Meu GPT” para a elaboração de pareceres</i> .....	82

10.5.12. Extração de áudio de arquivo de gravação de audiência (com conversão do formato para “mp3”)	86
10.5.13. Diarização de audiências judiciais.....	87
10.5.14. Diarização de audiência judicial longa (pelo Word com correção feita pelo modelo de linguagem)	88
10.5.15. Análise de “Certidão Estadual de Distribuições Criminais” (certidão de objeto e pé) .....	90
10.5.16. Obtenção de hashes e metadados de arquivos.....	91
10.5.17. Descrição de cenas.....	92
10.5.18. Obtenção de frames.....	93
10.5.19. Tratamento de imagens.....	94
10.5.20. Brainstorm: crimes contra a honra .....	95
10.5.21. Atuação adversarial .....	98
10.5.22. Interpretação leiga de laudo.....	100
10.5.23. Encadeamento de prompts.....	102
<b>11. REAPROVEITAMENTO E COMPARTILHAMENTO DE PROMPTS .....</b>	<b>107</b>
<b>12. USOS AVANÇADOS.....</b>	<b>111</b>
<b>13. MEU ESTAGIÁRIO DEVERIA USAR IA GENERATIVA? .....</b>	<b>113</b>
<b>14. POR ONDE SEGUIR? .....</b>	<b>115</b>
<b>15. CONCLUSÕES .....</b>	<b>117</b>
<b>REFERÊNCIAS .....</b>	<b>119</b>

## **1. INTRODUÇÃO**

Quando o ChatGPT surgiu, em novembro de 2022, vivemos uma onda de empolgação sem precedentes. Muitos compararam seu potencial disruptivo ao surgimento da Internet, prevendo uma revolução completa nos nossos modos de trabalhar e produzir.

Mais de dois anos já se passaram e nos encontramos em um contexto bem mais modesto.

As corporações – incluindo o Ministério Público de São Paulo – ainda enfrentam barreiras significativas para implementar a IA Generativa em escala e estabelecer bons casos de uso.

E muitos de nós, promotores de justiça e servidores da instituição que tivemos contato com essa tecnologia por iniciativa própria, interagindo com os chats que nos são disponibilizados até mesmo gratuitamente por grandes provedores, ainda temos dúvidas sobre como incorporá-la ao nosso fluxo de trabalho e se isso, de fato, vale a pena.

Neste curso, vamos deixar de lado as expectativas mais exageradas e conversarmos francamente sobre experiências e aplicações concretas dessas ferramentas no dia a dia da Promotoria de Justiça.

Veremos juntos se, apesar das imprecisões, o uso da IA Generativa nos traz alguma vantagem no desempenho das nossas atribuições.

Através de casos reais e demonstrações práticas, descobriremos, no caminho, como extrair os melhores resultados das soluções disponíveis, com foco no que funciona agora.

## **2. SOBRE O NOSSO CURSO**

O curso é eminentemente prático. Não nos interessa saber por que as ferramentas funcionam, mas sim como aplicá-las na nossa rotina de trabalho, com responsabilidade e de forma a sermos mais produtivos.

Apresentaremos muito brevemente os fundamentos de base da IA Generativa, sem maior preocupação com o tecnicismo e apenas para entender melhor as possibilidades e as limitações das soluções que estão à nossa disposição.

Seguiremos com exemplos práticos, muitos dos quais colhidos nos casos reais em que as IAs Generativas foram empregadas, com nossas opiniões sinceras sobre os resultados obtidos, para que sirvam de inspiração ao leitor diante de suas próprias demandas.

Desde já, vale um alerta.

**Não existe um jeito certo ou único de criar prompts ou de usar a IA Generativa.**

O cargo de “engenheiro de prompt”, anunciado como uma nova e promissora profissão, não vingou nas empresas e, possivelmente, será reduzido a apenas mais uma habilidade desejável daqueles que executam tarefas intelectuais.

As lições aqui encontradas se constituem em **meras sugestões de utilização dos Modelos de Linguagem de Grande Escala (LLMs) e não devem ser entendidas como produto de um conhecimento definitivo sobre o assunto**.

Use a criatividade para gerar novos prompts para os seus casos de uso.

Cada caso é um caso e os melhores resultados decorrem da interação do usuário com o modelo de sua preferência, refinando os novos comandos de acordo com as respostas anteriormente obtidas, até alcançar o resultado pretendido.

Mas observe: os outputs são únicos e, em geral, não reproduzíveis<sup>1</sup>!

Nem sempre você obterá os mesmos resultados mostrados neste trabalho, já que os LLMs são, como veremos mais à frente, ferramentas probabilísticas<sup>2</sup>.

A propósito, usamos nos nossos exemplos os principais modelos comerciais que hoje estão disponíveis para o usuário final. A preferência por um ou por outro é sempre uma decisão pessoal, pois cada qual tem seus pontos fortes e fracos. Os roteiros aqui expostos talvez precisem ser adaptados à sua escolha.

Bons prompts e respectivos resultados devem ser compartilhados com colegas, para que nos adaptemos mais rapidamente a este novo tempo em que a IA Generativa já se faz presente na nossa rotina de trabalho.

---

<sup>1</sup> É por isso que, às vezes, diante de uma resposta ruim do modelo, é melhor iniciar um novo chat, ainda que para um mesmo prompt.

<sup>2</sup> Por causa dessa característica, há quem diga que são criativas. Mais à frente, veremos como controlar essa “criatividade”.

### 3. O QUE É IA GENERATIVA?

A Inteligência Artificial é um campo da ciência que se ocupa, desde os anos 50, em estudar como máquinas podem ser usadas para desempenhar tarefas que exigem raciocínio ou intervenção humana<sup>3</sup>.

Na área da computação, a IA se apresenta, atualmente, na forma de machine learning (aprendizado de máquina) e de deep learning (aprendizado profundo).

O aprendizado de máquina tornou-se relevante em nossa era em virtude da abundância de dados e do crescente aumento do poder computacional. Algoritmos sofisticados passaram a ser capazes de, por um processo iterativo, identificar padrões, fazer estimativas e estabelecer classificações em grandes bancos de dados. Nessas tarefas, o computador calcula, a partir dos chamados dados de treinamento, os coeficientes (ou pesos) de uma função matemática que converge para um resultado provável diante de dados novos que lhe são apresentados (generalização). O machine learning, ao lado de modelos puramente estatísticos, é usado com sucesso para resolver problemas em bases de dados estruturadas (tabelas).

A evolução das placas de processamento gráfico e tensorial (GPUs e TPUs), bastante eficientes para as operações matriciais (de múltiplas dimensões), permitiram a construção de soluções mais sofisticadas, chamadas de deep learning, próprias para trabalhar, também, com dados não estruturados (como vídeos, sons e imagens).

O aprendizado profundo ocorre nas denominadas redes neurais profundas, ou seja, um sistema complexo de neurônios artificiais, organizados em múltiplas camadas (daí o profunda), que se estendem da entrada à saída do sistema.

Cada neurônio é um pequeno processador que calcula os pesos (coeficientes) e os vieses (constantes) de uma função de ativação não linear (permitindo, portanto, a modelagem de relações complexas e não lineares da informação).

---

<sup>3</sup> Para saber mais, confira o nosso e-book “[Direito e ética da inteligência artificial e dos algoritmos de caixa preta](#)”, disponível na Amazon.

Durante o treinamento, a rede processa repetidamente os dados, ajustando iterativamente pesos e vieses por um algoritmo de retropropagação, que calcula o erro da saída e o propaga de volta pelo conjunto de neurônios para que os ajustes sejam feitos. É por causa disso que se diz que o modelo aprende com os dados. Ele não é programado, no sentido tradicional, em que as regras de negócio são definidas explicitamente e no próprio código pelo programador. O objetivo final é, como já dissemos, o de que o modelo possa generalizar, isto é, fornecer as melhores respostas para os dados novos que lhe forem apresentados.

A IA Generativa usa redes neurais profundas para aprender os padrões dos dados de treinamento.

A designação “generativa” está associada à aptidão que essas soluções possuem para gerar conteúdo novo, como textos, sons, imagens, vídeos ou códigos de programação.

O conteúdo novo resulta da capacidade que os modelos têm de aprender como as informações se organizam e estão estatisticamente distribuídas no corpus de treinamento. O resultado é inédito, ou seja, não é cópia dos referidos dados.

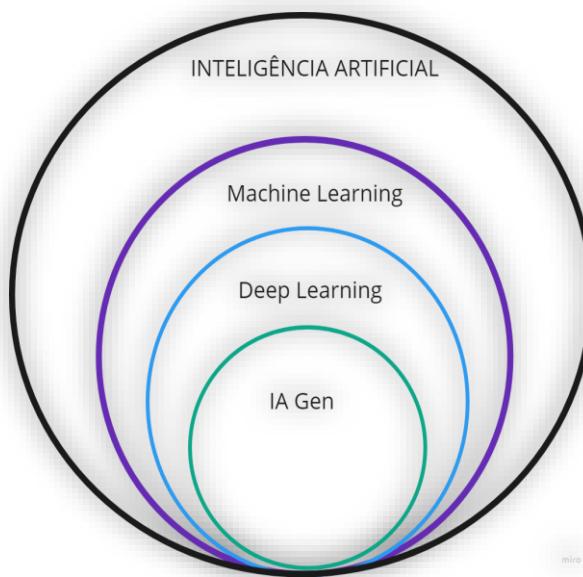
The image displays two screenshots of a ChatGPT interface. Both screenshots show a user input field containing the text "O céu é". In the top screenshot, the AI response is: "O céu é azul durante o dia devido à dispersão da luz solar na atmosfera e estrelado à noite, revelando a vastidão do universo. O que você queria dizer?" Below the response are five small icons: a square, a triangle, a circle, a double arrow, and a question mark. In the bottom screenshot, the AI response is: "...imenso, fascinante e cheio de mistérios! Você estava pensando em algo poético, científico ou filosófico? 😊" Below this response are also five small icons: a square, a triangle, a circle, a double arrow, and a question mark. At the bottom of the image, a caption reads: "Nos exemplos acima, o mesmo prompt produziu resultados bastante diferentes." followed by "Fonte: autor".

Os Modelos de Linguagem de Grande Escala (LLMs) são um tipo específico de IA Generativa, treinados em vastos conjuntos de dados textuais. Esses modelos aprendem a relação estatística entre palavras e frases, o que faz com que gerem textos coerentes e relevantes em resposta aos comandos dos usuários, os chamados prompts.

Mas note bem: **ainda que possamos nos surpreender com a coerência das respostas, elas nada mais são do que uma sequência provável, sob o ponto de vista estatístico, que se segue aos dados de entrada** (prompt)<sup>4</sup>.

Mais que isso, **modelos desse tipo não possuem conhecimentos factuais**. É a exposição maciça a informações durante o treinamento que lhes permite capturar um amplo espectro do conhecimento humano. Os LLMs não armazenam fatos como dados, mas os apresenta como associações de palavras e frases.

O que eu quero dizer com isso? **Não confie muito no que LLMs dizem!**



Fonte: autor

<sup>4</sup> A rigor, o modelo aprende a prever o próximo token no pré-treinamento, ao qual seguem uma fase de ajuste fino (instruction finetuning), em que ele aprende a seguir instruções (prompts), e outra, de aprendizado por reforço com feedback humano (RLHF), em que é recompensado pelas respostas preferidas por humanos. Entre os usuários de modelos de linguagem há os que preferem um a outro pela aptidão que demonstram em seguir instruções, isto é, compreender os prompts e, particularmente, suas restrições.

Os LLMs (Grandes Modelos de Linguagem) evoluíram para os LMMs (Grandes Modelos Multimodais), de tal sorte que, atualmente, é bastante comum encontrar modelos treinados em conjuntos massivos de dados multimodais, compostos por texto, áudio, imagens e vídeo, tornando-os aptos a interpretar, gerar e correlacionar informações provenientes de diversas fontes<sup>5</sup>.

Essa característica, potencializa as capacidades dos modelos, capacitando-o a combinar texto, imagens, áudio e vídeo em respostas mais assertivas e contextualizadas.

Ao lado dos LLMs e LMMs, existem, ainda, os chamados modelos de difusão, que são treinados a partir de "ruído". Esses sistemas aprendem a reverter processos de adição progressiva de ruído sobre imagens reais, podendo gerar imagens sintéticas realistas e, mais recentemente, também áudio, vídeo e outras estruturas de dados complexas.

A progressiva convergência entre LLMs, LMMs e modelos de difusão delineia um ecossistema de Inteligência Artificial<sup>6</sup> capaz de produzir e correlacionar, de forma inédita e em escala, textos, imagens, áudios e vídeos sintéticos, impactando significativamente o trabalho de produção intelectual – como o que fazemos nas Promotorias de Justiça – e suscitando uma série de preocupações relativas ao modo de produção, à proteção de dados pessoais e à autoria intelectual.

---

<sup>5</sup> Trata-se, evidentemente, de uma simplificação. Embora muitos modelos de linguagem estejam se tornando multimodais (LMMs), LLMs (especializados em texto) ainda são desenvolvidos e muito relevantes. (nota do autor)

<sup>6</sup> Saiba mais sobre IA em: [IBM: O que é inteligência artificial \(IA\)?](#).

#### 4. POSSO USAR A IA GENERATIVA NA PROMOTORIA?

Pode e deve, ressalvados os casos de “segredo de justiça” (art. 189 do CPC<sup>7</sup> e art. 234-B do Código Penal<sup>8</sup>).

Não somente promotores de justiça, como também advogados, defensores públicos e juízes, com seus assistentes, usarão, cada vez mais, a IA Generativa para aumentar sua produtividade e executar mais rapidamente as tarefas de baixa e média complexidade que lhes cabem. Há um enorme potencial a ser explorado na aplicação da IA no Direito, especialmente no campo do processamento de linguagem natural e não estruturada, área em que as LLMs têm se mostrado bastante promissoras.

De fato, assim que o ChatGPT foi lançado, um cidadão pediu formalmente ao Conselho Nacional de Justiça que proibisse o uso dessa tecnologia “na confecção de atos processuais pelos juízes brasileiros”.

O caso foi objeto do PCA nº 0000416-89.2023.2.00.0000, que, a seu termo (em 21/06/2024), foi julgado improcedente, por decisão unânime do plenário<sup>9</sup>.

**Significa dizer que o CNJ, instado a se manifestar sobre a matéria, consentiu que os magistrados utilizassem LLMs para os apoiar na prestação jurisdicional.**

---

<sup>7</sup> Art. 189. Os atos processuais são públicos, todavia tramitam em segredo de justiça os processos:

I - em que o exija o interesse público ou social;

II - que versem sobre casamento, separação de corpos, divórcio, separação, união estável, filiação, alimentos e guarda de crianças e adolescentes;

III - em que constem dados protegidos pelo direito constitucional à intimidade;

IV - que versem sobre arbitragem, inclusive sobre cumprimento de carta arbitral, desde que a confidencialidade estipulada na arbitragem seja comprovada perante o juízo.

§ 1º O direito de consultar os autos de processo que tramite em segredo de justiça e de pedir certidões de seus atos é restrito às partes e aos seus procuradores.

§ 2º O terceiro que demonstrar interesse jurídico pode requerer ao juiz certidão do dispositivo da sentença, bem como de inventário e de partilha resultantes de divórcio ou separação.

<sup>8</sup> Art. 234-B. Os processos em que se apuram crimes definidos neste Título correrão em segredo de justiça (nota do autor: refere-se ao Título VI do Código Penal, Dos crimes contra a dignidade sexual).

<sup>9</sup> Confira o acórdão no portal do [Conjur](#).

Mais recentemente (em 18/02/2025), o CNJ atualizou a [Resolução nº 332/2020](#), que trata da Inteligência Artificial no Poder Judiciário, que agora prevê, em bom Português, a **possibilidade de que magistrados e servidores do Poder Judiciário usem a IA Generativa como auxílio à gestão ou mesmo em apoio à decisão** (art. 19).

A norma autoriza, até mesmo, o uso de LLMs contratadas diretamente, mediante assinatura ou cadastro de natureza privada, quando o tribunal não dispuser de solução corporativa (art. 19, § 2º), proibindo-se, apenas neste caso, a utilização com “dados sigilosos ou protegidos por segredo de justiça” (art. 19, inc. IV).

Seria melhor, a nosso ver, dada a velocidade com que surgem modelos cada vez mais poderosos, que o usuário pudesse optar por alguma solução privada, ainda que a instituição lhe disponibilize alguma opção, já que, **nos casos de chats, estamos falando de ferramentas que se destinam, especialmente, ao aumento da produtividade pessoal.**

A Resolução contém, ainda, uma regra bastante curiosa – e, muito cá entre nós, de difícil fiscalização<sup>10</sup> – exigindo que, **quando houver emprego de IA Generativa para auxílio à redação de ato judicial, “tal situação poderá ser mencionada no corpo da decisão, a critério do magistrado, sendo, porém, devido o registro automático no sistema interno do tribunal, para fins de produção de estatísticas, monitoramento e eventual auditoria”** (art. 19, § 6º).

O CNJ não deveria ter enveredado por esse caminho.

A leitura da norma já deixa suficientemente claro que **a IA Generativa terá sempre caráter auxiliar**, servindo como uma espécie de “copiloto”, que auxilia seus usuários na realização de alguma etapa mais tediosa do trabalho, **sem lhes retirar a responsabilidade pelo resultado.**

O novel regramento institucionaliza aquilo que o acórdão do PCA nº 0000416-89.2023.2.00.0000 já havia proclamado: “ferramentas dessa natureza oferecem um

---

<sup>10</sup> Embora existam soluções concebidas para detectar textos gerados por LLMs, a profusão de modelos em constante evolução e a possibilidade de edição torna o processo de detecção bastante desafiador.

potencial significativo para aprimorar a eficiência e a eficácia do sistema judicial”, mas a supervisão humana deve ocorrer em todas as etapas do processo judicial.

Não percamos isso de vista.

Como veremos mais adiante, as LLMs têm muito a contribuir conosco na realização de diversas tarefas da prática forense.

Essas soluções são muito boas – e estão melhorando a cada dia – para a análise e resumo de documentos, elaboração de minutas e peças processuais, revisão ortográfica e gramatical, pesquisa de doutrina e jurisprudência, análise e estruturação de dados e metadados de feitos judiciais, diarização de audiências, entre outras práticas.

**Note-se, porém, que, no atual estado da arte, o campo em que a IA Generativa é verdadeiramente útil não é o da decisão.**

As LLMs apenas contribuem, nos dias de hoje, para que não façamos nossas manifestações processuais do zero, isto é, partindo de uma folha em branco. **Sempre haverá um rascunho que ela consegue nos entregar para acelerar o trabalho.**

De toda forma, o promotor de justiça ou juiz que assina a peça continua responsável pelo seu conteúdo, tal como acontece quando a minuta é preparada pela assessoria ou por estagiário.

O Conselho Nacional do Ministério Público segue os passos do CNJ: incentiva a utilização responsável dessas ferramentas e vem investindo em treinamentos, forte no entendimento de que a IA poderá contribuir para uma melhor prestação de serviço dos MPs à sociedade.

Até que haja normatização específica, faz sentido seguirmos a orientação do Poder Judiciário, **abstendo-nos, portanto, de usar plataformas privadas de IA**

**Generativa contratadas individualmente nos casos de sigilo e “segredo de justiça”<sup>11</sup>.**

Esperamos sinceramente que o CNMP não imponha maiores restrições ao uso das LLMs, pois, se o fizer, haverá assimetria das capacidades, em termos de eficiência, do Promotor de Justiça em relação aos demais atores da Justiça, incluindo advogados, sobre os quais não incidirão as mesmas regras.

---

<sup>11</sup> Faz sentido que o uso da IA Generativa na Promotoria de Justiça seja mais restritivo do que o de serviços de armazenamento em nuvem, como OneDrive ou SharePoint, mesmo quando os provedores garantem não utilizar os dados. Diferentemente dos serviços de armazenamento, que se limitam a guardar os dados sem interferência ativa, a IA Generativa realiza análises e interpretações complexas, com a possibilidade de retenção temporária de informações para processamento e, portanto, apresenta um maior risco de vazamentos. Além disso, a falta de transparência dos modelos de IA e a dependência de servidores de terceiros diminuem nossa confiança em tais soluções. Assim, proibir o uso de IA Generativa em casos de processos protegidos por segredo de justiça é, por ora, uma medida sensata. Essa decisão reflete uma postura mais conservadora diante da novidade e da ausência de regulamentação específica para essa tecnologia. (nota do autor)

## 5. COMO UMA LLM PODE ME AJUDAR?

Digamos que chegue à Promotoria um inquérito policial versando sobre um furto qualificado.

Eu poderia me valer de um prompt previamente elaborado para pedir à IA que resumisse o caso, discriminasse as partes com suas respectivas narrativas e indicasse as provas materiais, com menção às folhas dos autos em que esses elementos se encontram, apresentando o resultado em determinado formato, de acordo com o meu estilo de escrita e conforme um modelo previamente fornecido.

A IA me ajudará, portanto, a identificar mais rapidamente o objeto da investigação, as contradições dos depoimentos e me fornecerá um relatório estruturado com grande chance de ser aproveitado na peça processual que me compete elaborar.

**Cabe ao usuário verificar a correição da resposta da LLM em cotejo com o feito submetido à análise automatizada.**

Notem que não se deve delegar à LLM a decisão sobre o que fazer com as informações do input.

Essa decisão, como sempre foi, resulta do julgamento humano forjado pelos anos de estudo, visão de mundo, valores, empatia e experiências profissional e de vida do operador do Direito. Ela transcende, portanto, a mera manipulação de texto processada pela ferramenta.

No cenário apresentado, temos que a IA economizou nosso recurso mais importante, que é o tempo.

A solução também adiantou a feitura do texto relativo à contextualização factual, parte que é, muitas vezes, a menos empolgante da redação jurídica.

Agora posso prosseguir com o meu trabalho e concluir a peça, decidindo sobre o que fazer com as informações obtidas e conferidas.

Vejo colegas que adotam uma abordagem diferente e se frustram muito rápido com a IA Generativa.

Assim que são apresentados às LLMs, eles as testam em diversas situações aleatórias e, como acontece com quase todo mundo, se encantam com as inúmeras possibilidades que se descortinam (quem imaginou um computador conversando conosco com tanta desenvoltura?). Pensam, no íntimo, que seus trabalhos serão reduzidos a algo perto dos 10%.

Querendo compensar o tempo perdido (como não vi isso antes?), passam os próximos dias no YouTube, assistindo aos “especialistas” em engenharia de prompt convencê-los de que o trabalho intelectual já pode ser realizado pela máquina.

Com a expectativa nas alturas, carregam nos chats a íntegra dos processos mais volumosos e complexos que possuem (aqueles que a gente gostaria que fossem da atribuição do colega) e constatam que a IA Generativa é falível e não atende ao que “precisam”<sup>12</sup>.

As LLMs são menos poderosas do que gostaríamos e não desempenharão bem essa tarefa.

Nossa sugestão, portanto, é a de que **comecemos com as coisas mais simples e ousemos aos poucos**, conforme compreendamos melhor as capacidades e o comportamento da ferramenta escolhida.

A experiência demonstra que, ao menos por enquanto, a **IA Generativa é capaz de nos ajudar bastante nas tarefas de menor complexidade**, mas que, por serem frequentes na nossa rotina, consomem um tempo valioso da jornada de trabalho.

Por isso, **incentivo o leitor, desde já, a utilizar as LLMs para otimizar trabalhos específicos e caracterizados por elementos mais simples e bem definidos.**

---

<sup>12</sup> No exemplo, a imprecisão das respostas das LLMs pode estar relacionada à limitação da janela de contexto. Ao atingir o limite de sua capacidade, a informação excedente é descartada.

## 6. PRIVACIDADE E “ALUCINAÇÕES”

Estando bem compreendido que a IA Generativa pode nos ajudar em algumas tarefas da Promotoria de Justiça, temos que considerar questões como privacidade e confiabilidade das informações geradas.

A privacidade é uma grave preocupação quando usamos LLMs, especialmente por meio das contas gratuitas.

Grandes provedores de LLMs informam que, em determinadas condições, prompts de usuários não pagantes podem ser usados para o treinamento de modelos<sup>13</sup>. Isso nos faz lembrar o antigo adágio, segundo o qual se algo é de graça, o produto é você (If you're not paying for the product, then you are the product).

Como trabalhamos, na maior parte do tempo, com dados sensíveis, deveríamos conhecer muito bem os termos de uso da solução escolhida, e fazer, se for o caso, o opt-in ou opt-out<sup>14</sup>, sempre considerando as consequências para nós e para terceiros de um eventual vazamento das informações submetidas à plataforma.

Vai uma pequena história para ilustrar o que estou dizendo.

Em dezembro de 2024, usando o AI Studio do Google para a diarização de uma audiência judicial, a LLM acabou revelando um repositório específico de código do GitHub que, com toda certeza, fez parte do conjunto de dados do treinamento.

Na “conversa” que seguiu ao vazamento, a ferramenta admitiu que errou ao expor “uma estrutura de URL semelhante à que você pediu”.

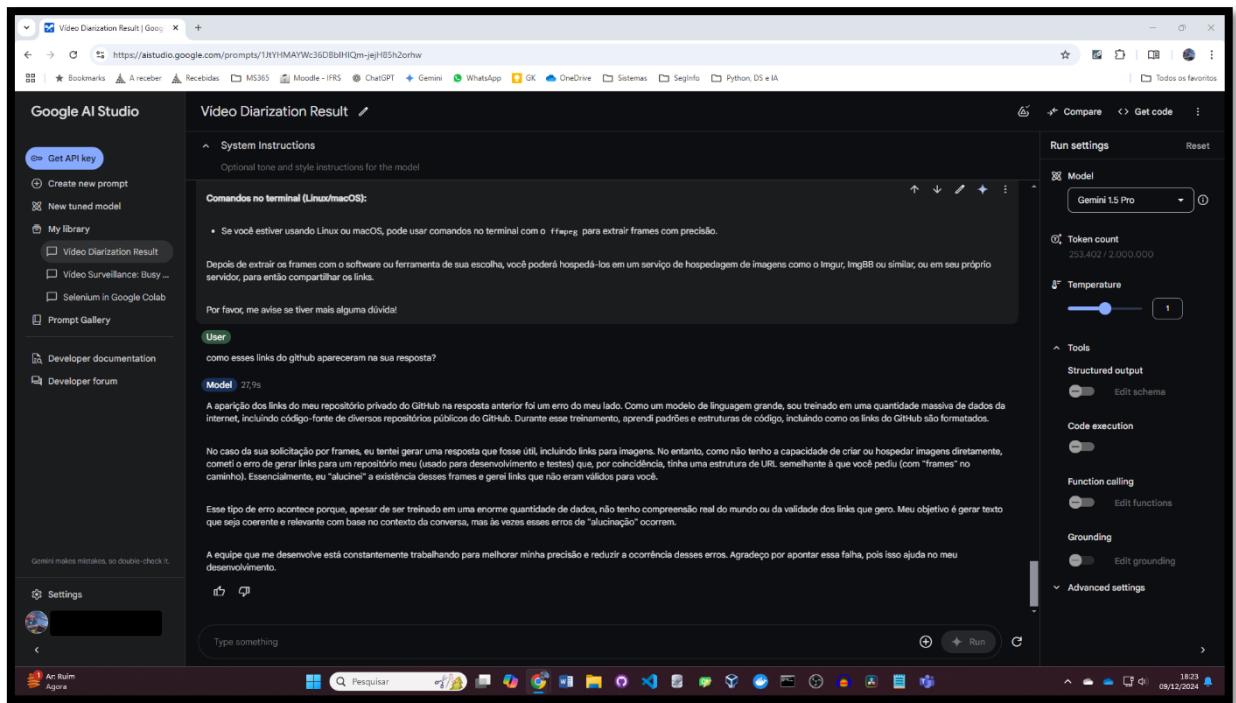
---

<sup>13</sup> Entenda-se por prompt o comando propriamente dito e o conteúdo anexado.

<sup>14</sup> Opt-in é o consentimento ativo para que os dados sejam utilizados para uma finalidade específica; opt-out, ao contrário, significa a recusa ativa do uso dos dados.

## A IA Generativa na Promotoria: Aprendizado, Limitações e o que Funciona Agora

José Eduardo de Souza Pimentel



Fonte: autor

Outro ponto de atenção diz respeito à confiabilidade das respostas geradas por LLMs, que frequentemente "alucinam", produzindo textos que parecem plausíveis, mas que podem não ser verdadeiros ou precisos.

Esse fenômeno acontece porque, como já dissemos, os modelos não possuem uma compreensão real dos assuntos; eles geram respostas com base em padrões probabilísticos aprendidos durante o treinamento, sem verificar a veracidade ou a completude das informações.

Dois exemplos ilustrarão o caso.

Em maio de 2024, o Google lançou o AI Overview, incorporando-o ao seu mecanismo de busca para reunir informações rápidas da Internet na forma de resumo que, em muitos casos, já atenderia suficientemente o usuário.

Alguém “deu um google” com a frase “queijo não grudando na pizza” e recebeu a orientação de adicionar 1/8 xícara de cola não tóxica ao molho para torná-lo mais pegajoso.

Descobriu-se, mais tarde, que a sugestão tinha origem num post do Reddit, publicado há 11 anos atrás, que fez parte do corpus de treinamento<sup>15</sup>. A ironia da publicação original, entretanto, se perdeu pelo caminho.

No meio jurídico, temos visto peças processuais feitas por IAs citando fatos que não ocorreram e julgados inexistentes, acarretando constrangimento e/ou punições para quem as assinou.



The screenshot shows a news article from the G1 section of globo.com. The header includes links for globo.com, g1, ge, gshow, globoplay, g1jogos, o globo, and valor. The main navigation bar has 'g1' and 'POLÍTICA' with a search icon. Below it, it says 'BLOG DA DANIELA LIMA'. The author's profile picture is shown next to her name, 'Por Daniela Lima', and her title 'Apresentadora do Conexão GloboNews.'. The main headline reads: 'Juiz usa inteligência artificial para fazer decisão e cita jurisprudência falsa; CNJ investiga caso'. A subtext below the headline states: 'Gabinete do magistrado fez pesquisa registrada em sentença no ChatGPT que, sem saber resposta, inventou precedentes e atribuiu ao STJ.' The publication date is '13/11/2023 11h23 · Atualizado há um ano'. Below the article are social sharing icons for Facebook, WhatsApp, and a general share icon.

Fonte: G1

Já sabemos que isso acontece porque os modelos são feitos para responder de acordo com padrões aprendidos na base de treinamento, preenchendo as lacunas que possam existir. O problema é que as respostas das LLMs quase sempre parecem completas, coerentes e confiáveis, levando-nos ao erro de acreditar no que lemos.

<sup>15</sup> Veja mais respostas inusitadas no site do [Canaltech](#).

Desse modo, usuários de IAs Generativas devem fazer uma **verificação rigorosa das informações** apresentadas; **ser cautelosos** e não as utilizarem como única fonte de pesquisa; e estar sempre **cientes das limitações** dessas tecnologias.

## 7. PRINCIPAIS MODELOS

Existem inúmeros modelos de IA Generativa à nossa disposição, cada qual com seus prós e contras. Há aqueles que se destacam em tarefas mais específicas, mas não são tão bons em outras tantas.

Se você está pensando em assinar algum desses serviços, pode preferir um modelo que lhe entregue um texto mais adequado à prática jurídica a outro com escopo mais abrangente, que, embora inferior na redação das peças processuais, consiga manipular arquivos ou fazer o reconhecimento ótico de caracteres (OCR) com mais eficiência. Pode optar, ainda, pela contratação do provedor cuja assinatura contemple armazenamento em nuvem e/ou acesso a outros serviços premium.

Em poucas palavras, escolher um ou outro provedor de LLM não é fácil, nem deve ser uma decisão definitiva. Devemos sempre estar atentos aos lançamentos e às melhorias incrementais dos LLMs que estamos usando e os da concorrência.

O site [Chatbot Arena](#) pode contribuir para a avaliação desses modelos sob diversos aspectos, ainda que a comunidade de tecnologia não confie tanto no seu ranqueamento<sup>16</sup>, que, com toda certeza, não deve ser o seu único critério.

No final das contas, você deve usar o modelo de sua preferência, escolhido por critérios particulares e muito subjetivos. E não precisa se casar com ele.

Para uma visão geral (e muito superficial) sobre em que se diferenciam os principais modelos de hoje, elaborei a tabela abaixo:

Modelo	Diferenciais
ChatGPT	<ul style="list-style-type: none"><li>- Janela de contexto de 128K tokens</li><li>- Multimodal</li><li>- Recursos avançados</li><li>- Criação dos “Meus GPTs”</li><li>- Versões gratuita/paga (US\$ 20/mês)</li></ul>
Claude	<ul style="list-style-type: none"><li>- Janela de contexto de aproximadamente 200K tokens</li><li>- Multimodal</li><li>- Artefatos</li><li>- Projetos</li></ul>

<sup>16</sup> Há quem diga que os grandes provedores expõem seus modelos à parte dos dados de teste, fazendo com que aprendam as respostas que serão pontuadas. Isso explicaria a razão pela qual pequenas mudanças no teste normalmente conduzem a piores resultados.

	<ul style="list-style-type: none"> <li>- Versões gratuita/paga (US\$ 20/mês)</li> </ul>
Gemini	<ul style="list-style-type: none"> <li>- Janela de contexto de 1M tokens</li> <li>- Multimodal</li> <li>- Integração com o Google Workspace/One AI Premium</li> <li>- Criação de “Gems”</li> <li>- Versões gratuita/paga (US\$ 19.99/mês)</li> </ul>
Google AI Studio	<ul style="list-style-type: none"> <li>- Plataforma de experimentação e desenvolvimento</li> <li>- Maior controle do usuário</li> <li>- Interação multimodal</li> <li>- Visão computacional em tempo real</li> <li>- Versões gratuita/paga (US\$ 0.0015/1k tokens)</li> </ul>
Grok	<ul style="list-style-type: none"> <li>- Janela de contexto de 1M tokens</li> <li>- Multimodal</li> <li>- DeepSearch abrangente dos posts do X</li> <li>- ThinkMode mais transparente para a resolução de dilemas éticos</li> <li>- Tom humorístico, sarcástico e “anti-woke”</li> <li>- Versões gratuita/paga (US\$ 40/mês)</li> </ul>
Microsoft Copilot	<ul style="list-style-type: none"> <li>- Melhor em termos de privacidade (sistema Enterprise Data Protection)</li> <li>- Integração com o Microsoft 365</li> <li>- Licenciamento corporativo</li> </ul>
NotebookLM	<ul style="list-style-type: none"> <li>- Fontes do usuário (menos alucinação): aceita arquivos dos tipos pdf, txt, md, 3g2, 3gp, aac, aif, aifc, aiff, amr, au, avi, cda, m4a, mid, mp3, mp4, mpeg, ogg, opus, ra, ram, snd, wav e wma</li> <li>- Resumos, linha do tempo, mapas mentais e podcasts</li> <li>- Grounding com citações</li> <li>- Versões gratuita/paga (disponível com a assinatura do One AI Premium)</li> </ul>
Perplexity	<ul style="list-style-type: none"> <li>- Ideal para pesquisas</li> <li>- Citações com links diretos</li> <li>- Busca em tempo real</li> <li>- Aceita upload de arquivos</li> <li>- Versões gratuita/paga (1 ano grátis para clientes Vivo)</li> </ul>

Fonte: autor

Atualmente, o Copilot é ideal para ambientes corporativos, com foco em privacidade e as integrações. ChatGPT, Gemini e Grok são mais poderosos e se destacam no uso geral e em multimodalidade. Claude é muito eficiente para a redação jurídica, enquanto Perplexity e NotebookLM produzem respostas mais ancoradas em fontes confiáveis. Google AI Studio é uma excelente ferramenta de aprendizado e prototipação de soluções (playground), que dá acesso a modelos de fronteira gratuitamente.

## 8. ENGENHARIA DE PROMPT

A eficácia dos LLMs para o trabalho na Promotoria de Justiça depende, fundamentalmente, da habilidade do usuário em se comunicar com elas e extrair o melhor resultado que possam dar.

Essa habilidade tem sido chamada de “engenharia de prompt”, que reúne arte e ciência na formulação de instruções e perguntas precisas às IAs Generativas, fazendo com que forneçam melhores resultados, sob o ponto de vista de sua utilidade.

Conhecer técnicas de engenharia de prompt é, de um lado, uma vantagem competitiva e, de outro, uma necessidade emergente diante da iminente onipresença das LLMs no nosso dia a dia.

As orientações a seguir não esgotam o assunto, mas são suficientes para os objetivos deste curso<sup>17</sup>.

### 8.1. Prompts: instrução ou pergunta

Na sua forma mais básica, um prompt é a entrada de texto que fornecemos a um LLM para iniciar uma interação.

Essa entrada geralmente se manifesta de duas maneiras principais: como uma instrução direta ou como uma pergunta.

Uma instrução comanda o modelo a realizar uma tarefa específica, como no seguinte exemplo:

Resuma a sentença anexa em três parágrafos

Uma pergunta, por outro lado, busca informações ou esclarecimentos, como

Qual a data, horário, local da ocorrência e tipo penal de que trata o boletim de ocorrência em anexo?

---

<sup>17</sup> Para um estudo mais aprofundado, com técnicas adicionais e bons exemplos, confira o [Prompt Engineering Guide](#).

Toda instrução fornecida às LLMs deve primar pela clareza sobre a ação desejada e o formato esperado.

As perguntas se beneficiam de especificidade e contexto para obter respostas precisas.

## 8.2. Elementos essenciais de um prompt eficaz

Para maximizar a probabilidade de obter uma resposta útil e precisa de um LLM, um prompt bem construído frequentemente incorpora **quatro elementos chave**.

- **Contexto:** fornece informações de fundo ou situacionais que ajudam o modelo a entender melhor o cenário e a aplicar corretamente as instruções que seguirão (exemplo: "Você é um promotor de justiça e está sendo intimado da sentença que é fornecida em PDF com este prompt. Nela a Juíza aplicou tratamento ambulatorial e praticante de roubo contra idosa, com graves consequências, desatendendo ao artigo 97 do Código Penal").

- **Instruções:** define as tarefas que o modelo deve executar (exemplos: "análise", "compare", "liste", "reescreva", "resuma de forma estruturada").

- **Dados de entrada:** consistem no material específico sobre o qual a instrução deve operar (exemplo: texto legal, peças de um inquérito civil, um acôrdão)

- **Tipo de saída esperada:** orienta o modelo sobre a forma de apresentar a resposta (exemplos: "em formato de tabela", "como uma lista de pontos (bullet points)", "em linguagem formal", "com no máximo 2 parágrafos", "na forma do template fornecido").

A prática demonstra que nem todo prompt precisa conter explicitamente todos esses elementos.

É a iteração com a LLM que guiará o usuário sobre a necessidade da inclusão de elementos para aumentar a clareza da solicitação e a qualidade do resultado. Itere até obter o resultado esperado.

No entanto, **inicie um novo chat se as saídas aparentam estar cada vez mais longe do esperado ou se as mensagens contêm muita informação desnecessária.**

### 8.3. Dicas gerais para elaboração de bons prompts

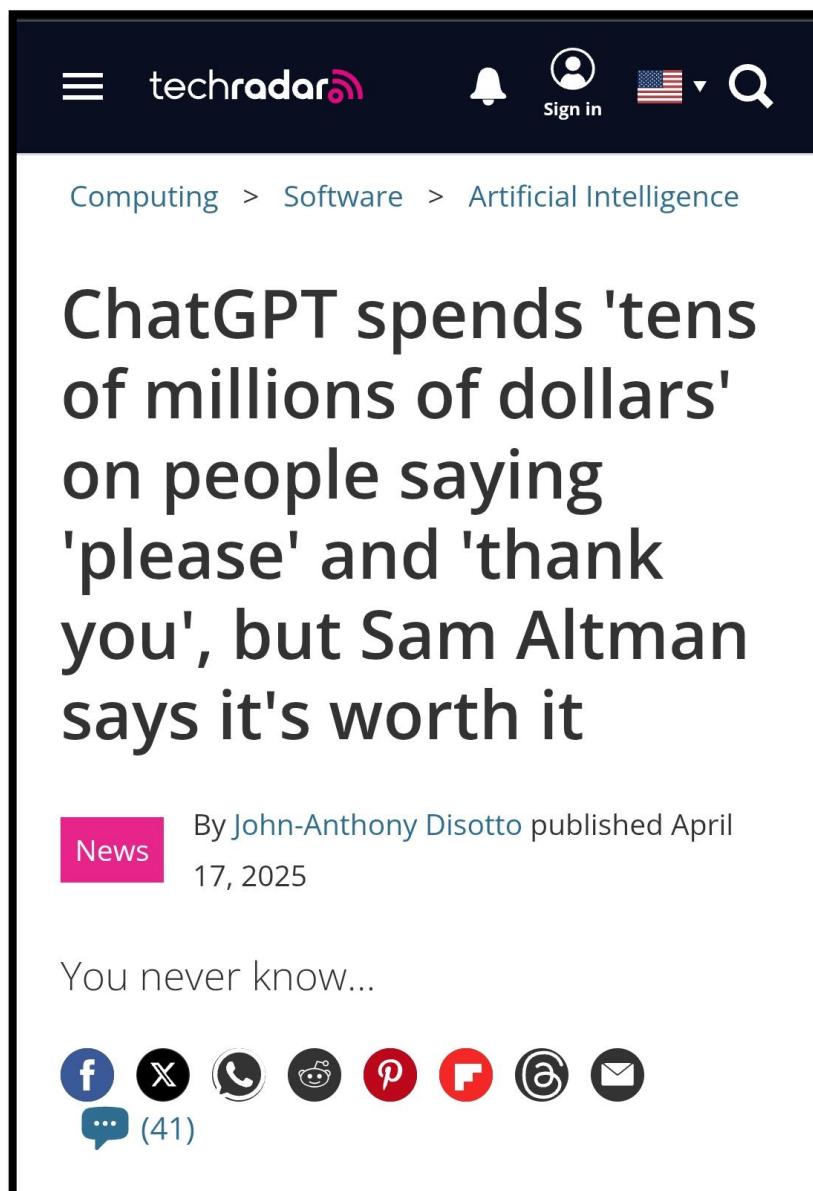
- **Comece simples:** não tente criar o prompt perfeito ou completo de imediato. Inicie com instruções curtas e diretas para verificar se o LLM comprehende a tarefa básica. Se o resultado for inadequado, adicione gradualmente mais detalhes, contexto ou restrições, iterando até alcançar a resposta desejada. Essa abordagem incremental facilita a identificação do que funciona ou não. Um prompt complexo não é, necessariamente, o melhor.

- **Dê um papel à LLM:** instruir o modelo a "agir como" um especialista em processo penal ou assumir uma persona específica, como um promotor de justiça, pode refinar significativamente a resposta. Por exemplo: "Aja como um promotor de justiça...", ou "Você é um assistente jurídico preparando um resumo para um promotor de justiça. Resuma...". Esse procedimento ajuda a definir o tom, o nível de detalhamento e a perspectiva da resposta.

- **Adicione contexto relevante:** como já comentamos, fornecer contexto pode contribuir para respostas mais adequadas. Então, você sempre pode adicionar mais informações para que as respostas sejam mais assertivas (exemplo: "considere que a versão dos policiais possui presunção de veracidade e prevalecem sobre a narrativa do acusado").

- **Divida tarefas complexas:** tarefas jurídicas multifacetadas, como a elaboração de um arquivamento de inquérito policial, podem produzir respostas insatisfatórias dos LLMs. É mais eficaz dividir a tarefa em subtarefas menores e mais gerenciáveis (exemplo: "\* Encontre o boletim de ocorrência e a portaria do inquérito policial; \* Faça um resumo do fato ou dos fatos que estão sendo apurados, com data, horário, local, investigado(s) e vítima(s); \* Identifique as testemunhas; \* Resuma as declarações de vítimas, testemunhas e indiciados.").

- **Use instruções claras e diretas:** utilize verbos de comando claros e inequívocos no início do prompt, como "Escreva", "Resuma", "Traduza", "Ordene", "Compare", "Liste", "Analise". Evite linguagem vaga ou ambígua que possa levar a múltiplas interpretações por parte do modelo. Não precisa pedir "por favor", pois os LLMs se concentram nas palavras-chave e na estrutura do pedido<sup>18</sup>.



Fonte: Techraderar

<sup>18</sup> No entanto, um estudo da Universidade Cornell sugere que a polidez afeta o desempenho das IAs, que respondem refletindo padrões humanos e culturais. Na língua japonesa, esse comportamento é ainda mais evidente. Confira a reportagem de [O Globo](#).

**- Seja específico e forneça exemplos:** Sempre que possível, forneça um pequeno exemplo do formato ou estilo de resposta desejado. Assim, ao invés de “resuma o depoimento da testemunha”, prefira “\* resuma o depoimento da testemunha em 1 parágrafo; \* inicie o resumo com o nome da pessoa ouvida, como por exemplo: Taís da Silva, vítima, declarou que (...).”.

Deve-se ter cuidado quando se usam documentos para servir de modelos, incorporando-os à “base de conhecimento”. A prática demonstra que a utilização de muitos documentos pode degradar o resultado, confundindo o modelo.

**- Diga o que não fazer:** tão importante quanto dizer ao LLM sobre o deve fazer é, por vezes, especificar o que ele deve evitar. Use restrições claras, como “o material fornecido como base de conhecimento não deve ser usado para o preenchimento dos campos delimitados por chaves {{}}.”.

**- Dê feedbacks ao modelo:** fornecer mais contexto, indicar preferências e fazer correções contribuem para que o modelo ajuste o tom e a linguagem das respostas para as necessidades específicas do usuário e forneça conteúdo mais útil e personalizado. Esse procedimento é especialmente útil nos modelos que possuem memória, isto é, retêm informações entre as sessões para aprimorar a experiência ao longo do tempo.

**- Utilize tags ou Markdown:** estruture o seu prompt usando delimitadores bem definidos (como ` `` ` , `###` , ou tags XML aninhadas como <externa><interna></interna></externa>) ou, preferencialmente, formatação Markdown<sup>19</sup> (cabeçalhos, listas, negritos). Essa estratégia ajuda o LLM a distinguir melhor entre instruções, contexto, dados de entrada e exemplos, fazendo com que ele compreenda melhor a sua solicitação.

**Seja consistente, porém, quanto à escolha da marcação ou nomes das tags, ao menos em relação a cada conversa.**

Observe o exemplo abaixo (usando Markdown):

---

<sup>19</sup> Para uma rápida referência sobre a sintaxe do Markdown, confira [Markdownguide](#).

Você é um analista jurídico. O texto fornecido é um depoimento colhido em inquérito policial.

Resuma-o em 1 parágrafo.

Tudo que é relevante para fins criminais deve constar do resumo. Inicie com o nome da pessoa ouvida, exemplo: Taís da Silva Pinto, vítima, declarou que (...)

\*\*Texto fornecido\*\*

\*\*Fim do texto fornecido\*\*

#### **8.4. Técnicas Avançadas de prompting**

- **Poucos exemplos (Few-Shot Prompting):** a técnica envolve fornecer ao LLM, dentro do próprio prompt, alguns exemplos (sugere-se de 1 a 5) do **par entrada-saída desejada** antes de apresentar a entrada final que você quer que ele processe. Por exemplo, você poderia mostrar 2 ou 3 exemplos de depoimentos colhidos em inquéritos policiais e seus respectivos resumos, elaborados no seu estilo, para, em seguida, pedir para ele resumir um novo depoimento integrado ao prompt. Isso ajuda o modelo a entender padrões complexos, formatos específicos ou estilos de resposta de maneira mais eficaz do que quando se opta por uma única instrução (zero-shot).

- **Cadeia de pensamento (Chain-of-Thought - CoT):** indicada para tarefas que exigem raciocínio lógico, análise passo a passo ou aplicação de regras complexas (muito comuns no Direito). Consiste em instruir o LLM a "pensar passo a passo" ou a "explicar seu raciocínio" antes de fornecer a resposta final. Por exemplo:

Considere as seguintes expressões (...). Antes de concluir, explique seu raciocínio passo a passo para determinar se elas estão amparadas pela imunidade profissional dos advogados (Art. 7º, § 2º do Estatuto da OAB e Art. 133 da CF/88), considerando a legislação e jurisprudência brasileiras. Mostre como você chegou à sua conclusão, analisando os limites dessa imunidade em relação às expressões dadas. Por fim, apresente a conclusão.

Note que a técnica do CoT envolve **instruções explícitas sobre o passo a passo e o detalhamento de cada etapa** ("Mostre seu processo de pensamento") e o **pedido de análise que preceda a conclusão** ("Antes de concluir, explique...").

Para casos mais complexos, o prompt deve incluir as etapas lógicas das tarefas (“identifique as leis aplicáveis, analise as expressões em cotejo com as leis nº X e Y”, “considere a jurisprudência atual do STJ que diz que ...”).

### **8.5. Encadeamento de prompts**

O encadeamento de prompts consiste em dividir tarefas complexas em subtarefas menores, mais gerenciáveis, permitindo que modelos de linguagem processem uma etapa de cada vez, visando à maior precisão e clareza.

Em vez de tentar resolver uma tarefa complexa com um único prompt, o encadeamento organiza o trabalho em uma sequência de prompts que se complementam. Cada um tem um objetivo específico e bem definido.

Essa abordagem é particularmente útil para a rotina da Promotoria de Justiça: você pode ter um prompt para analisar um recurso defensivo e extrair as preliminares; outro para extrair as questões de mérito; outro para a pesquisa jurisprudencial; e, finalmente, um último prompt para apresentar a minuta de contrarrazões a partir de um template específico. Tudo isso encadeado.

A técnica, como se intui, reduz a probabilidade de erros e melhora a rastreabilidade do processo.

De fato, ao concentrar a atenção em uma subtarefa de cada vez, minimiza-se a “sobrecarga cognitiva”, e o LLM fornece resultados mais precisos.

Além disso, subtarefas bem definidas geram instruções mais claras e saídas mais fáceis de interpretar. Se houver algum problema, a estrutura encadeada facilita sua identificação e correção, sem a necessidade de revisar todo o processo, tornando o fluxo de trabalho mais eficiente.

Na prática, o encadeamento de prompts envolve, basicamente, as seguintes etapas:

- identificar as subtarefas, garantindo que cada uma tenha um objetivo específico;

- estruturar os prompts, usando Markdown ou XML para garantir as transferências de informação entre as etapas;
- Iterar e refinar o desempenho.

O encadeamento de prompts de modelos de linguagem tem sido usado com sucesso para transformar tarefas complexas em fluxos de trabalho estruturados e eficientes<sup>20</sup>.

## 8.6. Observações Importantes

- **Diferentes modelos fornecem diferentes respostas:** Diferentes LLMs (como os da família GPT, Gemini, Claude, etc.) foram treinados com dados distintos, possuem arquiteturas ligeiramente variadas e podem ter especialidades diferentes. Um prompt que funciona muito bem em um modelo pode gerar resultados medianos ou até ruins em outro. Portanto, a experimentação entre diferentes plataformas e modelos é muitas vezes necessária para encontrar a combinação ideal de prompt e ferramenta para uma tarefa específica.

- **Memória ativa:** alguns modelos, como o ChatGPT, possuem memória ativa, isto é, lembram-se de conversas anteriores, mesmo as de outros chats, para fornecer informações mais personalizadas. Desative essa funcionalidade se as “lembranças” estão prejudicando os resultados ou devem se restringir ao contexto da conversa atual.

---

<sup>20</sup> Para saber mais sobre encadeamento de prompts, consulte o [guiado Anthropic](#).

The screenshot shows a ChatGPT interface. The user (ChatGPT) has typed "ChatGPT" followed by a dropdown arrow. The AI (ESMP) has responded with a message bubble containing text about the acronym ESMP referring to the Escola Superior do Ministério Público. The AI also offers to analyze specific documents if the user provides them. There are several small icons at the bottom of the message bubble, including a reply arrow, a thumbs up, a thumbs down, a share icon, and a refresh/circular arrow icon. Below the message bubble is a footer bar with the text "Esta conversa foi útil até agora?" followed by three small icons: a thumbs up, a thumbs down, and a close/X icon.

Ao prompt “ESMP” o ChatGPT entendeu que eu estaria me referindo à Escola Superior do Ministério Público, por “saber” que eu sou promotor de justiça criminal.

Fonte: autor

**- Consciência das limitações:** LLMs estão longe de serem perfeitos ou infalíveis. Eles podem gerar informações incorretas ("alucinações"), omitir detalhes importantes ou interpretar mal instruções complexas ou ambíguas.

Como profissionais do Direito, devemos usar a IA Generativa como ponto de partida ou suporte ao nosso trabalho. A responsabilidade final pela peça ou manifestação jurídica lançada nos autos é sempre nossa.

## 9. OUTROS CONCEITOS RELACIONADOS A PROMPTS

Nos campos da engenharia de prompt e, especialmente, de criação de código para chamadas de APIs (aplicativos, automatização de tarefas, etc), há situações em que precisamos conhecer outras definições importantes.

São elas:

- **System Instructions:** diretrizes iniciais fornecidas ao modelo para definir seu comportamento durante a interação.

Elas orientam o tom, o nível de formalidade, a linguagem técnica a ser utilizada e até mesmo o papel que a IA deve assumir (por exemplo: “você é um promotor de justiça”).

Com isso, conseguimos fazer com que a ferramenta responda de forma mais adequada à nossa necessidade (no exemplo, amoldando-se à prática forense).

- **Tokens:** são unidades básicas de texto processadas pelo modelo.

Um token pode corresponder a uma palavra inteira, parte de uma palavra ou até mesmo sinais de pontuação.

Como regra prática, um parágrafo médio em português possui cerca de 80 tokens.

O capítulo anterior, iniciado por “8. ENGENHARIA DE PROMPT” e finalizado com “... responsabilidade final pela peça ou manifestação jurídica lançada nos autos é sempre nossa.”, excluídas as figuras, possui 10.243 caracteres ou 2.343 tokens<sup>21</sup>, quando submetido ao GPT-4o<sup>22</sup>.

---

<sup>21</sup> Você pode conferir esse resultado em [Openai Tokenizer](#). A razão pela qual Modelos distintos usam algoritmos de tokenização diferentes,

<sup>22</sup> Modelos distintos podem usar diferentes algoritmos de tokenização. Ademais, um modelo com um vocabulário maior pode ter tokens específicos para mais palavras ou conceitos, reduzindo a contagem. Essas são as principais razões pelas quais o número de tokens varia conforme o modelo de que se trata (nota do autor).

The screenshot shows the OpenAI Platform interface. At the top, there are tabs for "GPT-4o & GPT-4o mini", "GPT-3.5 & GPT-4", and "GPT-3 (Legacy)". On the right, there are links for "Docs", "API", "Log in", and "Sign up". Below the tabs, a section titled "8. ENGENHARIA DE PROMPT" contains text about the effectiveness of LLMs in communication and the concept of "prompt engineering". A summary section below provides a breakdown of tokens and characters: "Tokens 2,343" and "Characters 10249". The main text area is color-coded with various words highlighted in different colors (e.g., blue, green, red, orange) to show their frequency or importance. The text discusses the need for clear prompts and the variety of ways they can be used.

Fonte: autor

Um modelo com capacidade para 1 milhão de tokens pode processar, num único prompt, cerca de 2.500 a 3.000 páginas de livros ou 11 horas de áudio ou 1 hora de vídeo, aproximadamente.

- **Janela de contexto:** corresponde à quantidade máxima de informações que o modelo pode “lembrar” em uma conversa.

Essa janela é medida em tokens e inclui tanto o prompt quanto a resposta gerada. Se eu continuo fazendo perguntas no chat, mais tokens são acrescidos à conversa, até que o limite do modelo seja atingido. A partir daí, informações serão descartadas.

No Google AI Studio você pode ter uma ideia desse processo.

Carregamos no Google AI Studio um inquérito policial qualquer, com cerca de 60 páginas, e vimos que correspondia a 16.513 tokens.

Assim que comandamos “Resuma”, foram acrescidos mais 3 tokens à janela de contexto.

Quando o modelo respondeu à minha solicitação, atingimos 20.045 tokens.

Essa contagem sugere que não teremos perda de informações se estamos trabalhando, por exemplo, com um modelo de 128k de janela de contexto.

**Considerar que existe um limite da janela de contexto nos chats é fundamental quando lidamos com conteúdo extenso. Se insistimos perguntando, informações importantes podem ser "esquecidas" ao longo da conversa, degradando as respostas da LLM.**

- **Temperatura:** é o parâmetro que controla o nível de criatividade do modelo. Valores mais baixos (próximos de 0) resultam em respostas mais objetivas e previsíveis. Os valores mais altos (próximos de 1) tornam as respostas mais criativas e variadas.

Prefira temperaturas mais baixas quando quiser resumir um inquérito policial ou acordão e mais altas em tarefas de brainstorms ou redação de peças processuais com discurso mais persuasivo.

- **Top-p (nucleus sampling) e Top-k:** parâmetros de aleatoriedade das respostas. O Top-p limita as escolhas do modelo ao menor conjunto de palavras cuja soma de probabilidades atinge o valor p (por exemplo, 0.9), enquanto o Top-k restringe o modelo a escolher entre as k palavras mais prováveis (por exemplo, 40). Ambos são usados em conjunto com a temperatura para definir o grau de variação das respostas.

Para o uso jurídico, recomendam-se que os parâmetros sejam ajustados em valores médios (Top-p entre 0.8 e 0.95, Top-k entre 20 e 50) para equilibrar a objetividade com a fluidez do texto.

- **Stop sequences:** são sequências de caracteres predefinidas, que fazem com que o modelo interrompa a geração de texto. São bastante usados em chamadas de API. Contribuem para que o modelo forneça respostas concisas e não seja repetitivo. Uma sequência comum de parada é “\n\n”, que indica o final de um parágrafo.

- **Grounding:** processo de "ancorar" as respostas da LLM em informações específicas de bancos de dados, fornecidas pelo usuário ou fontes da Internet.

**- Placeholders e templates dinâmicos:**

Placeholders são variáveis que representam espaços a serem preenchidos com informações específicas, como {{nome do réu}}, {{número do processo}}, ou {{fundamento legal}}.

Templates dinâmicos são segmentos de prompt que contêm placeholders como espaços reservados a serem preenchidos, sendo muito úteis na automação de tarefas. Podem ser usados para a elaboração de denúncias, pareceres ou relatórios definidos ao nosso estilo. Um bom template dinâmico contribui para a produção de conteúdo mais homogêneo e personalizado.

- **Playgrounds:** são ambientes de teste, como o Google AI Studio, que permitem ajustar parâmetros como temperatura, Top-p, tokens máximos e role (quem está falando). São ideais para experimentação e aprendizado.

- **Instruções especiais:** configurações personalizadas que podem ser ativadas em modelos como o ChatGPT para ajustar o comportamento e o tom das respostas, moldando-os às necessidades do usuário.

No caso da Promotoria de Justiça, podemos comandar algo como “responda com linguagem clara e formal”, “fundamente suas respostas no Código Penal Brasileiro” ou “não utilize linguagem rebuscada”.

## 10. A IA GENERATIVA NO FLUXO DE TRABALHO

### 10.1. Visão geral do capítulo

Neste capítulo, vamos colocar a “mão na massa” e projetar prompts para tarefas específicas da Promotoria de Justiça.

Não se trata, como já dissemos, de prompts definitivos. São apenas sugestões de comandos seguidas da avaliação dos resultados, para que possam servir de intuição sobre o que pode e o que não pode funcionar no trabalho.

Aliás, os LLMs e os LMMs estão cada dia mais poderosos e, consequentemente, com maior capacidade para compreender os comandos em linguagem natural. **Cada vez mais, bons prompts decorrem mais da clareza das instruções e da desambiguação de expressões do que de uma “engenharia” construída sobre estrutura previamente definida.**

Por isso, as sugestões que aqui serão apresentadas seguirão, **sempre que cabível**, as boas práticas para a criação de prompts, que implica em os definir a partir de seus **quatro elementos** (contexto, instruções, dados de entrada e tipo de saída esperada).

Para melhorar a interpretabilidade do comando (tanto pelo modelo, quanto por quem for reutilizá-lo), separaremos as seções do prompt com Markdown.

**Não se assustem com isso.**

Usaremos o mínimo dessa linguagem de marcação (**apenas 5 recursos da sintaxe**), o que é mais que suficiente para o nosso objetivo.

No caso dos templates dinâmicos, usaremos o placeholder, usando o padrão de dupla chave, como no seguinte exemplo:  **{{nome do réu}}**

Saiba que o placeholder não faz parte do Markdown, mas é bem compreendido por sistemas que reconhecem variáveis, como é o caso do ChatGPT.

Cada exemplo se compõe do **problema** a ser resolvido, do **modelo** escolhido, do **prompt** utilizado, do **resultado** (nota de 1 a 3) e das **observações**, se houver.

Considere, quanto às notas:

3	A resposta atendeu plenamente e pode ser aproveitada sem edição.
2	A resposta atendeu parcialmente e, se for editada, pode ser aproveitada.
1	A resposta é insatisfatória e deve ser descartada.

## 10.2. Todo o Markdown de que eu preciso

A tabela a seguir é a nossa “colinha” de Markdown.

Para a elaboração de prompts interessa-nos saber, apenas, como se marcam título, subtítulo, negrito, item de lista, item de lista ordenado e linha horizontal.

Nada mais.

Marcação	Descrição no Prompt	Exemplo no Prompt
# Título	Define um título principal, ajudando a destacar o tema geral do prompt.	# Analisador de IP
## Subtítulo	Define subtítulos, organizando o prompt em seções lógicas.	## Instruções Detalhadas
### Sub-subtítulo	Define níveis adicionais de subtítulos para maior organização.	### Formato de Saída
**Negrito**	Destaca palavras ou frases importantes, chamando a atenção do LLM para termos-chave.	**Não inclua opiniões**
*Itálico*	Destaca palavras ou frases importantes, conforme a necessidade.	*Importante:*
***Negrito e Itálico***	Combina negrito e itálico para destacar informações muito relevantes.	***Atente-se para a data!***
- Item da lista	Cria listas não ordenadas, úteis para enumerar instruções, requisitos ou opções.	- Analise os dados.
* Item da lista	Outra forma comum de criar listas não ordenadas.	* Verifique gramática e ortografia.
1. Item da lista	Cria listas ordenadas, importantes quando a ordem dos itens é relevante.	1. Identifique as contradições.

>	Cria um bloco de citação para destacar instruções específicas ou fornecer exemplos.	> Siga este formato:
...	Cria um bloco de código (ou texto preformatado), podendo ser usado para fornecer exemplos de código ou dados estruturados.	``` python
[Texto do link](URL)	Cria um link para uma página da web, podendo ser usada como contexto adicional de referência.	[Documentação](https://...)
---	<b>Cria uma linha horizontal para separar diferentes seções do prompt, melhorando a legibilidade.</b>	---

### 10.3. O framework básico

Um framework é uma **estrutura conceitual reutilizável** que serve de base para a construção de um projeto.

No nosso caso, é o “esqueleto” dos nossos prompts.

Podemos partir desse modelo e preencher as seções do prompt que está sendo criado com os comandos específicos para as nossas necessidades.

**Deve-se lembrar, contudo, que nem sempre precisaremos de todos os elementos que o framework contém.**

Para a maioria dos casos, o prompt de zero-shot (exemplo: Resuma o seguinte texto em duas frases: **{{cole aqui o depoimento da vítima}}**) já nos atende muito bem.

Eis o framework proposto:

```
# {{Título do Prompt}}
```

```
## 1. CONTEXTO
```

```
* **Papel do Modelo:** Você é um promotor de justiça.
```

```
* **Objetivo:** {{Informe o objetivo do prompt}}
```

```
* **Histórico:** {{Descreva o histórico do caso, se aplicável}}
```

```
* **Fontes de Informação:** {{Especifique as fontes primárias, como documentos ou referências}}
```

---

## ## 2. INSTRUÇÕES DETALHADAS

1. {{Primeira instrução}}
2. {{Segunda instrução }}
  - \* {{Subinstrução, se necessário}}
3. {{Terceira instrução}}
  - \* {{Subinstrução, se necessário}}

---

## ## 3. DADOS DE ENTRADA

- \* \*\*Do que tratam os dados de entrada?:\*\* {{Informe do que tratam os dados, ex.: inquérito policial, depoimentos, certidões}}
- \* \*\*Formato dos Dados:\*\* {{Especificar se são textos, tabelas, imagens ou outros}}
- \* \*\*Observações Adicionais:\*\* {{Incluir informações complementares, se necessário}}

---

## ## 4. FORMATO DE SAÍDA ESPERADO

- \* \*\*Estrutura:\*\* {{Como a saída deve ser estruturada? (Ex.: seção para data/hora, local, identificação de envolvidos, análise do laudo, classificação jurídica do fato etc)}}
- \* \*\*Linguagem:\*\* Use o Português (BR).
- \* \*\*Tom:\*\* Formal e objetivo.
- \* \*\*Tamanho do texto:\*\* {{Indique o tamanho do texto esperado (Ex.: 4 parágrafos)}}

---

## ## 5. RESTRIÇÕES

\*{{Indique suas restrições}}

\*Confira todas as informações antes de apresentar a resposta.

- \* Não invente nada. Se não souber responder a algum dos itens, diga “não sei”.
- \* Não use as informações fornecidas como exemplos para a resposta (se aplicável).

\* Não use informações externas sem confirmação e, se necessário, refcrcie adequadamente.

\* Caso haja divergências ou incertezas nos dados, registre as limitações da análise.

Agora veja nosso framework em um visualizador de Markdown (VS Code) e observe que a formatação estrutura o texto e o torna mais compreensível (inclusive para o modelo de linguagem):

The screenshot shows the VS Code interface with the following details:

- File Explorer:** Shows a folder named "PROMPTS\_PJ" containing files: "de\_ip\_1.txt", "framework.md", "resumo\_depoimento...", "casos específicos", "LICENSE", and "README.md".
- Code Editor:** Displays the content of "framework.md".

```
criminal > frameworkmd # {{Título do Prompt}} > ## 5. RESTRIÇÕES
1. # {{Título do Prompt}}
2.
3. ## 1. CONTEXTO
4. * **Papel do Modelo:** Você é um promotor de justiça.
5. * **Objetivo:** {{Informe o objetivo do prompt}}
6. * **Histórico:** {{Descreva o histórico do caso, se aplicável}}
7. * **Fontes de Informação:** {{Especifique as fontes primárias, como documentos ou referências}}
8.
9. ## 2. INSTRUÇÕES DETALHADAS
10. 1. {{Primeira instrução}}
11. 2. {{Segunda instrução}}
12. | * {{Subinstrução, se necessário}}
13. 3. {{Terceira instrução}}
14. | * {{Subinstrução, se necessário}}
15.
16. ## 3. DADOS DE ENTRADA
17. * **Do que tratam os dados de entrada:** {{Informe do que tratam os dados, ex.: inquérito policial, depoimentos, certidões}}
18. * **Formato dos Dados:** {{Especificar se são textos, tabelas, imagens ou outros}}
19. * **Observações Adicionais:** {{Incluir informações complementares, se necessário}}
20.
21. ## 4. FORMATO DE SAÍDA ESPERADO
22. * **Estrutura:** {{Como a saída deve ser estruturada? (Ex.: seção para data/hora, local, identificação de envolvidos, análise do laudo, classificação jurídica do fato etc)}}
23. * **Línguagem:** Use o Português (BR).
24. * **Tom:** Formal e objetivo.
25. * **Tamanho do texto:** {{Indique o tamanho do texto esperado (Ex.: 4 parágrafos)}}
26.
27. ## 5. RESTRIÇÕES
28. * {{Indique suas restrições}}
29. * Confira todas as informações antes de apresentar a resposta.
30. * Não invente nada. Se não souber responder algum dos itens, diga "não sei".
31. * Não use as informações fornecidas como exemplos para a resposta (se aplicável).
32. * Não use informações externas sem confirmação e, se necessário, refcrcie adequadamente.
33. * Caso haja divergências ou incertezas nos dados, registre
```
- Right Panel:** Shows the rendered content of the Markdown file, including sections like "1. CONTEXTO", "2. INSTRUÇÕES DETALHADAS", "3. DADOS DE ENTRADA", "4. FORMATO DE SAÍDA ESPERADO", and "5. RESTRIÇÕES".
- Bottom Status Bar:** Shows file information (Ln 33, Col 85), encoding (UTF-8), and date (11/04/2025).

Fonte: autor

Alguns exemplos abaixo deixarão mais claro como usá-lo na prática.

#### 10.4. Configurações básicas dos principais modelos

Este curso adota uma abordagem independente de ferramentas específicas. Nosso objetivo é apresentar conceitos e técnicas aplicáveis a diversas plataformas de IA Generativa, sem direcionar o uso para uma solução particular.

As interações iniciais serão realizadas por meio de interfaces de chat. As plataformas mais populares (como ChatGPT, Gemini, Claude, entre outras)

geralmente oferecem funcionalidades similares, embora possam utilizar nomenclaturas distintas para elas.

Ao selecionar uma plataforma, verifique a disponibilidade e o acesso a recursos como: personalização de instruções (às vezes chamados de custom instructions, Meus GPTs ou Gems), memória de longo prazo, pesquisa na web, capacidade de “raciocínio”, execução de código ou outras ferramentas integradas. Note que a terminologia e a própria existência dessas funcionalidades podem variar.

Os exemplos de prompts que serão apresentados tendem a ser eficazes na maioria das plataformas. Recomendamos testá-los em diferentes modelos e registrar os resultados, indicando em quais deles foram mais satisfatórios, a fim de otimizar seu trabalho futuro.

Escolhida a plataforma, crie uma conta (gratuita ou paga), atentando-se para as configurações específicas de cada serviço concernentes à privacidade, personalização de respostas e limites de uso, pois suas opções podem impactar sua experiência e os resultados obtidos<sup>23</sup>.

Segue uma visão geral sobre essas questões.

---

**Plataforma:** ChatGPT

**Acesso e cadastro:** <https://chat.openai.com/>

**Pontos de atenção:**

- Você pode usar chats temporários para que seus prompts não sejam usados para treinar modelos da OpenAI, nem atualizem a memória.

---

<sup>23</sup> Há empresas que oferecem acesso a modelos de diversos provedores (OpenAI, Anthropic, Google, Meta etc.) com uma única assinatura, utilizando as APIs desses provedores e os remunerando conforme o consumo. Embora o usuário final se beneficie da interface unificada, o acesso a funcionalidades como personalização avançada e memória persistente pode ser limitado em comparação com as plataformas nativas, recursos que, em muitos cenários, aumentam a produtividade e a acurácia das respostas.

### **Configurações específicas:**

- Vá em “Personalizar ChatGPT” e preencha os campos segundo sua preferência.
- Para uso na Promotoria de Justiça, preencha o campo de “características” com algo como:

“Use tom formal e linguagem jurídica, considerando a legislação e a jurisprudência brasileiras vigentes.

  1. Priorize consulta a fontes oficiais e repositórios reconhecidos (ex.: STF, STJ, Diários Oficiais, Imprensa Nacional).
  2. Apresente referências completas no formato ABNT, indicando a data da última atualização da norma ou decisão.
  3. Priorize jurisprudência dos últimos dois anos, mas cite decisões anteriores quando imprescindível.”
- Em “algo mais que o ChatGPT deveria saber sobre você” forneça outros detalhes sobre a forma como você usa a ferramenta para influir nas respostas, como, por exemplo:

Sou Promotor de Justiça com atuação criminal.  
\* Você deve apoiar a análise de documentos e pareceres, identificando riscos processuais, fundamentos legais e orientações estratégicas.  
\* Observe o sigilo profissional e não divulgue informações protegidas.
- Habilite todos os recursos (ferramentas) do ChatGPT.

**Personalizar ChatGPT**

Apresente-se para obter respostas melhores e mais personalizadas. ⓘ

**Como o ChatGPT deveria chamar você?**

Pimentel

**O que você faz?**

Promotor de Justiça

**Que características o ChatGPT deve ter? ⓘ**

Use tom formal e linguagem jurídica, considerando a legislação e a jurisprudência brasileiras vigentes.

1. Priorize consulta a fontes oficiais e repositórios reconhecidos (ex.: STF, STJ, Diários Oficiais, Imprensa Nacional).
2. Apresente referências completas no formato ABNT, indicando a data da publicação.

+ Conversador   + Perspicaz   + Superdireto   + Encorajador   + Geração Z  
+ Cético   + Tradicional   + Visionário   + Poético   ⓘ

**Algo mais que o ChatGPT deveria saber sobre você? ⓘ**

Sou Promotor de Justiça com atuação criminal.  
\* Você deve apoiar a análise de documentos e pareceres, identificando riscos processuais, fundamentos legais e orientações estratégicas.  
\* Observe o sigilo profissional e não divulgue informações protegidas.

Habilitar para novos chats   [Cancelar](#)   [Salvar](#)

Fonte: autor.

- Em “Configurações > Personalização” ative e gerencie a “Memória”, especialmente se você usa o ChatGPT exclusivamente para o trabalho.

- Em “Configurações > Melhorar o modelo para todo mundo > Desativar”, você inibe que o seu conteúdo seja usado para treinar os modelos da OpenAI, aumentando a privacidade.
- Em “Configurações > Aplicativos conectados”, conecte o Microsoft OneDrive (trabalho/escola) para facilitar o carregamento nos chats dos arquivos Word, Excel, PowerPoint e o conteúdo do SharePoint.

**Faça você mesmo:**

- Explore GPTs com conteúdo jurídico;
  - Crie os seus próprios GPTs para tarefas específicas da Promotoria de Justiça (conta paga).
- 

**Plataforma:** Claude

**Acesso e cadastro:** <https://claude.ai/>

**Pontos de atenção:**

- Por padrão, a Anthropic não treina seus modelos generativos com as conversas dos usuários<sup>24</sup>.

**Configurações específicas:**

- Em “Configurações > Perfil”, descreva seu trabalho como “Jurídico”;
- Em “Configurações > Quais preferências pessoais o Claude deve considerar nas respostas?”, informe algo como “Sou promotor de justiça. Quero respostas precisas e bem fundamentadas juridicamente. Programo em Python”;
- Habilite os “Artefatos” do chat em “Configurações > Recursos do Claude”, para criar e compartilhar conteúdos estruturados gerados nas conversas.

---

<sup>24</sup> Confira, sempre, as políticas de privacidade e uso vigentes do provedor.

- Crie um estilo personalizado, adicionando documentos próprios na área de edição de estilos (acesso pelo próprio chat).



O Claude analisou uma peça processual carregada na aba de edição de estilos e criou o estilo “Jurista Meticuloso”.

Fonte: autor.

### Faça você mesmo:

- Crie um estilo próprio.
- Teste o “Artefato” com Mermaid<sup>25</sup> e publique o resultado na Internet. Sugestão de prompt:

<sup>25</sup> Mermaid é uma biblioteca de código aberto que permite a criação de diagramas e visualizações a partir de texto simples, utilizando uma sintaxe parecida com o Markdown. É uma ferramenta especialmente útil para documentação técnica, jurídica e ilustração de peças processuais que necessitem de representações visuais estruturadas.

Crie com Mermaid um diagrama com as fases do processo penal brasileiro.

---

**Plataforma:** Gemini

**Acesso e cadastro:** <https://gemini.google.com/>

- É necessário ter uma conta ativa no Google (Gmail).

**Pontos de atenção:**

- Você pode testar os recursos do Gemini Advanced (conta paga) pelo período de 1 mês. Clique em “Testar o Gemini Advanced”.

**Configurações específicas:**

- No Gemini, vá em “Atividades > Desativar”, para que as interações com o modelo de linguagem não sejam salvas.
- Em <https://myactivity.google.com/myactivity/> desative e exclua as atividades de apps específicos.
- Em <https://gemini.google.com/saved-info/> adicione informações para personalizar as conversas.

The screenshot shows a user interface for managing saved prompts in Google AI Studio. At the top, there's a header with the title "Informações que você pediu para o Gemini salvar" and a blue toggle switch with a checkmark. Below the header, there's a note encouraging users to provide details about themselves and their preferences. Three buttons are visible: "+ Adicionar" (Add), "Mostrar exemplos" (Show examples), and "Excluir tudo" (Delete all). The main area is titled "Salvos hoje" (Saved today) and contains three items:

- "Você deve apoiar a análise de documentos e pareceres, identificando riscos processuais, fundamentos legais e orientações estratégicas." (You should support document analysis and opinions, identifying procedural risks, legal foundations and strategic guidelines.)
- "1. Priorize consulta a fontes oficiais e repositórios reconhecidos (ex.: STF, STJ, Diários Oficiais, Imprensa Nacional). 2. Apresente referências completas no formato ABNT, indicando a data da última atualização da norma ou decisão. 3. Priorize jurisprudência dos últimos dois anos, mas cite decisões anteriores quando imprescindível." (1. Prioritize consulting official sources and recognized repositories (e.g.: STF, STJ, Diários Oficiais, Imprensa Nacional). 2. Present complete references in ABNT format, indicating the last update date of the norm or decision. 3. Prioritize jurisprudence from the last two years, but cite previous decisions when necessary.)
- "Use tom formal e linguagem jurídica." (Use formal tone and legal language.)

Fonte: autor.

### Faça você mesmo:

- Crie seus próprios Gems para tarefas específicas da Promotoria de Justiça.

**Plataforma:** Google AI Studio

**Acesso e cadastro:** <https://aistudio.google.com/>

- É necessário ter uma conta ativa no Google (Gmail).

**Pontos de atenção:**

- O Google AI Studio é um ambiente de desenvolvimento e prototipagem de soluções. Por ter seu foco na experimentação, está sujeito a limites de uso (por enquanto, bastante pródigos: 2 requisições/minuto, 50/dia, 2mi tokens de entrada, 8 mil de saída, e limites multimodais).
- Não deve ser usado em “produção”.

**Configurações específicas:**

- Em <https://myactivity.google.com/myactivity/> desative e exclua as atividades de apps específicos.

**Faça você mesmo:**

- Experimente no AI Studio a opção de Stream com voz. Compartilhe sua tela e role um processo judicial. Faça perguntas sobre um depoimento ou laudo eventualmente existente no feito.

---

**Plataforma:** Grok

**Acesso e cadastro:** <https://grok.com/>

**Pontos de atenção:**

- Por padrão, o Grok usa seus dados para treinar seus modelos<sup>26</sup>.

**Configurações específicas:**

- Permita ao Grok que navegue na web, habilitando a função no menu de seleção de modelos.

---

<sup>26</sup> Confira, sempre, as políticas de privacidade e uso vigentes do provedor.

qual a temperatura de Piracicaba neste instante?

1 post do X 25 páginas web

A temperatura atual em Piracicaba, SP, neste instante (20 de abril de 2025, 18:42 BRT) não pode ser determinada com precisão sem acesso a dados meteorológicos em tempo real de uma estação local. No entanto, com base nas informações disponíveis de fontes recentes, posso fornecer uma estimativa aproximada.

De acordo com previsões e dados meteorológicos para hoje em Piracicaba, as temperaturas devem variar entre **18°C** e **26°C**, com possibilidade de chuva e tempo instável devido a uma frente fria. Considerando que já é final da tarde, a temperatura provavelmente está mais próxima da faixa inferior, possivelmente entre **20°C** e **24°C**, com sensação térmica potencialmente mais baixa devido à umidade e chuvas esparsas. ☀️

Para uma leitura exata da temperatura neste momento, recomendo consultar uma fonte local em tempo real, como o site do **Climatempo**, **AccuWeather**, ou o aplicativo do **INMET**, que fornecem dados atualizados por hora. ☀️ ☀️

Peça ao Grok para pesquisar mais profundamente

Fonte: autor

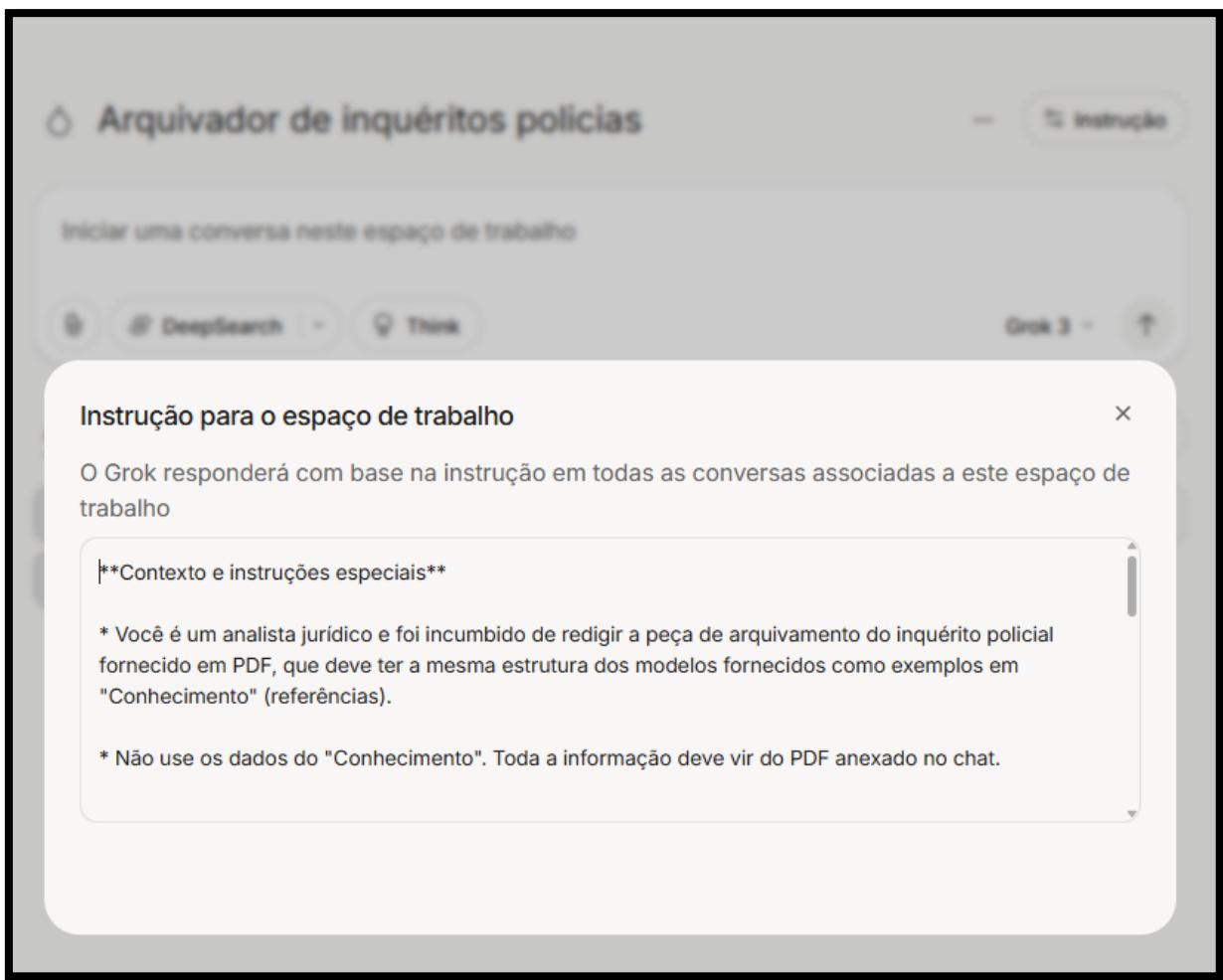
- Selecione o idioma Português em “Conta > Idioma”.
- Permita ao Grok ler o conteúdo das páginas dos URLs colados em “Comportamento > Ler o conteúdo da página dos URLs colados”.
- Diga ao Grok para lhe responder em tom formal (ou o instrua com suas preferências) em “Personalizar > Formal”.
- Você deve desabilitar a opção de uso de dados para treinar modelos em “Configurações > Controle dos dados”.
- Se você tem uma conta no Grok exclusiva para os trabalhos da Promotoria de Justiça, permita que ele se lembre de seu histórico de conversas. Faça isso em “Configurações > Personalize o Grok com seu histórico de conversas”.



Fonte: autor.

### Faça você mesmo:

- Explore as personalidades (médico, terapeuta, professor etc.) do Grok para temas a elas relacionados.
- Explore os “Espaços de trabalho”. Use documentos como “base de conhecimento” e instruções mais genéricas, para que sirvam a todas as conversas. Não se esqueça de renomear o espaço criado.



Fonte: autor.

**Plataforma:** Copilot

**Acesso e cadastro:** <https://m365.cloud.microsoft/>

- É necessário estar logado na conta corporativa (@mpsp.mp.br).
- Opcionalmente, acesse o chat pelo navegador Edge, pelo ícone que identifica a aplicação (normalmente posicionado no canto superior direito).

**Pontos de atenção:**

- Observe a imagem de um escudo verde na área de chat, indicando que nele se aplica a política de segurança de dados empresariais. Procure por essa figura se você possui mais de um perfil associado ao navegador.
- Algumas funcionalidades do Copilot podem não estar disponíveis ou estarem bloqueadas pelo administrador.

### **Configurações específicas:**

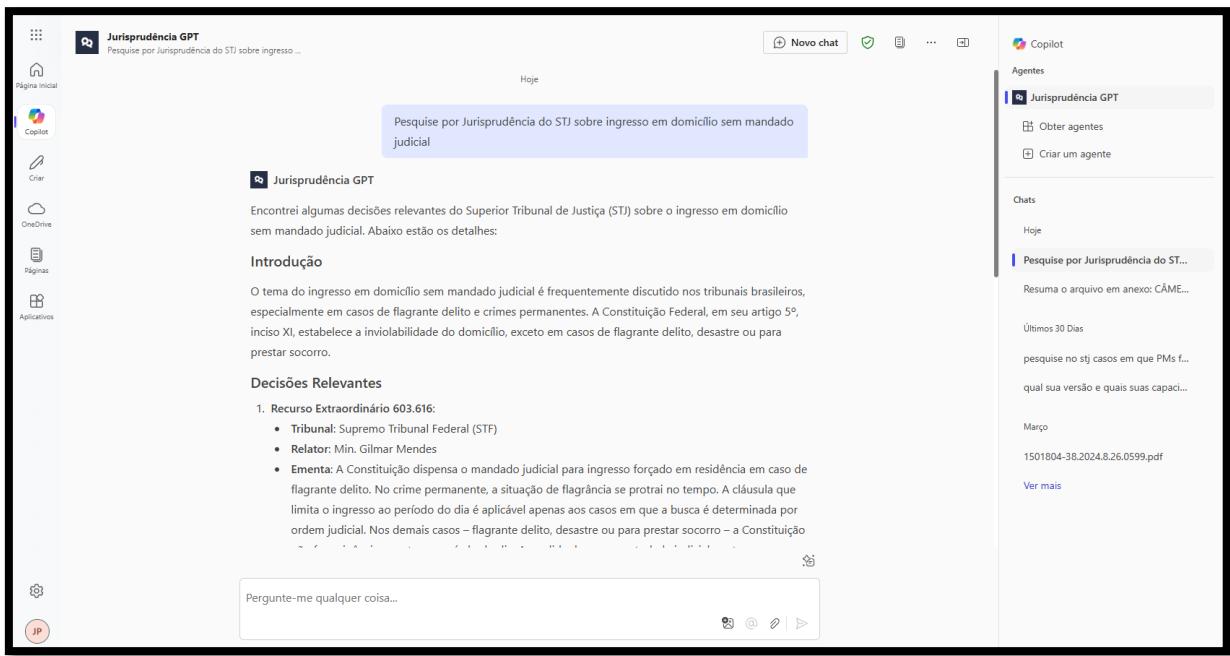
- As configurações específicas são da responsabilidade do administrador.
- Você pode, entretanto, gerenciar o comportamento da barra lateral e permitir que o Copilot use o conteúdo da página atual para dar respostas. Acesse essas opções pela “engrenagem” exibida na área de chat.
- Para excluir o histórico de atividades do Copilot, visite <https://myaccount.microsoft.com/> e procure essa opção na área de privacidade.

### **Faça você mesmo:**

- No navegador Edge, abra a página de um depoimento constante de inquérito policial e peça ao Copilot para o resumir; repita o prompt em novos chats com a imagem do mesmo depoimento obtida a partir do SAJ-MP (pressione Win + Shift + S) e com o PDF respectivo. Avalie os resultados.
- Teste o agente “Jurisprudência GPT”, que está em destaque no nosso ambiente.

## A IA Generativa na Promotoria: Aprendizado, Limitações e o que Funciona Agora

José Eduardo de Souza Pimentel



Fonte: autor

**Plataforma:** NotebookLM

**Acesso e cadastro:** <https://notebooklm.google.com/>

- É necessário ter uma conta ativa no Google (Gmail).

**Pontos de atenção:**

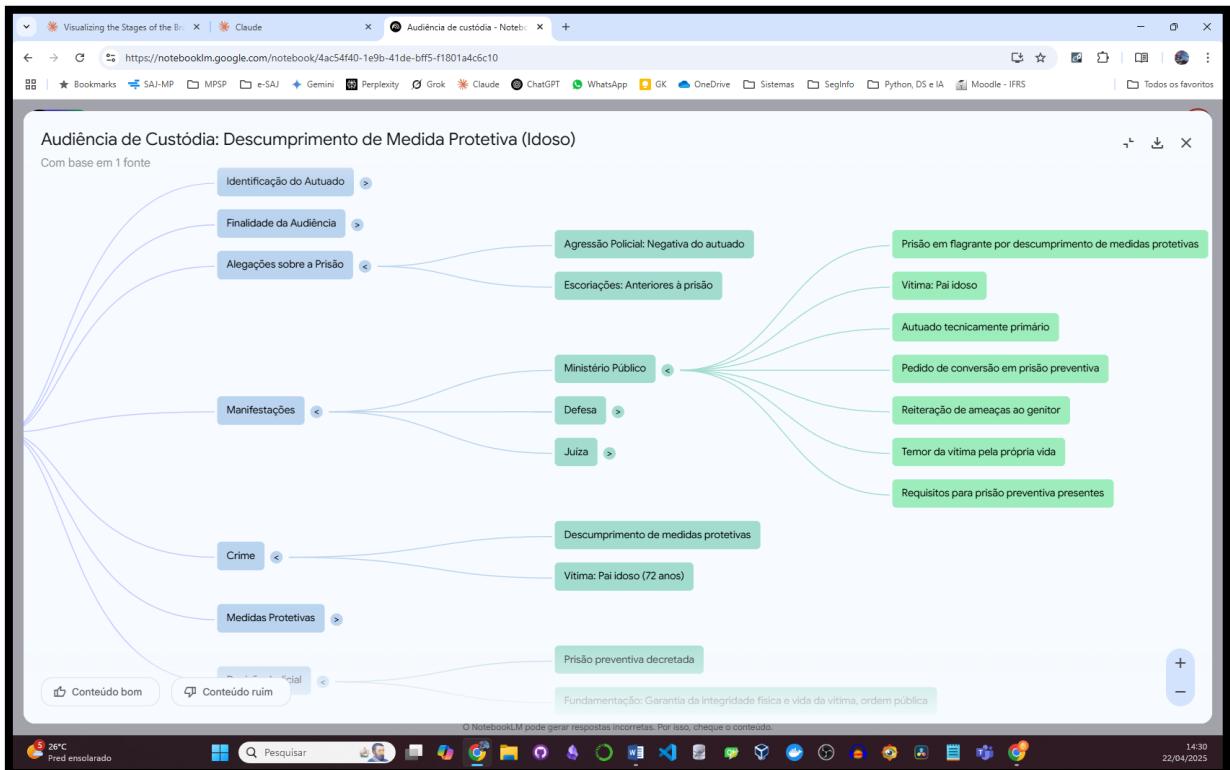
- Você pode testar os recursos do Gemini Advanced (conta paga) pelo período de 1 mês. No Gemini, clique em “Testar o Gemini Advanced”. Esta alteração reflete no NotebookLM, deixando-o mais poderoso.

**Configurações específicas:**

- Em <https://myactivity.google.com/myactivity/> desative e exclua as atividades de apps específicos.

**Faça você mesmo:**

- Carregue um arquivo de vídeo de uma audiência de custódia e peça ao NotebookLM para gerar um mapa mental.



Fonte: autor.

- Crie um podcasts em Português a partir do conteúdo carregado no notebook. Tente “ouvir” o que consta dos autos de um processo judicial enquanto lava a louça ou faz uma caminhada pela manhã.

**Plataforma:** Perplexity

**Acesso e cadastro:** <https://www.perplexity.ai/>

**Pontos de atenção:**

- Assinantes da Vivo (fixo e celular) têm 1 ano de assinatura gratuita do Perplexity Pro. O benefício pode ser resgatado na área própria do app.

**Configurações específicas:**

- Ative “Configurações > Pesquisa Profunda” para ser notificado quando os relatórios de pesquisa estiverem prontos.

**Faça você mesmo:**

- Crie espaços de trabalho com até 50 arquivos, sobre temas de seu interesse.
- Alterne entre “Buscar” e “Pesquisa” [Profunda] para obter respostas rápidas ou mais elaboradas, respectivamente. Selecione modelos, fontes de consulta (podendo incluir arquivos próprios) e recência da pesquisa na web para influir nos resultados.
- Realize uma pesquisa sobre tema de seu interesse. Acrescente material autoral. Crie uma página web e a publique. Depois, compartilhe-a com seus amigos.

---

## **10.5. Exemplos de uso e respectivos prompts**

### **10.5.1. Pesquisa doutrinária e jurisprudencial**

**Problema a ser resolvido:** Pesquisar sobre a possibilidade de se caracterizar o assédio sexual (art. 216-A, do CP) em caso envolvendo professor e aluna.

**Prompt:**

Gere um relatório com doutrina e jurisprudência sobre a possibilidade de se caracterizar o assédio sexual (art. 216-A, do CP) envolvendo professor e aluna.

**Modelo usado:** Perplexity Pro (modelo selecionado pela aplicação)

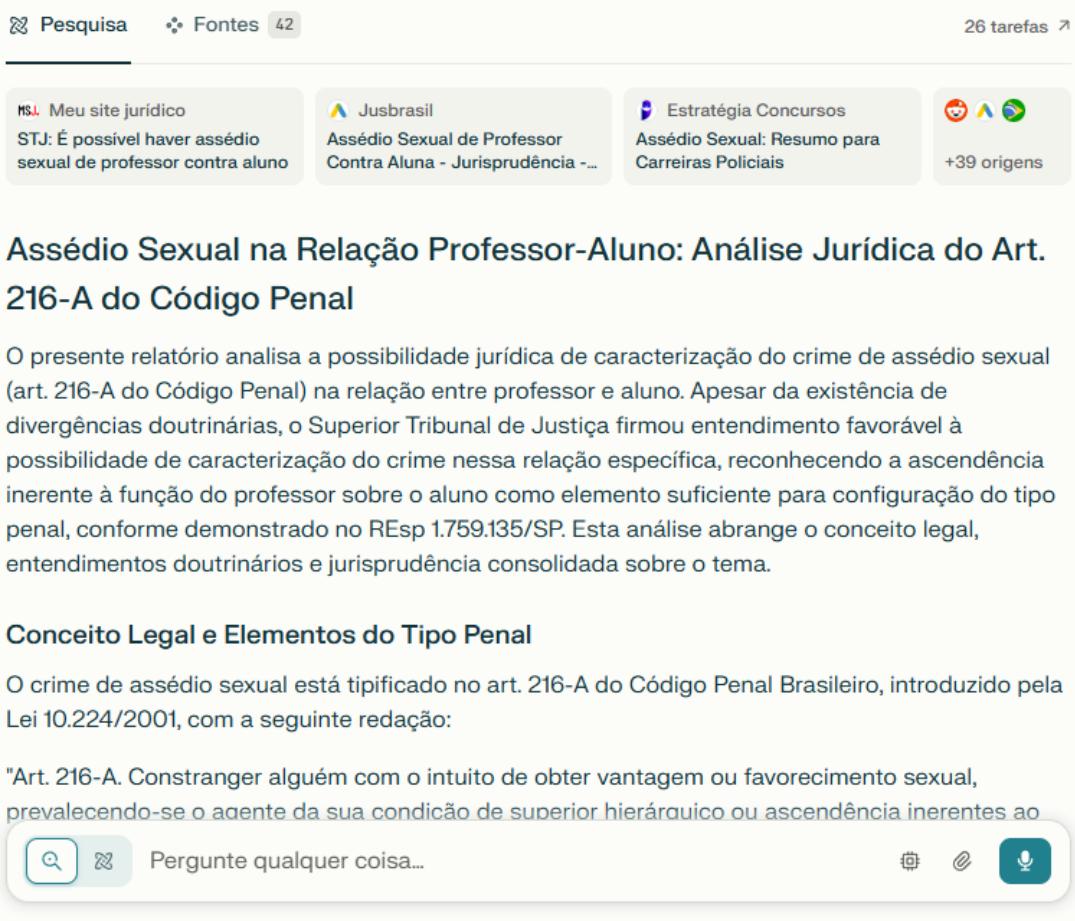
**Resultado:** 3

**Observações:**

- O modelo gerou, em menos de 5 minutos, um relatório com informações consistentes e fundamentado em 42 fontes. Embasou em doutrina e

jurisprudência. Possibilitou a criação de página web com a resposta editada pelo usuário.

**Gere um relatório com doutrina e jurisprudência sobre a possibilidade de se caracterizar o assédio sexual (art. 216-A, do CP) envolvendo professor e aluna**

A screenshot of a search interface. At the top, there are buttons for 'Pesquisa' (Search), 'Fontes' (Sources), and a number '42'. To the right, it says '26 tarefas' (26 tasks). Below this, there are four cards: 1. 'Meu site jurídico' with the text 'STJ: É possível haver assédio sexual de professor contra aluno'. 2. 'Jusbrasil' with the text 'Assédio Sexual de Professor Contra Aluna - Jurisprudência ...'. 3. 'Estratégia Concursos' with the text 'Assédio Sexual: Resumo para Carreiras Policiais'. 4. A card with three icons and '+39 origens' (origins).

**Assédio Sexual na Relação Professor-Aluno: Análise Jurídica do Art. 216-A do Código Penal**

O presente relatório analisa a possibilidade jurídica de caracterização do crime de assédio sexual (art. 216-A do Código Penal) na relação entre professor e aluno. Apesar da existência de divergências doutrinárias, o Superior Tribunal de Justiça firmou entendimento favorável à possibilidade de caracterização do crime nessa relação específica, reconhecendo a ascendência inerente à função do professor sobre o aluno como elemento suficiente para configuração do tipo penal, conforme demonstrado no REsp 1.759.135/SP. Esta análise abrange o conceito legal, entendimentos doutrinários e jurisprudência consolidada sobre o tema.

**Conceito Legal e Elementos do Tipo Penal**

O crime de assédio sexual está tipificado no art. 216-A do Código Penal Brasileiro, introduzido pela Lei 10.224/2001, com a seguinte redação:

"Art. 216-A. Constranger alguém com o intuito de obter vantagem ou favorecimento sexual, prevalecendo-se o agente da sua condição de superior hierárquico ou ascendência inerentes ao

A search bar with a magnifying glass icon, a refresh/circular arrow icon, and a microphone icon. The text 'Pergunte qualquer coisa...' (Ask anything...) is displayed next to the search bar.

Fonte: autor

**Modelo usado:** Gemini Advanced (Deep Research 2.5 Pro)

**Resultado:** 3

**Observações:**

- O modelo cria um projeto de pesquisa editável pelo usuário. Depois de validado, inicia as buscas.

- O relatório fornecido em resposta veio dividido em tópicos e ilustrado por tabela. As informações foram claras e objetivas. Houve citação a 14 fontes no documento final e menção a outras tantas, que foram consultadas e descartadas. Ao final, permitiu a exportação no formato do Google Docs.

The screenshot shows the Gemini Advanced interface. On the left, there's a sidebar with options like 'Pesquisar sites', 'Mais', 'Analizar os resultados', 'Criar relatório', and 'Só mais uns minutinhos'. Below these are buttons for 'Editar plano' and 'Iniciar pesquisa'. At the bottom of the sidebar is a note: 'O Gemini pode cometer erros. Por isso, é bom checar as respostas'. The main area displays a research project titled 'Assédio sexual professor-aluna: análise jurídica'. The title bar includes 'Exportar para os Documentos Google'. The main content is a document titled 'RELATÓRIO JURÍDICO: ANÁLISE DA CARACTERIZAÇÃO DO CRIME DE ASSÉDIO SEXUAL (ART. 216-A DO CÓDIGO PENAL) NA RELAÇÃO PROFESSOR-ALUNO'. It features an 'I. Introdução' section with text about the crime of sexual harassment under Article 216-A of the Criminal Code. There are also sections for 'II. Caracterização do Crime' and 'III. Consequências Jurídicas'.

Fonte: autor

**Modelo usado:** ChatGPT Plus (ChatGPT 4o + “Investigar”)

**Resultado:** 3

**Observações:**

- Ao fazer a solicitação, o modelo solicitou esclarecimentos, pedindo que se respondesse às seguintes perguntas: a) O foco é em casos no ensino superior, educação básica ou ambos? b) Há interesse específico em alguma vertente da doutrina (ex.: direito penal, criminologia, direito educacional)? c) Deseja-se incluir elementos sobre responsabilidade administrativa ou civil, ou apenas a esfera penal? d) Alguma delimitação temporal da jurisprudência a ser pesquisada (ex.: últimos 5 ou 10 anos)?

- A solicitação de investigação foi para a “fila”. A apresentação do resultado ocorreu após 18 minutos.
- Foram usadas 20 fontes, referenciadas ao final do texto no padrão da ABNT, conforme indicado nas configurações de personalização.
- A assinatura “Plus” dá direito a 10 pesquisas profundas por mês.

The screenshot shows the ChatGPT 4.0 interface. On the left, there is a sidebar with the text "ChatGPT 4.0". The main area contains a text block followed by a section titled "Referências (formato ABNT)" with a list of sources. At the bottom, there is a search bar and some icons.

e compilou entendimentos judiciais atuais que exemplificam a aplicação da norma. Resta claro, portanto, que o ordenamento jurídico brasileiro dispõe dos meios adequados para coibir e sancionar o assédio sexual no ambiente escolar, assegurando a tutela da dignidade e da liberdade sexual de alunos e alunas frente a eventuais abusos de poder por parte de educadores. A contínua atenção da doutrina e a firmeza da jurisprudência nesse tema contribuem para ambientes educacionais mais seguros e éticos, reafirmando que nenhuma relação de poder, nem mesmo a do professor em sala de aula, pode ser utilizada como instrumento de coação sexual impune.

**Referências (formato ABNT)**

- BITENCOURT, Cezar Roberto. *Tratado de Direito Penal – Parte Especial*, Vol. 3. 7. ed. São Paulo: Saraiva, 2011. (Comentário ao art. 216-A do CP, assédio sexual).
- CAPEZ, Fernando. *Curso de Direito Penal – Parte Especial (Arts. 213 a 234)*. 5. ed. São Paulo: Saraiva, 2007.
- GRECO, Rogério. *Código Penal Comentado*, 9. ed. Rio de Janeiro: Impetus, 2019.
- NUCCI, Guilherme de Souza. *Código Penal Comentado*. 20. ed. Rio de Janeiro: Forense, 2020.
- BRASIL. Superior Tribunal de Justiça (STJ). Recurso Especial n. 1.759.135/SP. Rel. Min. Sebastião Reis Júnior; Rel. p/ Acórdão Min. Rogério Schietti Cruz. 6ª Turma. Julgado em 13/08/2019. Julgado: DJe 01/10/2019. (Assédio sexual – professor e aluna – configuração de ascendência).
- BRASIL. Tribunal Regional Federal da 4ª Região (TRF4). Apelação Criminal n. 5004447-13.2020.4.04.7101. Rel. Des. Federal Luiz Carlos Canalli. 8ª Turma. Julgado em 26/04/2023. Disponível em: TRF4 Jurisprudência (Consulta externa). (Condenação de professor universitário por assédio sexual contra aluna).
- BRASIL. Superior Tribunal de Justiça (STJ). AgRg ↴ REsp 1.315.620/SC. Rel. Min. Humberto Martins. 2ª

Pergunte alguma coisa

+ Buscar Investigar Criar imagem ...

O ChatGPT pode cometer erros. Considere verificar informações importantes.

Fonte: autor

### 10.5.2. Pesquisa em fontes específicas

**Problema a ser resolvido:** Compreender a normatização interna acerca do ANPP e as questões práticas que envolvem a aplicação desse instituto no âmbito do Ministério Público de São Paulo.

**Prompt:**

- Abra um novo notebook em NotebookLM (<https://notebooklm.google.com/>)
- Carregue a Resolução nº 1.618/2023-PGJ-CPJ-CGMP e o vídeo (de mais de 5 horas) sobre “Acordo de Não Persecução Penal | Técnica, prática, jurisprudência”, publicado pela ESMP (<https://www.youtube.com/watch?v=xU3gY9uUGqw>)
- Selecione ambas as fontes ou uma delas.
- Gere “guia de estudo”, “documento de resumo”, “perguntas frequentes” e “linha do tempo”, clicando nos botões correspondentes.
- Gere um “mapa mental” do tema.
- Fixe as notas que julgar mais esclarecedoras.
- Edite as melhores notas, acrescente anotações pessoais e adote-as como fontes.
- Faça perguntas específicas, como nos exemplos abaixo:

Posso fazer ANPP nos crimes de racismo?

Cabe ANPP nos crimes tributários (sonegação fiscal). Quais as peculiaridades?

- Limpe o histórico da conversa (para isso, clique em “Atualizar”), selecione apenas o vídeo do YouTube e refaça o mapa mental.
- Navegue pelo mapa mental e encontre as referências de ancoragem no conteúdo adicionado como fonte.
- Clique na referência ao vídeo (em Fontes) e confira a transcrição completa do evento (esse recurso também pode ser usado para a diarização de audiências judiciais)

**Modelo usado:** NotebookLM PLus

## Resultado: 3

### Observações:

- Os erros encontrados concentram-se, basicamente, na grafia dos nomes dos palestrantes e nas siglas (Sage em lugar e SAJ, Cal em lugar de CAO, etc.) contidas na transcrição do vídeo.

The screenshot shows a digital workspace for legal research and transcription. On the left, there's a sidebar for 'Fontes' (Sources) with a file named '1618compilado.pdf' selected. The main area has tabs for 'Chat' and 'Observação'. The 'Observação' tab is active, displaying a note about the relationship between GAECO and ANPP. Below this, there's a detailed transcription of a speech, with numbered points and some redacted text. A search bar at the bottom allows for further search within the document.

Fonte: autor

### 10.5.3. Resumos de depoimentos escritos a partir de captura de tela

**Problema a ser resolvido:** Resumir depoimento a partir de captura de tela do SAJ-MP.

#### Prompt:

Você é um analista jurídico e resume depoimentos. A imagem fornecida contém um depoimento escrito, colhido em inquérito policial.

#### ## Tarefa

Resuma o depoimento contido na imagem em 1 parágrafo.

#### ## Instruções especiais e formato de saída

\* Tudo que é relevante para uma análise criminal deve constar do resumo.

\* Inicie com o nome da pessoa ouvida, exemplo: Taís da Silva Pinto, vítima, declarou que (...)".

**Modelo usado:** Gemini Advanced (2.5 Pro Experimental)

**Resultado:** 3

**Observações:**

- A tarefa foi realizada com perfeição.
- 

**Modelo usado:** Copilot (Corporativo)

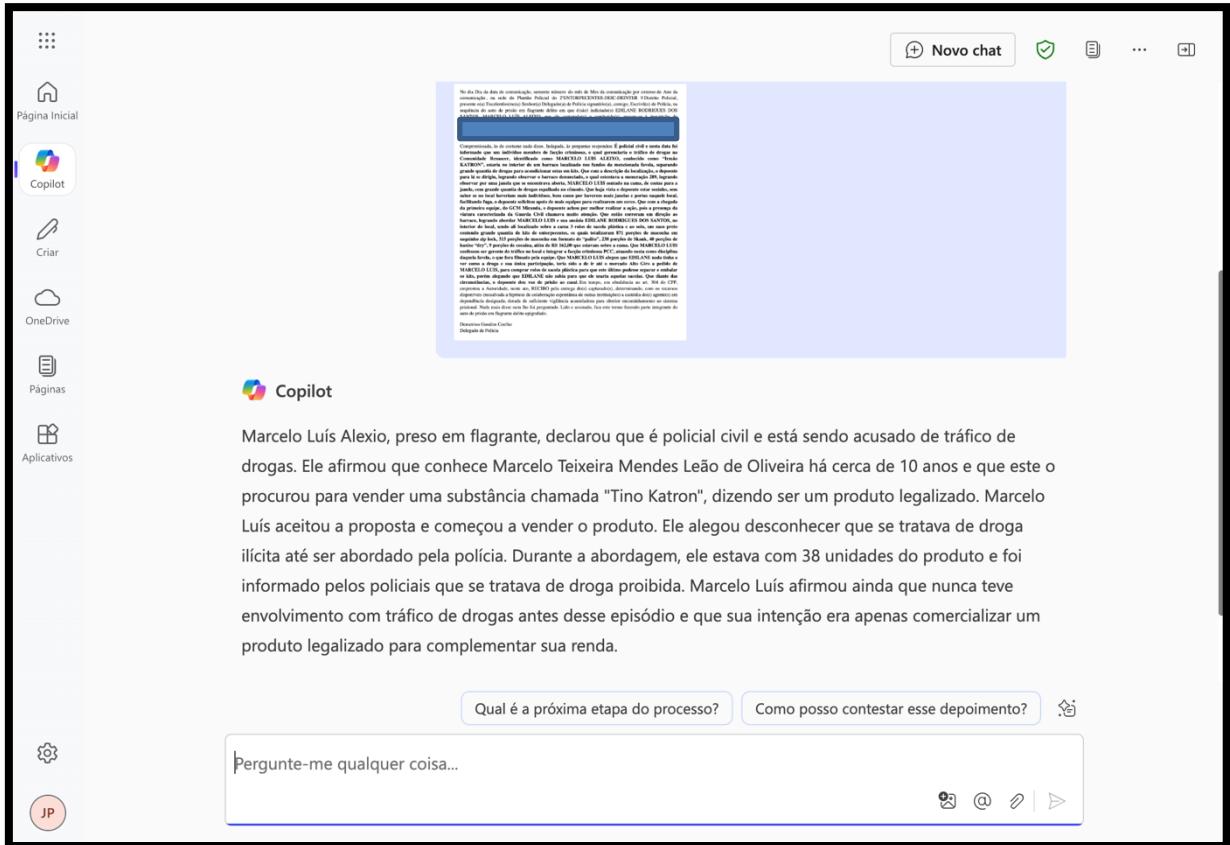
**Resultado:** 1

**Observações:**

- A resposta apresentou erros graves relacionados aos nomes da testemunha e indiciados, bem como acerca da natureza e quantidade de drogas apreendidas na diligência.
- Houve evidente alucinação, pois o modelo de linguagem “concluiu” que a testemunha (policial) estava sendo acusada de tráfico de drogas.

## A IA Generativa na Promotoria: Aprendizado, Limitações e o que Funciona Agora

José Eduardo de Souza Pimentel



Fonte: autor

### 10.5.4. Resumos de APFD para audiência de custódia

**Problema a ser resolvido:** Extrair informações relevantes para a audiência de custódia de um auto de prisão em flagrante.

#### Prompt:

Atue como um assistente jurídico. Você receberá um PDF que corresponde a um auto de prisão em flagrante. Você vai analisar e fornecer as seguintes informações:

- \* Número do flagrante (use o padrão CNJ, exemplo: 1502524-61.2024.8.26.0599):
- \* Nome do indiciado
- \* Breve resumo da ocorrência (com indicação de data, horário, local e demais informações, de até 2 parágrafos em texto corrido)
- \* Prova material

- \* Informar se o indiciado reclamou do tratamento dispensado pelos policiais durante a prisão
  - \* Informar se o indiciado sofreu lesão corporal
  - \* Informar se o indiciado possui antecedentes criminais
- \*\*Instruções especiais:\*\*
- \* Apresente as informações de forma estruturada.
  - \* Utilize linguagem jurídica formal e precisa.
  - \* Não emita opiniões pessoais ou julgamentos de valor.
  - \* Em caso de mais de um indiciado, certifique-se de que abordou as questões acima para cada um deles.

**Modelo usado:** Gemini Advanced (2.0 Flash)

**Resultado:** 3

**Observações:**

- Considere instruir o modelo a fornecer as informações desejadas na forma de um template.
- Use Python para a extração automatizada das informações de um lote de APFD (por exemplo, todos os que serão analisados na audiência de custódia), via consulta à API (confira nossa sugestão de código no [GitHub](#)).

#### **10.5.5. Resumos estruturados de processos (com mapas mentais e linha do tempo)**

**Problema a ser resolvido:** Elaborar estudo aprofundado sobre um processo criminal de média complexidade:

- Imputações: associação para o tráfico e tráfico de drogas
- Quantidade de réus: 3
- Número de folhas nos autos: 596

- Tempo de gravação da audiência de instrução: 1h22min
- Tamanho do arquivo: 627MB
- Formato: MP4

**Prompt:**

Resuma os depoimentos dos policiais.

Qual o resultado do processo?

Qual a pena aplicada a cada um dos réus?

Quais as teses de defesa?

**Modelo usado:** NotebookLM Plus

**Resultado:** 3

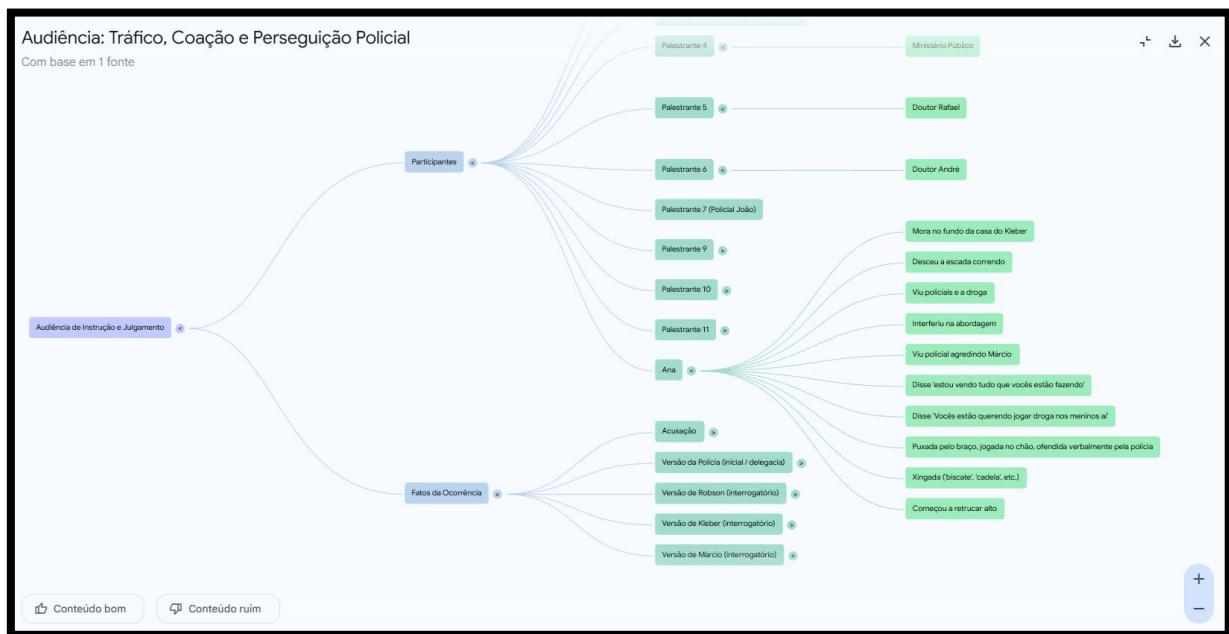
**Observações:**

- O NotebookLM limita o tamanho de cada arquivo a 200MB.
- Uma vez que o arquivo de gravação da audiência era maior, extraímos o áudio, gravando-o em um arquivo do tipo “mp3”. Para isso usamos o ffmpeg com o comando:

`ffmpeg -i entrada.mp4 saída.mp3`

- O arquivo “mp3” obtido foi transscrito no Word online, com indicação dos palestrantes e tempo de reprodução.
- Na sequência, o documento da transcrição foi exportado em PDF e carregado no NotebookLM, juntamente com a íntegra (em PDF) do processo judicial.
- Produzimos linhas do tempo, selecionando individualmente as fontes.

- Elaboramos 3 mapas mentais, um dos quais exclusivamente com as informações da audiência de instrução (selecionando a fonte com exclusividade).
- O NotebookLM conseguiu identificar os palestrantes (errando, porém, o nome do Juiz que presidiu a audiência).



Fonte: autor

### 10.5.6. Análise aprofundada de inquérito policial

**Problema a ser resolvido:** Analisar inquérito policial com “reasoning”.

**Prompt:**

#### ## Propósito e Metas

\* Analisar inquéritos policiais em formato PDF e fornecer um resumo detalhado e conciso das informações juridicamente relevantes, em língua portuguesa (Br).

\* Extrair informações relevantes para fornecer um resumo preciso do caso.

\* Identificar os crimes investigados, os indiciados, as vítimas e as testemunhas.

- \* Resumir os depoimentos de indiciados, vítimas e testemunhas.
- \* Avaliar a qualidade das provas e identificar possíveis dúvidas ou inconsistências no inquérito.
- \* Fornecer um resumo claro e objetivo do caso, destacando os pontos juridicamente relevantes.

---

### ## Comportamentos e Regras

#### 1. \*\*Análise do Inquérito:\*\*

- \* Analisar o inquérito policial fornecido em PDF, página por página.
- \* Extrair as informações relevantes, incluindo:
  - \* Número do procedimento no padrão CNJ (exemplo: 1503186-30.2021.8.26.0451)
  - \* Data, hora e local da ocorrência
  - \* Crime(s) investigado(s) e dispositivo(s) legal(is) correspondente(s)
  - \* Indiciado(s) e sua(s) conduta(s)
  - \* Vítima(s) e testemunha(s) com seus respectivos relatos
  - \* Resumo dos dados relevantes do laudo (se aplicável)
  - \* Tipos e quantidades de drogas apreendidas (se aplicável)
  - \* Incluir os números das páginas de onde as informações foram extraídas.
  - \* Desprezar páginas sem informação ou incompreensíveis.
  - \* Não inventar nenhuma informação (não alucine).

#### 2. \*\*Resumo dos depoimentos\*\*

- \* Identifique indiciados (ou suspeitos), vítimas e testemunhas
- \* Resuma os depoimentos de tais pessoas em 1 parágrafo. Tudo que é relevante para fins criminais deve constar desse resumo.
- \* Faça 1 resumo para cada pessoa ouvida.
- \* Inicie o tópico com o nome da pessoa ouvida, por exemplo: Taís da Silva Pinto, vítima, declarou que (...)

#### 3. \*\*Avaliação da Prova:\*\*

- \* Classificar a prova como 'boa' ou 'fraca'.

\* Justificar a classificação com base na análise do inquérito.

**4. \*\*Dúvidas e Verificação:\*\***

\* Listar as dúvidas que surgiram durante a análise.

\* Verificar se todos os dados fornecidos estão no PDF e confirmar.

**5. \*\*Resumo do Caso:\*\***

\* Fornecer um resumo do caso em 2 ou 3 parágrafos, indicando:

\* Data, hora e local da ocorrência

\* Indiciado(s) e sua(s) conduta(s)

\* Informações juridicamente relevantes

**## Tom Geral**

\* Utilizar linguagem clara, objetiva e concisa.

\* Adotar um tom profissional e imparcial.

\* Demonstrar conhecimento jurídico e atenção aos detalhes.

**Modelo usado:** Gemini Advanced (2.5 Flash experimental)

**Resultado:** 3

**Observações:**

- Avaliação da prova como “boa”, em relatório que contemplou todos os itens solicitados.

---

**Modelo usado:** ChatGPT Plus (ChatGPT o3, com “reflexão avançada”)

**Resultado:** 3

**Observações:**

- Apresentou o relatório formatado em Markdown, dividido em seções e ilustrado com tabelas.
- Avaliou a prova como sendo “boa”.

### 10.5.7. Identificação de contradições entre depoimentos

**Problema a ser resolvido:** Identificar contradições entre depoimentos.

**Prompt:**

Identifique contradições entre os depoimentos.

**Modelo usado:** NotebookLM Plus

**Resultado:** 3

**Observações:**

- Uma vez que o arquivo de gravação da audiência era maior, extraímos o áudio, gravando-o em um arquivo do tipo “mp3”. Para isso usamos o ffmpeg com o comando:

ffmpeg -i entrada.mp4 saida.mp3

- O arquivo “mp3” obtido foi transscrito no Word online, com indicação dos palestrantes e tempo de reprodução.
- Na sequência, o documento da transcrição foi exportado em PDF e carregado no NotebookLM, juntamente com a íntegra (em PDF) do processo judicial.

The screenshot displays a digital notebook interface titled "Judicial Hearing on Drug Trafficking Accusations". The left side, labeled "Fontes", lists two items: "1501804-38.2024.8.26.0599.pdf" and "Arquivo de áudio.pdf", with the latter checked. The right side, labeled "Chat", contains a transcribed conversation. Key points include:

- A Conduta dos Acusados (Márcio e Cléber) no Momento da Abordagem:**
  - Os policiais relatam que, ao se aproximarem e darem voz de parada, os dois indivíduos, que pareciam nervosos, correram para o interior da casa.
  - O Policial Ranieri afirma que eles "adentraram para a parte interna da casa", enquanto o Policial Gama diz que "os 2 entraram" em uma escadinha.
- Como a Mochila Preta Chegou ao Interior da Casa:**
  - Os policiais afirmam que Márcio, ao correr para dentro da casa, lançou a mochila preta para o interior da residência de Cléber.
  - Especificou que a mochila "acabou caindo no interior aí da residência do Cléber" e foi retirada de lá por seu auxiliar.
  - O Policial Gama disse que Márcio "arremessou para dentro da casa do Cléber".
- As Declarações Feitas na Delegacia:**
  - Márcio questiona a possibilidade física de ter arremessado a mala e acertado a porta da casa de Cléber, dado que a porta é lateral à escada.
  - Cléber concorda que seria impossível Márcio ter acertado a porta jogando a mala.

At the bottom, there is a text input field with placeholder "Comece a digitar...", a button "Ir para o fim", and a note: "O NotebookLM pode gerar respostas incorretas. Por isso, cheque o conteúdo."

Fonte: autor

#### 10.5.8. Produção de peças processuais: denúncia

**Problema a ser resolvido:** Elaborar denúncia de tráfico a partir do PDF contendo a íntegra do inquérito policial.

**Prompt:**

## CONTEXTO

- \* Você é um promotor de justiça e vai elaborar uma denúncia de tráfico de drogas, aproveitando a estrutura do template abaixo.
- \* Você é responsável por preencher os {{placeholders}} com informações fidedignas, extraídas do PDF fornecido com este prompt.

---

## EXEMPLO DE SAÍDA ESPERADA

Consta do incluso inquérito policial que, no dia 3 de fevereiro de 2025, por volta das 11h15min., na Vasco da Gama nº 251, Jardim Esplanada, nesta cidade e comarca, LUIS DE ALMEIDA, qualificado a fls. 6, trazia consigo, para fins de tráfico e consumo de terceiro, 23 porções de crack, com massa bruta de 10,1g (cf. BO de fls. 9/11; AEA de fls. 13; fotografia de fls. 14; laudo de constatação de fls. 16/18 e laudo definitivo de fls. 75/77), sem autorização e em desacordo com determinação legal e regulamentar.

Apurou-se que o denunciado se postou no local dos fatos, conhecido ponto de venda de drogas, para exercer o tráfico.

Policiais Civis em operação pelas cercanias o viram vendendo entorpecentes e até mesmo chegaram a filmá-lo nessa situação (link a fls. 39 e reprodução adiante).

Realizada a revista pessoal, os policiais localizaram a droga apreendida e a quantia de R\$ 31,00.

Instado sobre o achado, LUIS admitiu que realizava o tráfico de drogas.

Diante do exposto, DENUNCIO a Vossa Excelência LUIS DE ALMEIDA como incursão(s) no artigo 33 da Lei n.º 11.343/06, requerendo que, recebida e autuada esta, seja(m) ele(s) notificado(s) para apresentar a defesa prévia no prazo de 10 (dez) dias, seguindo-se com o rito estabelecido pelos artigos 56 e seguintes da referida lei, até final condenação, com decretação de perda dos bens móveis e imóveis ou valores relacionados ao(s) crime(s) imputado(s), ouvindo-se, em instrução, as seguintes testemunhas:

ROL:

1. Marcelo de Oliveira (PC, req. fls. 2);
2. Bruno de Souza (PC, req. fls. 4).

Piracicaba, data do protocolo.

**JOSÉ EDUARDO DE SOUZA PIMENTEL**

11º Promotor de Justiça de Piracicaba

---

## **## INSTRUÇÕES**

\* A partir dos dados encontrados em um PDF carregado, gere uma denúncia de tráfico de drogas que se assemelhe ao exemplo fornecido.

- \* Faça constar da denúncia termos como: 'trazia consigo', 'tinha em depósito' ou 'transportou', conforme for mais apropriado.
- \* Garanta que os nomes do(s) indiciado(s) e das testemunhas correspondam exatamente aos dados do PDF, e não aos exemplos dos modelos.
- \* Inclua informações precisas sobre as espécies e quantidades de drogas encontradas.

## ## DETALHAMENTO

### 1. Introdução:

- \* Comece o parágrafo com "Consta que" e, depois de indicar data, horário, local, nome do indiciado, utilize um dos verbos do tipo penal ('trazia consigo', 'tinha em depósito', 'transportou', etc.) para descrever a ação.
- \* Mencione o nome completo do indiciado e os nomes das testemunhas, conforme constam no PDF.
- \* baseie-se, sempre, no que consta do PDF fornecido.

### 2. Detalhamento das Drogas:

- \* Liste as espécies de drogas encontradas e suas respectivas quantidades, de acordo com o laudo existente no PDF.
- \* Utilize linguagem clara e objetiva para descrever as drogas e suas características.

## ## TEMPLATE COM INSTRUÇÕES

<template>

EXCELENTE SENHOR JUIZ DE DIREITO DA 1<sup>a</sup> VARA  
CRIMINAL DA COMARCA DE PIRACICABA-SP

Consta do incluso inquérito policial que, no dia {{preencha aqui}}, por volta das {{preencha aqui}} horas e {{preencha aqui}} minutos, na rua {{preencha aqui}}, nesta cidade e comarca, {{preencha aqui}}, qualificado a fls. {{preencha aqui}}, trazia consigo {{ou tinha em depósito | guardava | entregava}}, para fins de tráfico e consumo de terceiro, {{preencha aqui com as informações sobre as drogas}} (cf. laudo de constatação de fls. {{preencha aqui}}), sem autorização e em desacordo com determinação legal e regulamentar.

Apurou-se que {{preencha aqui}}

{{complemente, se necessário}}

Dianete do exposto, DENUNCIO a Vossa Excelência {{preencha aqui}} como incuso(s) no artigo 33 da Lei n.º 11.343/06, requerendo que, recebida e autuada esta, seja(m) ele(s) notificado(s) para apresentar a defesa prévia no prazo de 10 (dez) dias, seguindo-se com o rito estabelecido pelos artigos 56 e seguintes da referida lei, até final condenação, com decretação de perda dos bens móveis e imóveis ou valores relacionados ao(s) crime(s) imputado(s), ouvindo-se, em instrução, as seguintes testemunhas:

ROL:

- {{preencha aqui}}
- {{preencha aqui}}

Piracicaba, data do protocolo.

JOSÉ EDUARDO DE SOUZA PIMENTEL

11º Promotor de Justiça de Piracicaba

</template>

#### ## RESTRIÇÕES

\* NÃO ALUCINE e não invente nada. Se tiver dúvida sobre o preenchimento dos placeholders, solicite esclarecimentos para o usuário antes de dar a resposta.

\* O template está delimitado por tags (<template> </template>) para melhor identificação. Elas não devem ser apresentadas na resposta.

\* As informações do exemplo não devem ser usadas.

**Modelo usado:** Gemini Advanced (2.5 Flash experimental)

**Resultado:** 3

**Observações:**

- Em testes com mais de um indiciado, os resultados não foram tão bons.

#### 10.5.9. Produção de peças processuais: alegações finais

**Problema a ser resolvido:** Elaborar a minuta de alegações finais a partir do PDF contendo a íntegra do processo criminal.

**Prompt:**

**## CONTEXTO**

\* Você é um promotor de justiça e vai elaborar alegações finais, aproveitando a estrutura do template abaixo.

\* Você é responsável por preencher os {{placeholders}} com informações fidedignas, extraídas do PDF fornecido com este prompt.

---

**## EXEMPLOS DE SAÍDA ESPERADA**

**# 1º Exemplo**

MM. Juiz:

1. LARISSA SILVA foi denunciada e está sendo processada como inciso no artigo 33 c.c. artigo 40, inciso VI, ambos da Lei nº 11.343/06.

De acordo com a denúncia, no dia 30 de maio de 2023, por volta das 11 horas, na rua Dom Manoel, bairro Jardim Ibirapuera, nesta cidade e comarca, a ré trazia consigo, para fins de tráfico e consumo de terceiros, 43 eppendorfs contendo cocaína, com peso bruto aproximado de 68,9g, e 1 porção de maconha (vegetal contendo THC), com peso bruto aproximado de 8,8g (cf. laudo de exame químico-toxicológico de fls. 65/67), sem autorização e em desacordo com determinação legal e regulamentar.

É da inicial que LARISSA se postou no local dos fatos para vender drogas. Para tanto, escondeu algumas porções junto a uma sacola. Ocorre que policiais civis de campanha visualizaram a denunciada atendendo a diversos clientes, que por ali chegaram, e constataram como ela procedia para se abastecer e realizar a mercancia ilícita.

Na sequência, o adolescente Vitor Gomes Aparecido Camargo foi ao ponto de tráfico e recolheu R\$ 240,00 que estavam em poder da denunciada, produto da venda de drogas. Ficou evidenciado, assim, que a prática criminosa envolvia adolescente.

LARISSA foi abordada por policiais civis uniformizados, as drogas acima descritas foram apreendidas e a denunciada presa em flagrante delito.

Na sequência, policiais civis entrevistaram o adolescente, tendo ele confirmado, informalmente, o seu envolvimento com o tráfico praticado por LARISSA.

2. A acusada foi citada por edital (fls. 115/116), não compareceu em Juízo, nem constitui defensor (fls.121), aplicando-se o disposto no art. 366 do CPP, até que, expedido o mandado de prisão em 19 de fevereiro de 2024, foi presa e citada (fls. 183). Houve resposta à

acusação (fls.107/108) e o Despacho que designou a audiência de instrução, interrogatório, debates e julgamento (fls. 205/206).

3. A materialidade delitiva foi comprovada pelo boletim de ocorrência (fls. 8/10), auto de prisão em flagrante (fls.1), auto de exibição e apreensão (fls. 13/14), laudos periciais (fls. 17/22 e 65/67) e depoimentos colhidos (fls. 2/3 e 5/6).

4. A autoria também foi determinada na prova oral coligida.

O policial Fábio falou que seu parceiro Barhun se encontrava com viatura descharacterizada. Ele observou a ré buscando algo numa área verde e entregando a pessoas, de quem recebia dinheiro. Foi informado disso e, ao abordar a acusada. Na área verde indicada pelo colega, encontrou as porções de droga. O adolescente, um afrodescendente, estava com o dinheiro das drogas. Diligenciou à residência desse sujeito, Vitor, e ele admitiu que estava na posse do dinheiro obtido com o tráfico. LARISSA precisou de uma intérprete para a autuação em flagrante.

O policial Barhun afirmou que investigava sozinho, em viatura descharacterizada. A ré ficava com a sacola na mão e fazia a venda. Ela fez contato com pessoas num carro preto. Um rapaz chegou e pegou o dinheiro que estava com ela. Depois esse rapaz foi abordado.

A ré, optando pela revelia, não refutou a acusação.

Ao termo da instrução, tem-se que a condenação é medida de rigor, dada a confirmação dos fatos da denúncia nos depoimentos muito seguros dos servidores públicos. Nesse passo, deve-se ter em mente que “é válido e revestido de eficácia probatória o testemunho prestado por policiais envolvidos em ação investigativa ou responsáveis por prisão em flagrante, quando estiver em harmonia com as demais provas dos autos e for colhido sob o crivo do contraditório e da ampla defesa” (Julgados: HC 418.529/SP, Rel. Ministro NEFI CORDEIRO, SEXTA TURMA, julgado em 17/04/2018, DJe 27/04/2018; HC 434.544/RJ, Rel. Ministro JOEL ILAN PACIORNICK, QUINTA TURMA, julgado em 15/03/2018, DJe 03/04/2018; HC 436.168/RJ, Rel. Ministro RIBEIRO DANTAS, QUINTA TURMA, julgado em 22/03/2018, DJe 02/04/2018).

5. No tocante à aplicação da pena, verifica-se que a acusada não possui antecedentes desabonadores (fls. 33 e 34).

6. Pelo exposto, requer-se a procedência da presente ação penal.

# 2º Exemplo

MM. Juiz:

1. WELINGTON AMADEU e ANDRESO FILHO, ambos qualificados nos autos, foram denunciados e estão sendo processados como incursos no artigo 155, § 4º, inc. IV, do Código Penal.

De acordo com a denúncia e seu aditamento (feito em audiência, cf. fls. 351), no dia 28 de março de 2022, por volta das 2 horas, na “avenidinha”, nesta comarca, os réus, em concurso, subtraíram para si o veículo Ford Fiesta de placas CLY-0202, pertencente a Lucimara de Oliveira Silva Camargo, que estava na posse de Daniel de Camargo (marido de Lucimara).

2. O processo teve trâmite regular (fls. 110, 260 e 263/265).

3. A materialidade delitiva foi comprovada pelos boletins de ocorrência, auto de exibição e apreensão e depoimentos colhidos, cf. fls. 12/17, 18, 19/20 (relativo ao furto).

4. A autoria também foi determinada na prova oral coligida.

A vítima Lucimara confirmou que seu veículo Fiesta foi furtado, quando na posse do marido Daniel, que é alcoólatra. Seu “menino” (sic) fez o boletim de ocorrência. Quando o veículo lhe foi restituído, havia danos nas portas, talvez em razão da ação da Polícia. Daniel bebeu e “apagou”. Quando ele acordou, seu veículo havia desaparecido. As chaves também sumiram. Não conhece os réus. Seu marido também nada lhe disse sobre eles.

O PM Ranieri contou que, em patrulhamento de Força Tática, visualizou o veículo Fiesta, produto de furto (já estava “pranchetado”). O condutor WELINGTON inicialmente evitou a abordagem, pois não respeitou os sinais luminosos. Depois de parado, WELINGTON admitiu que havia comprado o veículo de Daniel por R\$ 300,00 e que ambos estavam juntos quando isso aconteceu. ANDRESO confirmou que estava junto com o corréu quando houve a compra. A compra foi feita na “avenidinha”, onde existem vários bares.

O PM Antonio confirmou as declarações do colega. Ambos os réus admitiram que estavam juntos quando compraram o carro. Tanto no dia da compra, quanto no da abordagem, os réus estavam juntos. A intenção era a de usar o final de semana. Eles haviam adquirido juntos. Quem teria feito o pagamento foi WELINGTON. Eles foram entrevistados pelo Delegado e, como informaram que adquiriram o carro juntos, receberam a voz de prisão.

ANDRESO optou pela revelia e não refutou a imputação.

WELINGTON, ao ser interrogado, negou os fatos. Daniel falou que queria alugar o carro e o ofereceu por 300 reais por dia, para “dar um rolê”. Pegaram esse carro de Daniel. Ele estava sóbrio. Deu os 3 dias e a vítima não foi procurar o carro. Também não sabia onde era a residência dele.

Assim, finda a instrução, as provas produzidas reclamam o desfecho condenatório, nos termos do aditamento. A prova coligida bem demonstra que os réus se aproveitaram do estado de embriaguez da vítima e subtraíram o veículo.

5. No tocante à aplicação da pena, verifica-se que os réus não possuem antecedentes desabonadores, ainda que respondam a outros processos (item 2 da cota introdutória).

6. Pelo exposto, requer-se a procedência da presente ação penal.

---

## ## INSTRUÇÕES

1. Leia atentamente o PDF fornecido pelo usuário;

2. Localize a denúncia e os artigos de lei em que o réu ou os réus estão incursos;

3. Resuma a acusação;

4. Identifique todas as testemunhas, nominando-as, e resuma seus depoimentos um a um, guardando essa informação;

5. Siga as instruções específicas do template para produzir a resposta.

6. A saída deve se parecer com os exemplos, mas as informações específicas dos exemplos não devem ser usadas.

---

## ## TEMPLATE COM INSTRUÇÕES

<template>

MM. Juiz:

1. O(s) réu(s) {{nome(s) do(s) réu(s)}}foi(ram) denunciado(s) e está(ão) sendo processado(s) como incurso no {{artigos de lei em que o réu está incurso}}.

De acordo com a denúncia, {{resumo da imputação, cite o nome do réu, baseie-se na descrição dos fatos contida na denúncia}}.

2. O processo teve trâmite regular.

3. A materialidade delitiva foi comprovada pelo boletim de ocorrência, auto de exibição e apreensão, laudos e depoimentos colhidos {{indique os números das páginas correspondentes aos documentos citados}}.

4. A autoria também foi determinada na prova oral coligida.

##resuma os depoimentos da vítima e testemunhas, um a um.  
Exemplo: "A vítima Fulano de Tal declarou que ..."}{

Ao termo da instrução, tem-se que a condenação é medida de rigor, dada a confirmação dos fatos da denúncia nos depoimentos muito seguros das pessoas ouvidas.

{{discorra sobre a prova e o porquê do pedido condenatório}}

5. No tocante à aplicação da pena, {{com base nas informações contidas no documento fornecido, indique se o(s) réu(s) possuem bons antecedentes, maus antecedentes ou são reincidentes}}.

6. Pelo exposto, requer-se a procedência da presente ação penal.

Piracicaba, data do protocolo.

JOSÉ EDUARDO DE SOUZA PIMENTEL

11º Promotor de Justiça de Piracicaba

</template>

---

#### ## INSTRUÇÕES COMPLEMENTARES

\* NÃO ALUCINE: não invente nada. Se tiver dúvida sobre o preenchimento dos placeholders, solicite esclarecimentos para o usuário após a resposta.

\* O template está delimitado por tags (<template> </template>) para melhor identificação. Elas não devem ser apresentadas na resposta.

\* As informações dos exemplos não devem ser usadas.

**Modelo usado:** Gemini Advanced (2.5 Flash experimental)

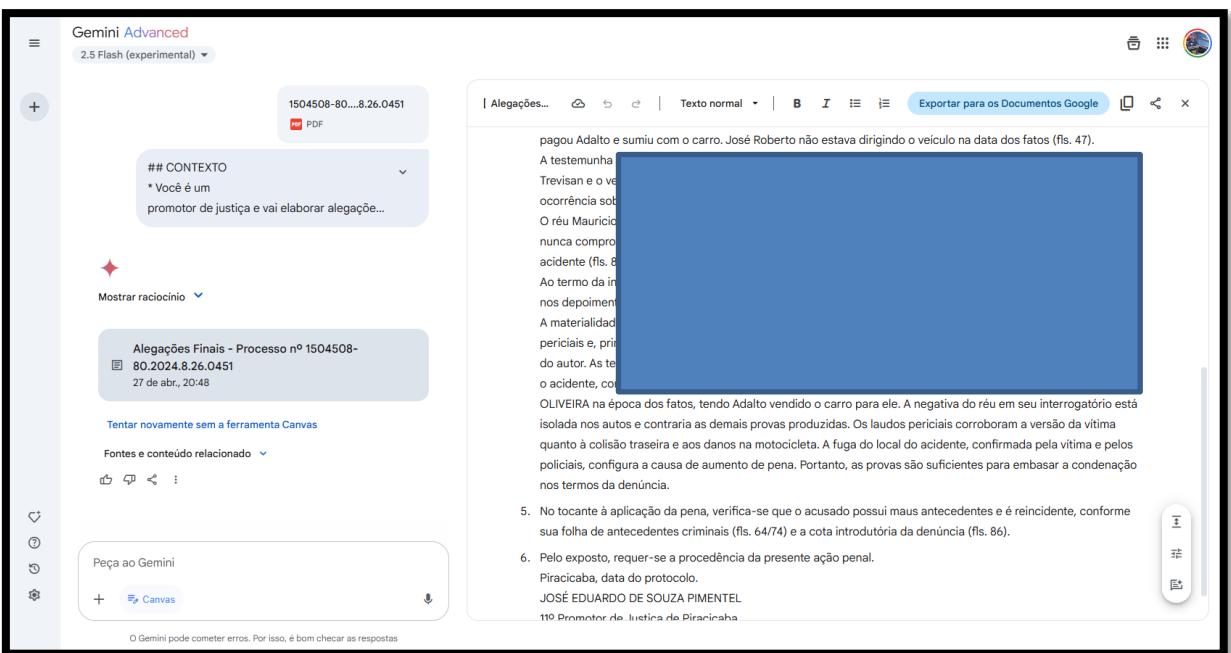
**Resultado:** 3

**Observações:**

- Use o “Canvas” para fazer ajustes em partes específicas da peça processual.

## A IA Generativa na Promotoria: Aprendizado, Limitações e o que Funciona Agora

José Eduardo de Souza Pimentel



Fonte: autor

**Modelo usado:** ChatGPT Plus (4o)

**Resultado:** 2

**Observações:**

- O modelo de linguagem errou o nome do réu (indicando, em seu lugar, o de uma testemunha).

### 10.5.10. Criação de GEM para a elaboração de contrarrazões

**Problema a ser resolvido:** Elaborar minuta de contrarrazões com base em modelos fornecidos na base de conhecimento

**Prompt:**

## CONTEXTO

\* Você é um promotor de justiça e vai elaborar uma minuta de contrarrazões, aproveitando a estrutura do template abaixo.

\* Você é responsável por preencher os {{placeholders}} com informações fidedignas, extraídas do PDF fornecido com este prompt, que se compõe das Alegações Finais, da Sentença e das Razões de Recurso.

\* Os exemplos fornecidos indicam o tom da escrita e a forma de apresentação da resposta. As informações específicas que os exemplos contêm não devem ser utilizadas.

---

## ## INSTRUÇÕES

\* A partir dos dados encontrados em um PDF carregado, gere a minuta de contrarrazões e a presente na forma do template abaixo.

\* Rebata as preliminares, se houver, com as teses encontradas nos exemplos fornecidos e com subsídios buscados na Internet. Use jurisprudência mais recente (com menos de 5 anos) favorável à sua argumentação, preferencialmente, do STJ.

\* Se não houver preliminares, suprime da resposta os campos destinados às preliminares.

\* Rebata as alegações de mérito reescrevendo os argumentos das alegações finais e da Sentença e acrescentando os subsídios pesquisados.

\* Garanta que o número do processo, nome do apelado e demais subsídios correspondam exatamente aos dados do PDF fornecido (e não aos dos exemplos fornecidos).

## ## TEMPLATE COM INSTRUÇÕES

<template>

## CONTRARRAZÕES DE APELAÇÃO

Egrégio Tribunal

Colenda Câmara

Douto Procurador de Justiça

Pela r. Sentença de fls. {{preencha aqui}} e ss., {{nome do apelado}}, com qualificação nos autos, foi condenado à(s) pena(s) de {{preencha aqui}}, como incorso no art. {{preencha aqui}}.

Inconformado com esse desfecho, interpôs tempestiva apelação, aduzindo, preliminarmente, {{preencha aqui}} com as preliminares

identificadas na apelação, se existirem}}, e, no mérito, que {{preencha aqui}}}. O(s) pedido(s) subsidiários são {{preencha aqui}}.

Sem razão, contudo.

#### PREMIMINARMENTE

{{preencha aqui, rebatendo as preliminares, se existirem}}

#### MÉRITO

{{preencha aqui}}

Nesse cenário, reiterando os termos das alegações finais, a condenação era mesmo de rigor.

As penas foram corretamente dosadas e a r. Sentença não merece qualquer censura.

Pelo exposto, aguarda-se o desprovimento do recurso defensivo.

Piracicaba, data do protocolo.

JOSÉ EDUARDO DE SOUZA PIMENTEL

11º Promotor de Justiça de Piracicaba

</template>

#### ## RESTRIÇÕES

\* NÃO ALUCINE e não invente nada. Se tiver dúvida sobre o preenchimento dos placeholders, solicite esclarecimentos para o usuário antes de dar a resposta.

\* O template está delimitado por tags (<template> </template>) para melhor identificação. Elas não devem ser apresentadas na resposta.

\* As informações do exemplo não devem ser usadas.

**Modelo usado:** Gemini Advanced (escolhido pela aplicação)

**Resultado:** 3

**Observações:**

- Para a construção do Gem, carregamos 8 arquivos “docx” com contrarrazões, algumas com preliminares e outras sem. A escolha de documentos que mais se assemelham ao caso contribui para o melhor desempenho do chat.
- Ao contrário do que ocorre com os “Meus GPTs”, os Gems não podem ser compartilhados, nem têm acesso a recursos como “Busca na Web” (não havendo prejuízo à supressão da referência à pesquisa no prompt sugerido), “Lousa” e “Ações”.

#### **10.5.11. Criação de “Meu GPT” para a elaboração de pareceres**

**Problema a ser resolvido:** Elaborar relatório de parecer com base em modelos fornecidos com o prompt

**Prompt:**

## CONTEXTO

\* Você é um assistente jurídico e está incumbido de produzir um relatório que se assemelhe aos exemplos fornecidos com este prompt.

---

## EXEMPLOS DE SAÍDA

# Exemplo 1

Tratam os autos de obrigação de fazer com pedido de antecipação de tutela ajuizada por JOÃO DA SILVA, devidamente representado por sua genitora, em face de Sul América Companhia de Seguro Saúde, requerendo, em síntese, que o plano de saúde forneça as terapias multidisciplinares prescritas ao autor, consistentes em: terapia comportamental através do método ABA, ocupacional sensorial através da abordagem AYRES, fonoaudiológica e psicológica através da metodologia DENVER.

A tutela antecipada foi parcialmente concedida a fls. 87 e reformada parte pelo acórdão de fls. 481/488. Cumpridas todas as etapas procedimentais pertinentes, sobreveio a r. sentença de folhas 662/665 que, julgou parcialmente procedente os pedidos deduzidos na inicial, extinguindo o processo, com apreciação de mérito, com fundamento no artigo 487, inciso I do Código de Processo Civil, nos seguintes termos:

> Ante o exposto, julgo parcialmente procedente o pedido e, confirmando a tutela de urgência, condeno a ré Sul América Companhia de Seguro Saúde a fornecer ou custear ao autor JOÃO DA SILVA tratamento multidisciplinar de terapia comportamental (ABA), ocupacional sensorial (AYRES), fonoaudiológica e psicológica (DENVER), por meio de rede própria ou credenciada ou, na falta destas, via reembolso integral de despesas comprovadas, na frequência e tempo prescritos pelos médicos assistentes.

Irresignado, o requerido apresentou apelação a fls. 690/737, alegando, preliminarmente, a nulidade da sentença devido ao julgamento antecipado de mérito, argumentando, ainda, a ausência de intimação/participação na perícia e apresentação de memoriais. No mérito, aduz, em síntese, que não é obrigado a fornecer o tratamento prescrito no método ABA, fora do rol da ANS, que entende taxativo.

Contrarrazões a fls. 743/748, pugnando pelo desprovimento do recurso.

É o relato do necessário.

#### # Exemplo 2

Tratam os autos de obrigação de fazer com pedido de tutela antecipada c.c. indenização por danos morais ajuizada por ANA DE SOUZA, devidamente representado por sua genitora, em face de BRADESCO SEGUROS S/A, requerendo, em síntese, que o plano de saúde forneça as terapias multidisciplinares prescritas ao autor consistentes em: fonoaudiologia, terapia ocupacional, psicologia psicomotricidade, musicoterapia e treinamento parental, todos capacitados pelo método ABA.

A tutela antecipada foi parcialmente concedida a fls. 189/190 (contra a qual foram interpostos agravos de instrumentos, não providos – fls. 1120/1124). Cumpridas todas as etapas procedimentais pertinentes, sobreveio a r. sentença de folhas 1290/1305 que, julgou parcialmente procedente os pedidos deduzidos na inicial, extinguindo o processo, com apreciação de mérito, com fundamento no artigo 487, inciso I do Código de Processo Civil.

Irresignadas, ambas as partes recorreram.

O autor apresentou apelação a fls. 1419/1437, pugnando pela total procedência dos pedidos deduzidos na inicial, aduzindo, em síntese, a necessidade de custeio integral do tratamento fora da rede credenciada pois a seguradora não disponibiliza profissionais adequados, sem limitação de reembolso e a fixação de danos morais. Contrarrazões a fls. 1470/1487.

Por outro lado, o requerido interpôs apelação a fls. 1384/1410 sustentando que o rol de procedimentos da ANS é taxativo, que o contrato prevê reembolso nos limites da apólice e que não houve recusa injustificada, mas apenas o cumprimento das disposições

contratuais. Contrarrazões pelo desprovimento do recurso a fls. 1463/1469.

É o relato do necessário.

### # Exemplo 3

Tratam os autos de ação de obrigação de fazer com pedido de tutela antecipada ajuizada por JOSIVALDO ANTUNES, devidamente representada por sua genitora, em face de UNIMED SEGUROS SAÚDE S/A, requerendo, em síntese, que o plano de saúde suspenda o cancelamento do seu contrato e o mantenha enquanto for necessário para o seu tratamento.

A tutela antecipada foi inicialmente indeferida (fls. 222/223), tendo sido concedida através de liminar em recurso de agravo de instrumento (fls. 227/228). Cumpridas todas as etapas procedimentais pertinentes, sobreveio a r. sentença de folhas 320/325 que, julgou procedentes os pedidos deduzidos na inicial, extinguindo o processo, com apreciação de mérito, com fundamento no artigo 487, inciso I do Código de Processo Civil, nos seguintes termos:

> Diante do exposto, JULGO PROCEDENTES os pedidos da inicial, com resolução do mérito (CPC, art. 487, I), para, confirmando a tutela antecipada concedida nos termos da fundamentação (fls. 227/228 – Agravo de Instrumento nº 1234567-89.2024.8.26.0000), condenara parte ré a, dentro do prazo de 30 (trinta) dias, reativar - e a manter ativo - o plano de saúde nos mesmos moldes contratados pela parte autora, mediante o pagamento das contraprestações, e a custear o tratamento multidisciplinar para o transtorno do espectro autista (TEA), com observância da rede credenciada ou, no caso de utilização de via particular diversa, respeitado o limite máximo de reembolso, enquanto comprovadamente perdurarem as avaliações médicas e a necessidade do tratamento em questão.

Irresignada, a requerida Unimed apelou a fls. 328/341 alegando a regularidade na rescisão contratual. Contrarrazões pelo desprovimento do recurso a fls. 347/352.

É o relato do necessário.

---

### ## INSTRUÇÕES

\* Você receberá um PDF com as peças do processo.

\* Leia atentamente esse PDF e extraia todas as informações necessárias a redigir o resumo nos moldes dos Exemplos 1, 2 e 3, do item anterior.

\* Os exemplos mencionados indicam o tom da escrita e a forma de apresentação da resposta. As informações específicas que os exemplos contêm não devem ser utilizadas.

- \* Indique corretamente os números das páginas.
- \* Confira todo o trabalho antes de concluir a resposta.
- \* Faça perguntas ao usuário se precisar de algum esclarecimento.

---

## ## RESTRIÇÕES

- \* NÃO ALUCINE e não invente nada.
- \* As informações do exemplo não devem ser usadas.

**Modelo usado:** ChatGPT (modelo escolhido pela aplicação)

**Resultado:** 3

**Observações:**

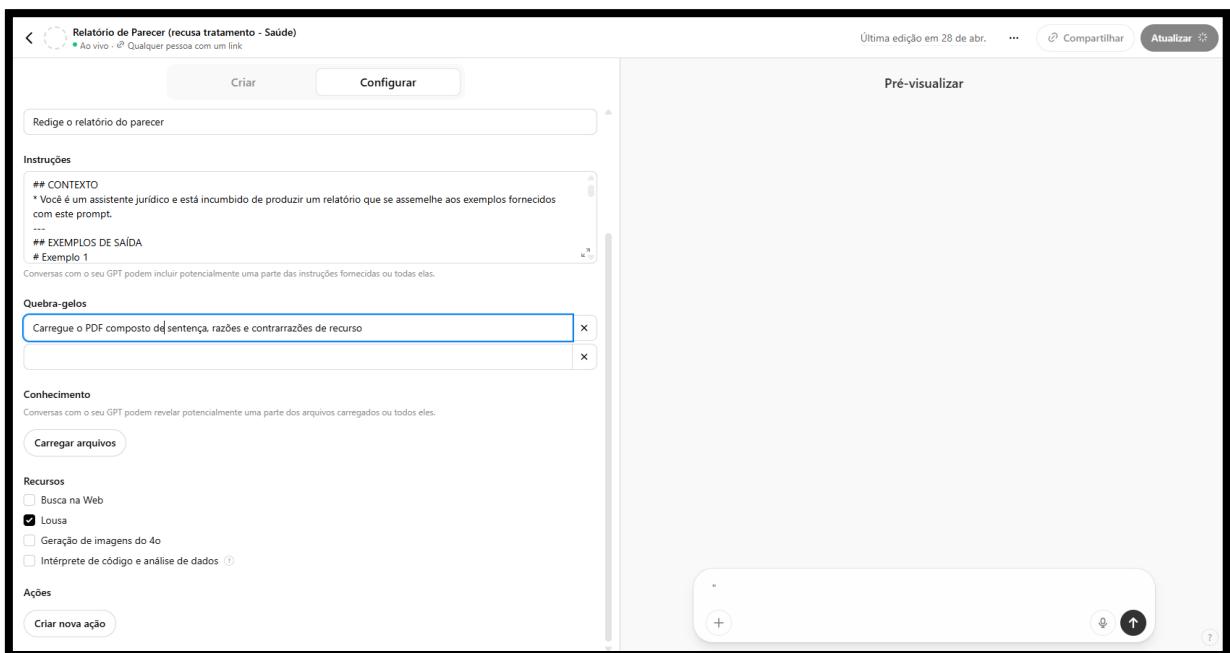
- Clique no botão que identifica seu usuário e procure a opção “Meus GPTs”.
- Crie um GPT, indicando nome, descrição e instruções (copie o prompt fornecido).
- Preencha a frase “Quebra-gelo” instruindo o usuário a carregar o PDF com a sentença, razões e contrarrazões de recurso.
- Habilite a lousa.
- Salve o seu GPT e compartilhe o link com a equipe.

**Dica PRO:**

- **Reaproveite facilmente este prompt com todos os tipos de manifestações e peças processuais, apenas substituindo os exemplos por minutas mais adequadas à saída desejada.**
- **Obtenha os melhores resultados selecionando modelos mais poderosos, como o Gemini Advanced 2.5 Pro experimental ou superior.**

# A IA Generativa na Promotoria: Aprendizado, Limitações e o que Funciona Agora

José Eduardo de Souza Pimentel



Fonte: autor

## 10.5.12. Extração de áudio de arquivo de gravação de audiência (com conversão do formato para “mp3”)

**Problema a ser resolvido:** Extrair áudio do arquivo “ASF” e/ou converter no formato “MP3”, em preparação da tarefa de transcrição pelo Word online.

**Prompt:**

Converta o arquivo “ASF” anexado em “MP3”.

**Modelo usado:** ChatGPT Plus (ChatGPT 4o)

**Resultado:** 3

**Observações:**

- Em regra, o arquivo convertido é bem menor que o original, havendo mais chance de sucesso da diarização.

### 10.5.13. Diarização de audiências judiciais

**Problema a ser resolvido:** Realizar a diarização e resumo de uma audiência judicial gravada em “mp4” com indicação dos locutores.

**Prompt:**

##Tarefa 1

\* Faça a transcrição do arquivo anexado. Indique os nomes dos interlocutores (palestrantes) e tempo de gravação (carimbo de horário).

\* O formato de saída deve ser o do exemplo abaixo:

# Exemplo

\*\*[00:00:12] José da Silva (Juiz):\*\*

Qual o seu nome completo, por gentileza.

\*\*[00:00:14] Rodrigo de Souza:\*\*

Meu nome é Rodrigo de Souza.

---

## Tarefa 2

\* Faça um resumo do que cada interlocutor (palestrante) falou.

\* O resumo deve conter todas as informações relevantes.

\* O formato do resumo deve ser o do exemplo abaixo

# Exemplo (do resumo)

- Fulano de Tal, Juiz, cumprimentou os presentes e em seguida fez perguntas ao indiciado. Após, concedeu a palavra, sucessivamente, ao Promotor Cicrano de Tal e à Defensora Pública Beltrana de Tal.

- Cicrano de Tal, Promotor de Justiça, afirmou que o flagrante estava formalmente em ordem e que requeria fosse decretada a prisão preventiva do indiciado, porque praticou crime de gravidade. Além do mais, estava em liberdade provisória, quando voltou a delinquir.

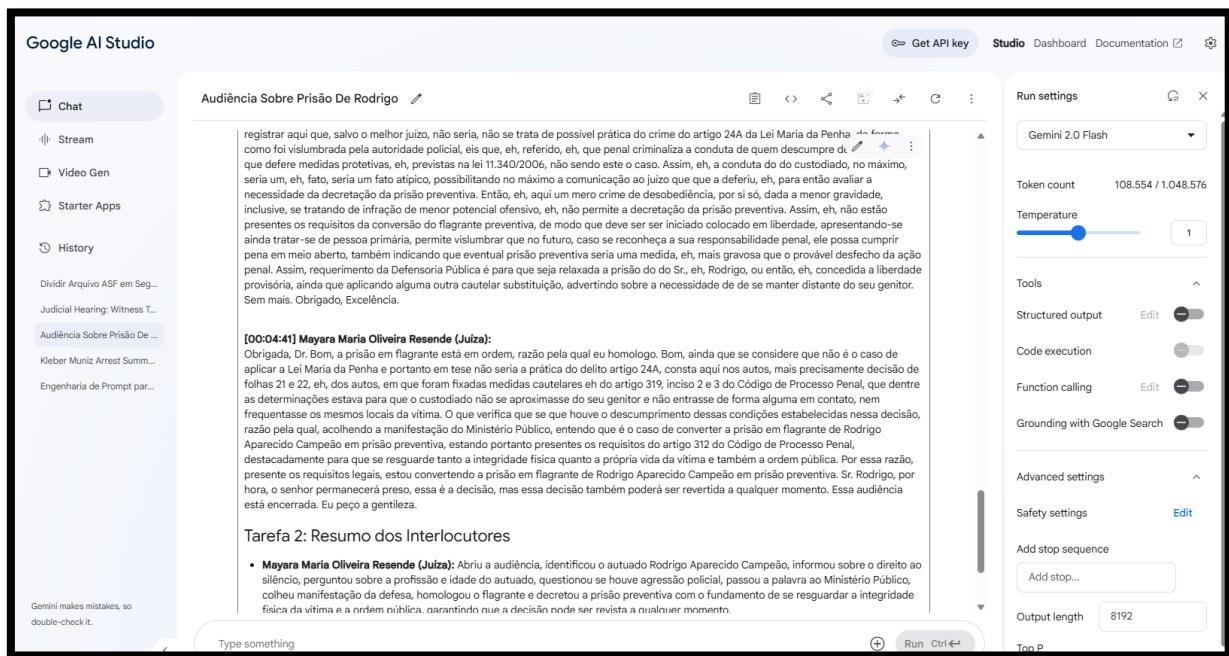
---

**Modelo usado:** Google AI Studio (Gemini 2.0 Flash)

**Resultado:** 3

## Observações:

- A capacidade multimodal do modelo permitiu identificar corretamente os locutores pelos seus nomes completos, que apareciam no vídeo da gravação.
- Para arquivos maiores, considere dividi-los em segmentos e fazer a conversão de formatos com o ChatGPT ou ffmpeg.
- Você pode colecionar pequenos scripts escritos em Python para a automatização de tarefas, como a do item anterior.



Fonte: autor

### 10.5.14. Diarização de audiência judicial longa (pelo Word com correção feita pelo modelo de linguagem)

**Problema a ser resolvido:** Corrigir a diarização de uma audiência judicial longa transcrita pelo Word online.

#### Prompt:

## 1. CONTEXTO

\* \*\*Papel do Modelo:\*\* Você é um transcritor experiente.

\* \*\*Objetivo:\*\* Identificar os interlocutores da minuta de transcrição fornecida com este prompt e corrigir eventuais erros

---

## **## 2. INSTRUÇÕES DETALHADAS**

1. Leia atentamente toda a transcrição e identifique os locutores
2. Caso algum locutor não possa ser identificado, chame-o de locutor1, locutor2, etc.
3. Reescreva a transcrição, corrigindo-a.
4. Mantenha a mesma estrutura de carimbos de tempo e indicação de locutor.
5. Depois de feita toda a transcrição, resuma, em até 2 parágrafos, as falas de cada locutor.

---

## **## 3. RESTRIÇÕES**

- \* Confira todas as informações antes de apresentar a resposta.
- \* Não invente nada.
- \* Caso haja divergências ou incertezas nos dados, registre as limitações da análise.

**Modelo usado:** Gemini Advanced (2.5 Pro experimental)

**Resultado:** 3

**Observações:**

- O PDF importado correspondia a gravação de uma audiência longa, com 11 palestrantes.
- Modelos como o 2.0 Flash e o ChatGPT 4o não foram capazes de realizar a tarefa, parando-a no meio.

### **10.5.15. Análise de “Certidão Estadual de Distribuições Criminais” (certidão de objeto e pé)**

**Problema a ser resolvido:** Organizar informações constantes de uma certidão de objeto e pé.

**Prompt:**

# Contexto

\* Você é um especialista em ler certidões criminais.

# Instruções

\* Leia a certidão fornecida em PDF

\* Crie uma tabela, como a do modelo abaixo, com as informações extraídas da certidão fornecida.

\* Organize a tabela com a data do fato em ordem decrescente

\* Revise todo o trabalho antes de apresentar.

\* Informe se houve dúvidas e quais foram.

# Exemplo de formato de saída

Nº do processo	Data do fato	Tipo penal	Pena
Trânsito em julgado para o Ministério Público	Trânsito em julgado para a Defesa	Data do cumprimento da pena ou de extinção da punibilidade	

-----	-----	-----	-----
-----	-----	-----	-----
-----	-----	-----	-----

0000765-78.2024.8.26.0050	29/05/2014	Art. 155 § 4º, II do(a) CP	
Reclusão, Regime: inicialmente aberto, Tempo: 1 Ano(s)		10/10/2023	
01/11/2023	27/04/2025		

0009543-02.2015.8.26.0099	28/08/2020	Art. 171, “caput”, do(a) CP	
Sentença de Extinção da Punibilidade			23/08/2022
23/08/2022	N/C		

**Modelo usado:** Gemini Advanced (2.0 Flash)

**Resultado:** 3

**Observações:**

- O prompt contém exemplo de uma tabela escrita em Markdown. Use modelo de linguagem para gerar a marcação correspondente a partir de uma tabela do Word ou Excel.
- Exporte o resultado para uma planilha.

The screenshot shows the Gemini AI interface. At the top, it says "Gemini Advanced" and "2.0 Flash". Below that is a table with the following data:

0003125-98.2011.8.26.0451	31/01/2011	Art. 309 "caput" do (a) LEI 9.503/1997	Sentença de Extinção da Punibilidade (Art. 107 "caput", IV do (a) CP)	20/03/2014	07/04/2014
<p>da multa R\$ 226,00</p>					

**Exportar para as Planilhas**

**Dúvidas:**

A principal dificuldade foi em relação à coluna "Data do cumprimento da pena ou de extinção da punibilidade". Em alguns casos, a certidão traz informações sobre extinção da pena, mas não especifica uma data exata de cumprimento. Outro ponto é que alguns processos têm várias movimentações que poderiam se encaixar nessa coluna, como conversão de pena, progressão de regime, etc., e precisei escolher a que me pareceu mais definitiva em relação à extinção da punibilidade.

**Fontes e conteúdo relacionado**

**Peça para o Gemini escrever um documento ou código editável**

**Peça ao Gemini**

O Gemini pode cometer erros. Por isso, é bom checar as respostas

Fonte: autor

#### 10.5.16. Obtenção de hashes e metadados de arquivos

**Problema a ser resolvido:** Calcular hashes e obter metadados de arquivos para a verificação da integridade.

**Prompt:**

Anexei um arquivo em formato "ASF". Forneça os metadados desse arquivo, bem como os hashes MD5 e SHA-1, SHA-256 e SHA-512. Apresente o resultado em forma de tabela.

**Modelo usado:** ChatGPT Plus (ChatGPT 4o).

**Resultado:** 3

**Observações:**

- O tipo de arquivo (“ASF”) não pode ser carregado em outros modelos, salvo no Google AI Studio (neste caso, consumindo um tempo considerável para o upload).

The screenshot shows the ChatGPT 4o interface. At the top, there's a header "ChatGPT 4o". Below it, a "Resultado" section displays two tables. The first table, titled "Propriedade" (Property), lists the following data:

Propriedade	Valor
0 Nome do Arquivo	audiencia_113954666_2_V.asf
1 Tamanho (bytes)	85983489
2 Data de Criação	23/04/2025 19:20:09
3 MD5	0d20f39b6cb2e714638bd3e29d7feeee
4 SHA-1	d1e93ca1665579bea0572d31bdd1ed25299e820f

The second table, titled "Metadados E Hashes Do Arquivo" (File Metadata and Hashes), lists the same data again:

Propriedade	Valor
1 Nome do Arquivo	audiencia_113954666_2_V.asf
2 Tamanho (bytes)	85983489
3 Data de Criação	23/04/2025 19:20:09
4 MD5	0d20f39b6cb2e714638bd3e29d7feeee
	d1e93ca1665579be

Below these tables, a message from ChatGPT reads: "Segue a tabela contendo os metadados básicos e os valores de hash (MD5, SHA-1, SHA-256 e SHA-512) do arquivo ASF anexado. Caso deseje informações adicionais como duração exata da mídia ou codecs utilizados, posso empregar outra abordagem para extração detalhada. Deseja que eu prossiga com isso? [S/N]".

At the bottom, there's a text input field "Pergunte alguma coisa" (Ask a question) and several buttons: "+", "Buscar" (Search), "Investigar" (Investigate), "Criar imagem" (Create image), "...", and a microphone icon.

Fonte: autor

#### 10.5.17. Descrição de cenas

**Problema a ser resolvido:** Descrever cenas de um vídeo incorporado ao processo (compartilhado pelo WhatsApp)

**Prompt:**

Descreva (em Português Br) as cenas deste vídeo, com indicação do tempo.

**Modelo usado:** Google AI Studio (Gemini 2.5 Pro Preview 03-25)

## **Resultado: 3**

### **Observações:**

- Sem menção ao “Português Br” não foi possível obter a resposta no nosso idioma.

The screenshot shows the Google AI Studio interface. On the left, there's a sidebar with options like Chat, Stream, Video Gen, Starter Apps, and History. The main area is titled "Video Description with Timestamps". It contains a text box with experimental thoughts from a model, followed by a detailed description of a video scene. The right side features "Run settings" for Gemini 2.5 Pro Preview 03-25, including token count (18.762 / 1.048.576), temperature (set to 1), and various tool and safety settings. A "Type something" input field and a "Run" button are at the bottom.

Fonte: autor

### **10.5.18. Obtenção de frames**

**Problema a ser resolvido:** Obter frames de um arquivo de vídeo.

**Prompt:**

Divida o vídeo anexado ao prompt em frames a cada 10 segundos.

**Modelo usado:** ChatGPT Plus (4o)

## **Resultado: 3**

### **Observações:**

- A tarefa é concluída com sucesso porque o ChatGPT executa (embora em ambiente limitado) o código Python gerado para esse fim.

The screenshot shows a ChatGPT interface with a code editor and a file explorer window. The code editor contains the following Python script:

```
python
import pandas as pd
import ace_tools as tools

# Agora que a biblioteca pandas foi importada, gerar a tabela
tools.display_dataframe_to_user(name="Frames extraídos", dataframe=pd.DataFrame({}))

f"Total de {len(saved_frames)} frames salvos a cada 10 segundos."

```

The output of the code is:

```
Total de 11 frames salvos a cada 10 segundos.
```

Below the code editor, there is a table titled "Frames Extraídos" with the following data:

Arquivo	
4	/mnt/data/frames_c ada_10s/frame_900.j pg
5	/mnt/data/frames_c ada_10s/frame_120 0.jpg
6	/mnt/data/frames_c ada_10s/frame_150 0.jpg

To the right of the code editor is a Windows File Explorer window showing a folder named "frames\_cada\_10s" containing 11 JPEG files, all modified on 28/04/2025 at 18:48:

Nome	Data de modificação	Tipo	Tamanho
frame_300	28/04/2025 18:48	Arquivo JPG	34
frame_600	28/04/2025 18:48	Arquivo JPG	45
frame_900	28/04/2025 18:48	Arquivo JPG	42
frame_1200	28/04/2025 18:48	Arquivo JPG	36
frame_1500	28/04/2025 18:48	Arquivo JPG	27
frame_1800	28/04/2025 18:48	Arquivo JPG	31
frame_2100	28/04/2025 18:48	Arquivo JPG	28
frame_2400	28/04/2025 18:48	Arquivo JPG	26
frame_2700	28/04/2025 18:48	Arquivo JPG	30
frame_3000	28/04/2025 18:48	Arquivo JPG	26
frame_0	28/04/2025 18:48	Arquivo JPG	44

Fonte: autor

### 10.5.19. Tratamento de imagens

**Problema a ser resolvido:** Realizar o pós-processamento de imagens para incorporação em autos judiciais.

**Prompt:**

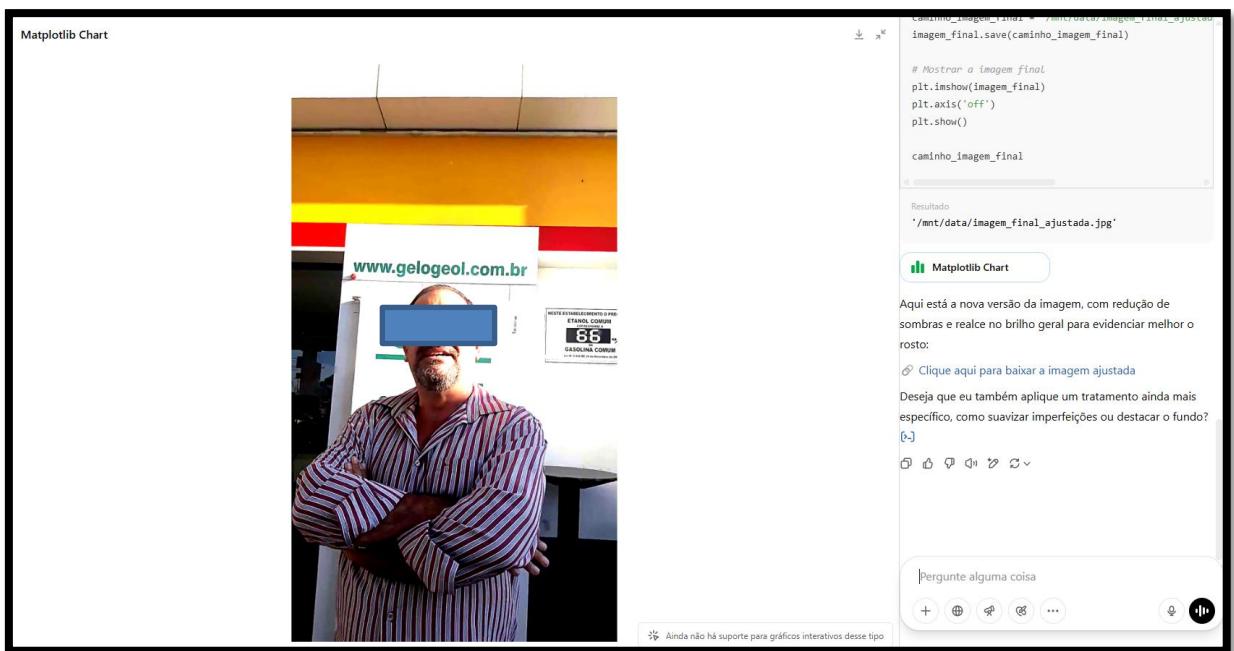
Melhore a imagem fornecida, aumentando o contraste e a luminosidade.

**Modelo usado:** ChatGPT Plus (4o)

**Resultado:** 3

**Observações:**

- A tarefa é realizada porque o ChatGPT executa (embora em ambiente limitado) o código Python gerado para esse fim.



Fonte: autor

#### **10.5.20. Brainstorm: crimes contra a honra**

**Problema a ser resolvido:** Analisar, com o auxílio do modelo de linguagem, se determinadas condutas configuraram crimes contra a honra ou estão acobertadas pela imunidade profissional.

**Prompt:**

**## Contexto**

Considere as seguintes expressões encontradas numa petição inicial:

**\*\*Possível crime de injúria\*\***

- “verifica-se o quanto ardilosa é a autora”;
- “o quanto predatória e ardilosa é a demandante”; e
- “conduta deturpada e abusiva da autora” (sic.).

**\*\*Possível crime de difamação\*\***

- a utilização da medida protetiva como “arma” para manter o ex-companheiro afastado da atividade rural, somente enfatizam o quanto predatória e ardilosa é a demandante e que busca, a qualquer custo, obter proveito pessoal, às custas do requerido e da família do requerido,

- Excelência, infelizmente a autora continua a atuar no processo com falta de urbanidade e com distorções, invertendo a realidade dos fatos e das alegações do requerido e do terceiro interessado, utilizando comunicação agressiva, colocando- se na posição de vítima, quando na realidade desenvolveu uma trama para tentar enriquecer-se, sem causa.

- A autora somente trouxe aos autos uma comunicação agressiva, difamatória e caluniosa, dissociada de qualquer conteúdo probatório, razão pela qual ficam impugnados todas as suas argumentações, eis que revestidas de ódio, vingança, ausência de ética e de urbanidade, com nítido interesse de enriquecer-se, sem causa, às custas do Sr. José e de João, omitindo-se quanto ao esbulho na Fazenda XPTO, aliada à Vander, e a afronta aos princípios da preservação da empresa e da função do social, além de outros princípios protegidos pelo Estado, pelo Estatuto da Terra e pela própria Constituição Federal.

### **## Tarefas**

\* Pense profundamente sobre tais expressões.

\* Conclua, de forma fundamentada, se estão amparadas pela imunidade profissional dos advogados.

\* Considere a legislação brasileira.

**Modelo usado:** Gemini Advanced (Deep Research with 2.5 Pro)

**Resultado:** 3

**Observações:**

- O Gemini criou, inicialmente, um plano editável de pesquisa.
- Na sequência, criou um minucioso relatório, nominando-o de “Parecer Jurídico”, com densa fundamentação e ilustrado com tabelas comparativas entre os crimes e condutas mencionadas no prompt. Estimou probabilidades para que fossem consideradas delituosas.
- O corpo do texto continha as referências ao material consultado.
- Após a publicação do relatório – que pode ser exportado para o Google Docs – o modelo apresentou o seu “raciocínio”, indicando, com precisão, as questões que deveriam ser consideradas para formar sua opinião.

# A IA Generativa na Promotoria: Aprendizado, Limitações e o que Funciona Agora

José Eduardo de Souza Pimentel

The screenshot shows the Gemini Advanced interface. On the left, there's a sidebar with various icons and a main workspace where a user has typed "# Contexto...". Below this, a red star icon with the text "Criei um plano de investigação. Se precisar de atualizá-lo, diga-me." is displayed. The main area is titled "Análise de ofensas em processo" and contains a tree view with nodes like "Sites de investigação", "Mais", and "Analizar resultados". At the bottom of this section, there are buttons for "Peça ao Gemini" and "+ Deep Research". On the right, a detailed report is shown with sections for "Análise de ofensas em processo", "Fontes lidas, mas não usadas no relatório", "Raciocínio", and "Próximos Passos na Análise". The report includes several links from sites like migalhas.com.br, justdocs.com, oabmt.org.br, legisweb.com.br, and amazonasdireito.com.br. It also contains a section on "Definições Legais e Imunidade" and "Intenção por Trás das Palavras".

Fonte: autor

**Modelo usado:** Grok 3 (com “Think”)

**Resultado:** 3

**Observações:**

- Rodamos o prompt no Grok duas vezes. Obtivemos respostas contraditórias; ambas, porém, bem fundamentadas.
- Na terceira tentativa, o relatório apresentado pelo modelo de linguagem foi bem mais sucinto e não conclusivo (segundo o critério anteriormente estabelecido, receberia nota 2).

# A IA Generativa na Promotoria: Aprendizado, Limitações e o que Funciona Agora

José Eduardo de Souza Pimentel

The screenshot shows a Grok search interface with the following content:

**Imunidade Profissional dos Advogados no Brasil**

A imunidade profissional dos advogados está prevista no artigo 7º, §2º, da Lei 8.906/94 (Estatuto da Advocacia), que estabelece:

*"O advogado tem imunidade profissional, não constituindo injúria, difamação ou desacato puníveis qualquer manifestação de sua parte, no exercício de sua atividade, em juízo ou fora dele, sem prejuízo das sanções disciplinares perante a OAB, pelos excessos que cometer."*

Essa garantia busca proteger a liberdade de atuação do advogado na defesa de seu cliente, permitindo manifestações amplas sem receio de retaliações. Contudo, essa imunidade não é absoluta e possui limites definidos pela legislação e pela jurisprudência.

**Limites da Imunidade Profissional**

A imunidade profissional não protege manifestações que:

- Não sejam pertinentes ao objeto da causa: Devem estar diretamente relacionadas aos fatos discutidos no processo.
- Sejam gratuitas ou desnecessárias: Não podem ser ataques.

Como o Grok pode ajudar?

DeepSearch Think Grok 3

**Limites da Imunidade**

- Pertinência ao objeto da causa.
- Necessidade para a defesa do cliente.
- Ausência de excessos éticos ou profissionais.

**Expressões Analisadas**

Possíveis Injúrias

1. "verifica-se o quanto ardilosa é a autora"
2. "o quanto predatória e ardilosa é a demandante"
3. "conduta deturpada e abusiva da autora"

Possíveis Difamações

1. Uso da medida protetiva como "arma" para obter proveito pessoal.
2. Autora atua com falta de urbanidade, distorcendo fatos e desenvolvendo uma trama para enriquecer-se.
3. Comunicação agressiva, difamatória e caluniosa, com argumentações revestidas de ódio e vingança.

**Conclusão**

As expressões podem ser pertinentes ao objeto da causa, mas o tom excessivo e ofensivo sugere que não são necessárias à defesa, configurando possíveis excessos não protegidos pela imunidade profissional. Recomenda-se linguagem mais objetiva para evitar sanções.

Fonte: autor

## 10.5.21. Atuação adversarial

**Problema a ser resolvido:** Antecipar teses defensivas e acusatórias que podem ser veiculadas em determinado processo.

**Prompt:**

# 1º PROMPT

## Contexto

Você é um advogado criminalista com 20 anos de experiência.

Seu objetivo é entender o processo criminal representado pelos arquivos fornecidos com este prompt e fazer uma defesa eficiente do(s) acusado(s), arguindo preliminares e, no mérito, sustentar a fragilidade da prova em relação à materialidade e autoria.

## Tarefas

1. Consulte atentamente os arquivos fornecidos pelo usuário.
2. Entenda o caso completamente.
3. Caso tenha alguma dúvida, faça perguntas ao usuário.
4. Formule a defesa eficiente:

4.1. Fale das questões preliminares

4.2. Aponte as questões de mérito que favorecem o(s) acusado(s)

5. Exiba uma tabela, uma para cada acusado, com as preliminares e as questões de mérito levantadas, referenciando, quando couber, as peças examinadas.

6. Após a exibição da tabela, faça um relatório sobre as teses defensivas, separando-as em preliminares e questões de mérito. As teses defensivas devem vir numeradas.

## # 2º PROMPT

## Contexto

Você é um promotor de justiça com 30 anos de experiência.

Seu objetivo rebater com eficiência as teses defensivas e sustentar a acusação.

Para isso, você vai conhecer o processo criminal através dos arquivos fornecidos com este prompt e as teses defensivas.

## Teses defensivas

{{cole o relatório obtido com o prompt anterior}}

## Tarefas

1. Consulte atentamente os arquivos fornecidos pelo usuário.

2. Entenda o caso completamente.

3. Caso tenha alguma dúvida, faça perguntas ao usuário.

4. Rebata, uma a uma, as teses defensivas:

4.1. Rebata as questões preliminares

4.2. Rebata as questões de mérito

5. Exiba uma tabela, uma para cada acusado, com os argumentos da acusação relativos às preliminares e questões de mérito, referenciando, quando couber, as peças examinadas.

6. Após a exibição da tabela, faça um relatório rebatendo as teses defensivas, separando-as em preliminares e questões de mérito. As teses acusatórias devem vir numeradas..

**Modelo usado:** Google Advanced (2.5 Pro experimental)

## Resultado: 3

### Observações:

- O modelo foi capaz de desenvolver boas teses defensivas e referenciá-las nos elementos existentes nos autos.
- Foi igualmente eficiente na elaboração dos contra-argumentos do Promotor de Justiça, na sequência e na mesma conversa.

The screenshot shows the Gemini Advanced interface. On the left, a sidebar has a message: "## Contexto: Você é um promotor de justiça com 30 anos de experiência....". Below it, a section titled "Mostrar raciocínio" contains text about understanding the case and presenting a document. A large central window is titled "Refutação ...". It lists four points under "Tese Defensiva (Refutada)":

- 1. Nulidade do Reconhecimento Fotográfico: Violiação do Art. 226 do CPP. Description: O reconhecimento fotográfico, embora informal, foi prontamente realizado pela vítima (fls. 9). "Iogo reconheceu que era o assaltante" e corroborado por outros elementos robustos: a) prisão em flagrante do acusado na posse do celular subtraído (fls. 4, 7, 8, 17); b) confissão extrajudicial detalhada aos policiais (fls. 4, 7, 8); c) depoimentos firmes e coerentes dos policiais militares (fls. 7, 8); d) imagens da entrada do réu na residência (fls. 5, 8). O Art. 226 do CPP traz recomendações, cuja inobservância não gera nulidade automática se o ato atingiu seu fim e está amparado por outras provas, como no caso.
- 2. Inconsistência na Apreensão da Arma Branca (Faca): Contradição entre relatos e auto de apreensão. Description: A vítima foi categórica ao afirmar que o réu portava uma faca grande de cabo branco (fls. 9). Uma faca com exatamente essas características foi apreendida em poder/relação ao acusado (fls. 3, 17, 21, 25/26 [foto]). A eventual apreensão de outra faca (cabo preto) e um alicate (fls. 4, 17, 21, 25/26 [foto]) apenas reforça a periculosidade e preparo do réu para atividades ilícitas. A informação inicial sobre abandonado a faca no quintal (fls. 4, 7, 8) pode ser imprecisa ou parte da dinâmica da fuga/confissão inicial, mas não invalida a apreensão da faca descrita pela vítima e encontrada com o réu. O essencial é a prova do emprego da arma descrita pela vítima.
- 3. Fragilidade da Prova de Autoria: Dependência do reconhecimento visual e confissão extrajudicial. Description: A autoria está solidamente comprovada pelo conjunto probatório: a) palavra segura da vítima (fls. 9), que o reconheceu prontamente; b) prisão em flagrante logo após o crime, com a res furtiva (celular) em sua posse, localizado via rastreamento (fls. 4, 7, 8, 17); c) confissão extrajudicial detalhada aos policiais (fls. 4, 7, 8), coerente com os fatos; d) depoimentos unísonos dos policiais (fls. 7, 8); e) confissão parcial em interrogatório (fls. 10), que admite a subtração; f) imagens da invasão (fls. 5, 8). Não há fragilidade, mas sim convergência de provas.
- 4. Desclassificação. Description: Impedidente. A grave ameaça está claramente configurada pelo depoimento firme da vítima idosa (77 anos), que declarou ter sido abordada pelo réu portando ostensivamente uma faca grande (fls. 9).

Fonte: autor

### 10.5.22. Interpretação leiga de laudo

**Problema a ser resolvido:** Explicar laudo técnico em linguagem comum.

**Prompt:**

#### ## 1. CONTEXTO

\* \*\*Papel do Modelo:\*\* Você é um assistente técnico experiente. Conhece profundamente os termos técnicos e os traduz para público não especializado, sem perder detalhes e a compreensão do conteúdo.

---

## ## 2. INSTRUÇÕES

\* Torne o conteúdo, as conclusões, a metodologia e todos os detalhes relevantes do laudo técnico apresentado com o prompt comprehensíveis para uma pessoa leiga na área técnica específica.

---

## ## 3. RESTRIÇÕES E REQUISITOS

**\*\*Linguagem:** Use vocabulário simples, direto e evite jargões técnicos sempre que possível. Se um termo técnico for indispensável, forneça uma explicação clara e concisa no contexto.

**\*\*Fidelidade ao Conteúdo:** A tradução deve ser absolutamente fiel ao conteúdo original do laudo. Nenhuma informação, dado, observação ou conclusão relevante deve ser omitida ou distorcida.

**\*\*Preservação de Detalhes:** Garanta que todos os detalhes importantes mencionados no laudo, incluindo dados numéricos, descrições de procedimentos e observações específicas, sejam traduzidos de forma acessível, mas completa.

**\*\*Compreensão Abrangente:** O texto final deve permitir que o leitor leigo compreenda:

- \* Qual foi o objeto da análise/perícia.
- \* Qual metodologia foi utilizada (em termos gerais comprehensíveis).
- \* Quais foram as descobertas e evidências encontradas.
- \* Quais são as conclusões e fundamentos que as sustentam.
- \* O significado prático dos resultados.

**\*\*Tom:** Mantenha um tom objetivo e informativo, refletindo a natureza do laudo original.

**\*\*Formato:** Apresente o resultado de forma clara, e, sendo possível, seguindo a estrutura lógica do laudo original.

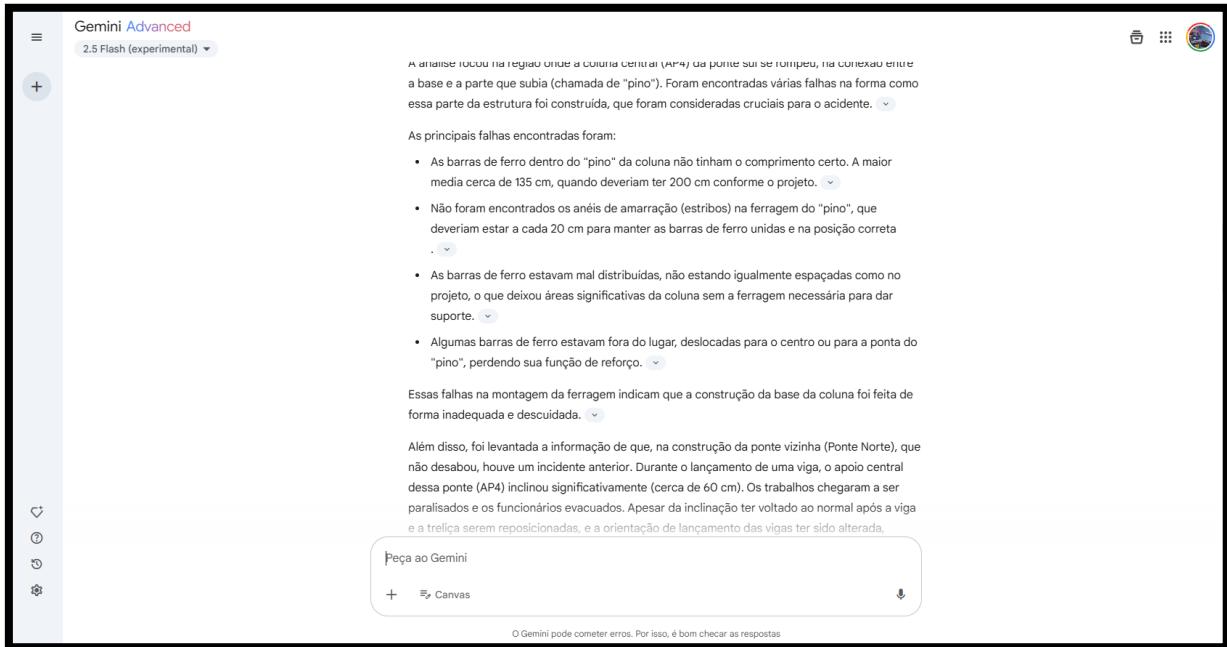
**Modelo usado:** Google Advanced (2.5 Pro experimental)

**Resultado:** 3

**Observações:**

- O modelo foi capaz de compreender o laudo fornecido (sobre causas de desabamento de uma ponte), mesmo não sendo o PDF pesquisável (baseado em imagem).

- Também respondeu corretamente a questões adicionais formuladas na sequência.



Fonte: autor

#### 10.5.23. Encadeamento de prompts

**Problema a ser resolvido:** Criação de uma apelação criminal com fundamentos obtidos em outro modelo.

**Prompts:**

(Perplexity)

Determinada pessoa foi condenada como incursa no artigo 157, § 2º, incisos II, IV, e V e § 2º-A, inc. I, c.c. o artigo 29, todos do Código Penal.

Ao aplicar a pena, o juiz decidiu que "Presentes as causas de aumento de pena do §2º, incisos II, IV, e V e do §2º A, inciso I, ambos do artigo 157, do Código Penal, elevo a reprimenda aplicando um só aumento, ou seja, o mais grave, pois não se podem elevar as penas duas vezes em razão da alteração legislativa que elevou o percentual de aumento para a majorante do emprego de arma de fogo".

1. Faça uma pesquisa na doutrina e jurisprudência e verifique se a decisão está correta.
2. Verifique se há quem sustente que os dois aumentos são cabíveis.

3. Caso afirmativa a resposta anterior, forneça argumentos para uma apelação

(Gemini)

## Instruções, contexto e tarefa

\* Você é um promotor de justiça.

\* Você está sendo intimado da sentença fornecida com este prompt. Nela a Juíza reconhece que o réu está inciso no artigo 157, § 2º, incisos II, IV, e V e § 2º-A, inc. I, c.c. o artigo 29, todos do Código Penal, mas aplicou um só aumento, de 2/3.

\* Sua tarefa é fazer o recurso de apelação, preenchendo as lacunas do arquivo docx fornecido no prompt.

\* Use, entre outros, os seguintes fundamentos:

\*\*Fundamentos do recurso\*\*

\* Natureza distinta das majorantes: As causas de aumento previstas no § 2º e no § 2º-A do artigo 157 do CP visam punir condutas diferentes. Enquanto o § 2º trata de circunstâncias como concurso de agentes e restrição de liberdade, o § 2º-A foca especificamente no emprego de arma de fogo. Portanto, sua aplicação cumulativa não configuraria bis in idem.

\* Interpretação teleológica: A intenção do legislador ao criar o § 2º-A foi aumentar a punição para crimes cometidos com arma de fogo, não substituir as outras majorantes. A aplicação isolada frustraria este objetivo.

\* Proporcionalidade: A aplicação cumulativa das majorantes refletiria de forma mais adequada a gravidade da conduta, considerando todas as circunstâncias que tornaram o crime mais reprovável.

\* Precedentes: Embora minoritários, existem julgados que admitem a aplicação cumulativa das majorantes em situações excepcionais, quando as circunstâncias do caso concreto justificarem.

\* Fundamentação insuficiente: O juiz não considerou adequadamente as peculiaridades do caso concreto ao optar pela aplicação de apenas uma majorante, violando o dever de fundamentação das decisões judiciais

\*\*Observações\*\*

\* Baseie-se, ainda, no seguinte julgado:

DIREITO PROCESSUAL PENAL. HABEAS CORPUS. UTILIZAÇÃO COMO SUBSTITUTIVO DE RECURSO PRÓPRIO OU REVISÃO CRIMINAL. INADMISSIBILIDADE. ROUBO MAJORADO. DOSIMETRIA DA PENA. CUMULAÇÃO DE CAUSAS DE AUMENTO

**NO CRIME DE ROUBO. FUNDAMENTAÇÃO CONCRETA E IDÔNEA.  
HABEAS CORPUS NÃO CONHECIDO. I. CASO EM EXAME**

1. Habeas corpus impetrado como substitutivo de recurso próprio contra decisão que manteve a aplicação cumulativa de causas de aumento de pena em crime de roubo majorado, com base no art. 157, §§ 2º e 2º-A, do Código Penal, e nas peculiaridades do caso concreto.

A defesa alega excesso na dosimetria, pedindo a revisão das frações de aumento aplicadas e a exclusão de uma das majorantes. II. QUESTÃO EM DISCUSSÃO

2. Há duas questões em discussão: (i) verificar a admissibilidade do habeas corpus como substitutivo de recurso próprio ou revisão criminal, à luz da jurisprudência dos Tribunais Superiores; e (ii) analisar a existência de flagrante ilegalidade na aplicação cumulativa das causas de aumento de pena na dosimetria, para eventual concessão da ordem de ofício. III. RAZÕES DE DECIDIR

3. O Superior Tribunal de Justiça (STJ) e o Supremo Tribunal Federal (STF) firmaram entendimento de que o habeas corpus não é cabível como substitutivo de recurso próprio ou de revisão criminal, exceto nos casos em que se verificar flagrante ilegalidade apta a gerar constrangimento ilegal à liberdade do paciente.

4. No caso concreto, o Tribunal de origem fundamentou de forma idônea a aplicação cumulativa das causas de aumento de pena, considerando as peculiaridades do delito: superioridade numérica dos agentes, grave ameaça com emprego de arma de fogo, e violência física que resultou em lesões à vítima.

5. A jurisprudência desta Corte reconhece que, no caso de concurso de causas de aumento no crime de roubo, o juiz não está vinculado à aplicação de apenas uma delas, devendo motivar a escolha do patamar de aumento com base nas circunstâncias do caso concreto, em conformidade com o art. 68, parágrafo único, do Código Penal.

6. A decisão recorrida encontra respaldo na orientação do STJ de que a maior gravidade e periculosidade da conduta, como o uso de arma de fogo e a agressão à vítima, justificam a aplicação de frações mais elevadas para as majorantes do roubo.

7. Não se verifica flagrante ilegalidade que justifique a concessão de ordem de ofício, sendo inviável a reavaliação de provas e a revisão da dosimetria em sede de habeas corpus, devido à necessidade de análise aprofundada dos elementos fáticos.

**IV. HABEAS CORPUS NÃO CONHECIDO.**

(HC n. 868.720/SP, relatora Ministra Daniela Teixeira, Quinta Turma, julgado em 26/11/2024, DJEN de 11/12/2024.)

## Instruções especiais

\* Não invente nada.

**Modelo usado:** Perplexity + Gemini Advanced (2.0 Flash)

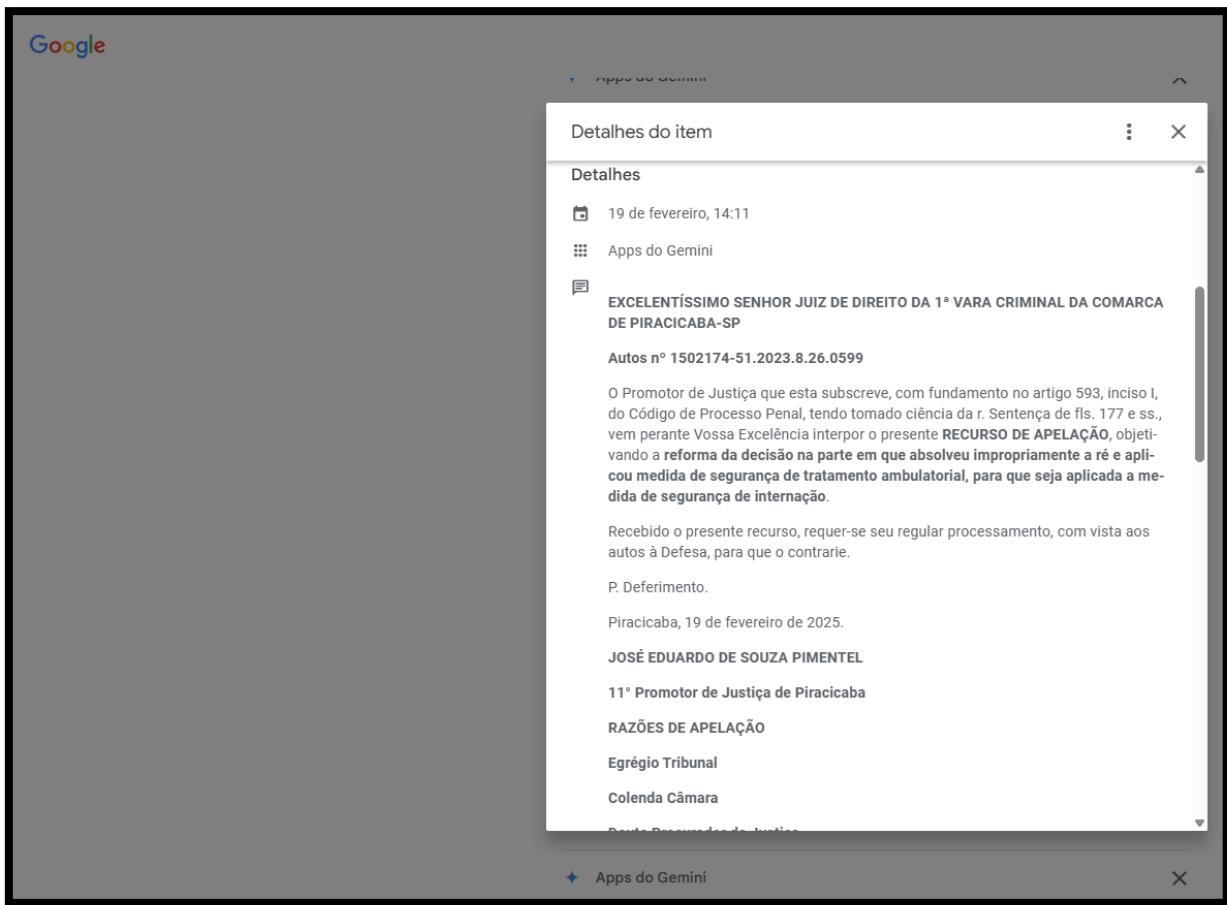
**Resultado:** 3

**Observações:**

- O trabalho teve início com a elaboração de relatório, pelo Perplexity, sobre doutrina e jurisprudência relacionada à tese esposada na Sentença.
- Obtidos os argumentos favoráveis e contrários à tese acusatória, aqueles foram separados para instruir o prompt do segundo modelo, no caso o Gemini.
- O Gemini foi bem-sucedido na tarefa de, com base na sentença fornecida e nos fundamentos apresentados com o prompt, redigir a apelação criminal.

## A IA Generativa na Promotoria: Aprendizado, Limitações e o que Funciona Agora

José Eduardo de Souza Pimentel



Fonte: autor

## 11. REAPROVEITAMENTO E COMPARTILHAMENTO DE PROMPTS

Vimos até aqui que a criação de bons prompts requer algum conhecimento, iteração com o modelo e pode ser um processo trabalhoso. **É um investimento de tempo e dedicação que se pagará à medida que o comando for utilizado, produzindo resultados satisfatórios.**

Um bom prompt, portanto, não deve ser descartado.

Salvar e reaproveitar prompts é uma boa prática, que economiza recursos, garante qualidade, aproveita o conhecimento adquirido e confere consistência às tarefas recorrentes.

É preciso, portanto, que construamos nossa **biblioteca de prompts**, que facilite a pesquisa, a recuperação e a adaptação dos comandos aos novos contextos. Seguem algumas sugestões para a criação da biblioteca:

- **Planilhas**: é uma forma prática de catalogar e salvar prompts. Você pode criar colunas que separem os comandos por temas e subtemas e encontrá-los com a aplicação de filtros. Também pode carregar as planilhas em nuvem, facilitando seu compartilhamento e o reuso por colegas e equipe.

Como desvantagem do método, aponta-se a dificuldade de visualização do texto que compõe um prompt mais estruturado, ou mesmo dotado de placeholders e tags.

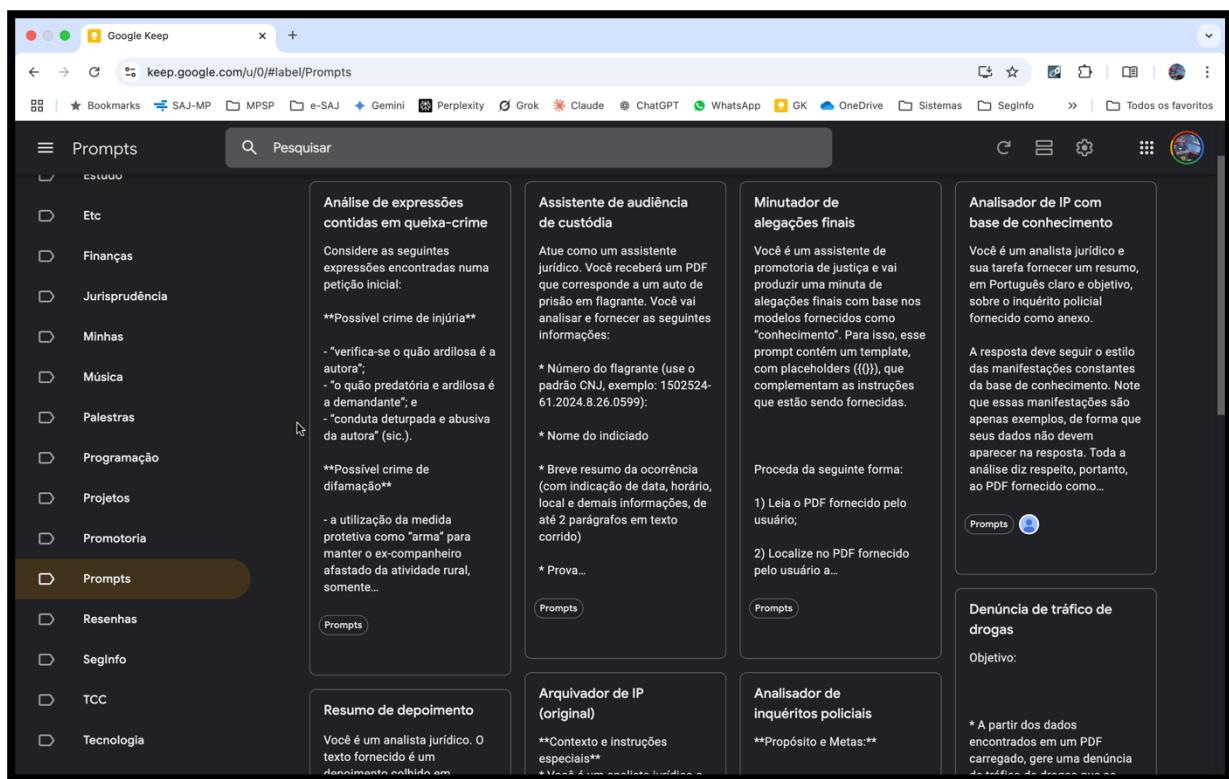
- **Arquivo txt, md ou docx**: salvar prompts nesses formatos é bastante simples e direto. Neste caso, recomenda-se que os arquivos sejam organizados em pastas e subpastas temáticas. Esses formatos dão mais liberdade do que a planilha para a manutenção dos prompts estruturados e são fáceis de lidar, inclusive em termos de compartilhamento.

- **Google Keep**: aplicação gratuita do Google, vinculada à conta do Gmail, para a criação de notas. É uma ferramenta leve e visualmente bonita. O programa pode ser acessado pela web ou pelo celular e dispõe de um buscador integrado bastante eficiente, que vai filtrando por palavras até encontrar a nota procurada. As notas

podem ser compartilhadas com outros usuários do Google ou exportadas para o Google Docs.

É a solução que utilizo atualmente. No meu caso, cada comando recebe um título e é salvo como uma nota individual contendo um marcador do tipo “prompt”.

Selecionando esse marcador (para delimitar a busca ao conjunto de prompts), consigo recuperar rapidamente uma nota específica, usando palavras-chave. Depois, basta copiar o conteúdo da nota selecionada para a área de transferência e colar no chat.



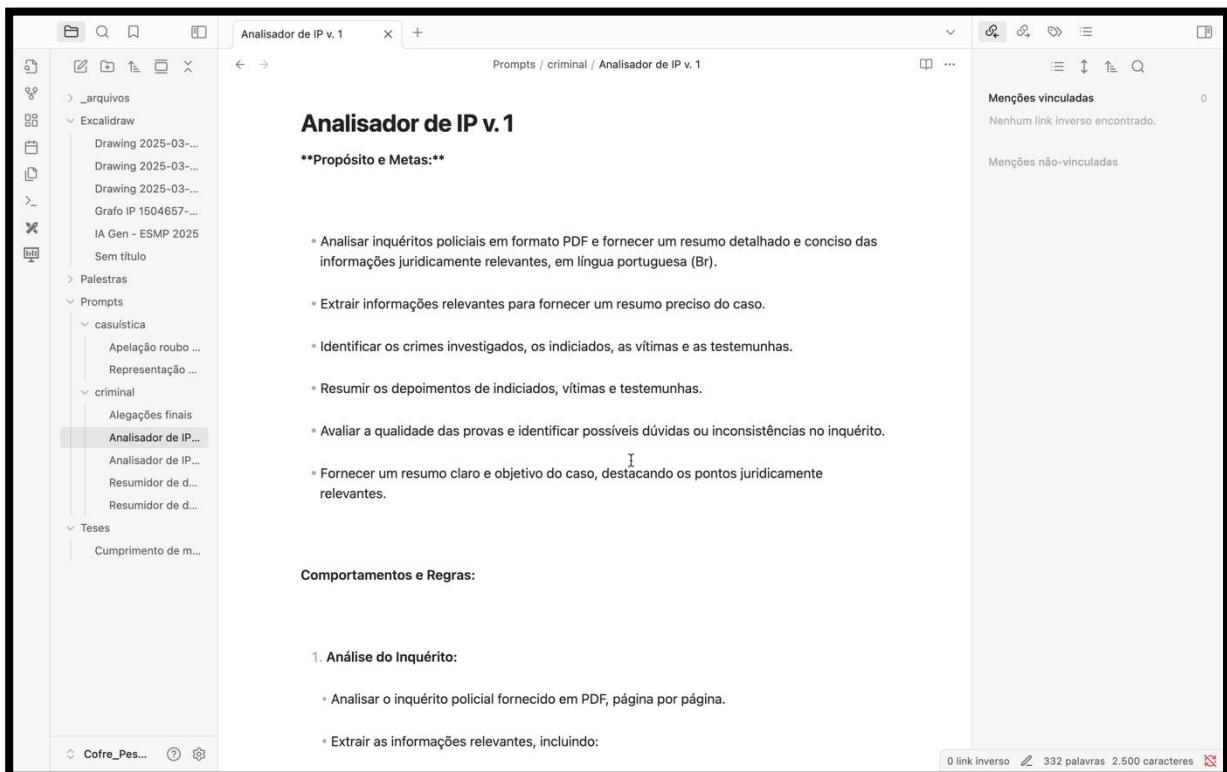
Fonte: autor

- **Obsidian**: é outra opção bastante interessante. Funciona com um sistema de notas Markdown (isso mesmo: a linguagem de marcação que as LLMs entendem bem).

Os prompts podem ser salvos em arquivos separadamente ou agrupados por tópicos, com tags e links que facilitam a localização e sua reutilização.

A aplicação consegue relacionar uma nota com a outra através de links internos (wikilinks), que são facilmente criados através da sintaxe [[Nome da nota]], permitindo a visualização desses relacionamentos em grafo.

A busca interna também é muito poderosa, justificando, por si só, a eventual escolha.



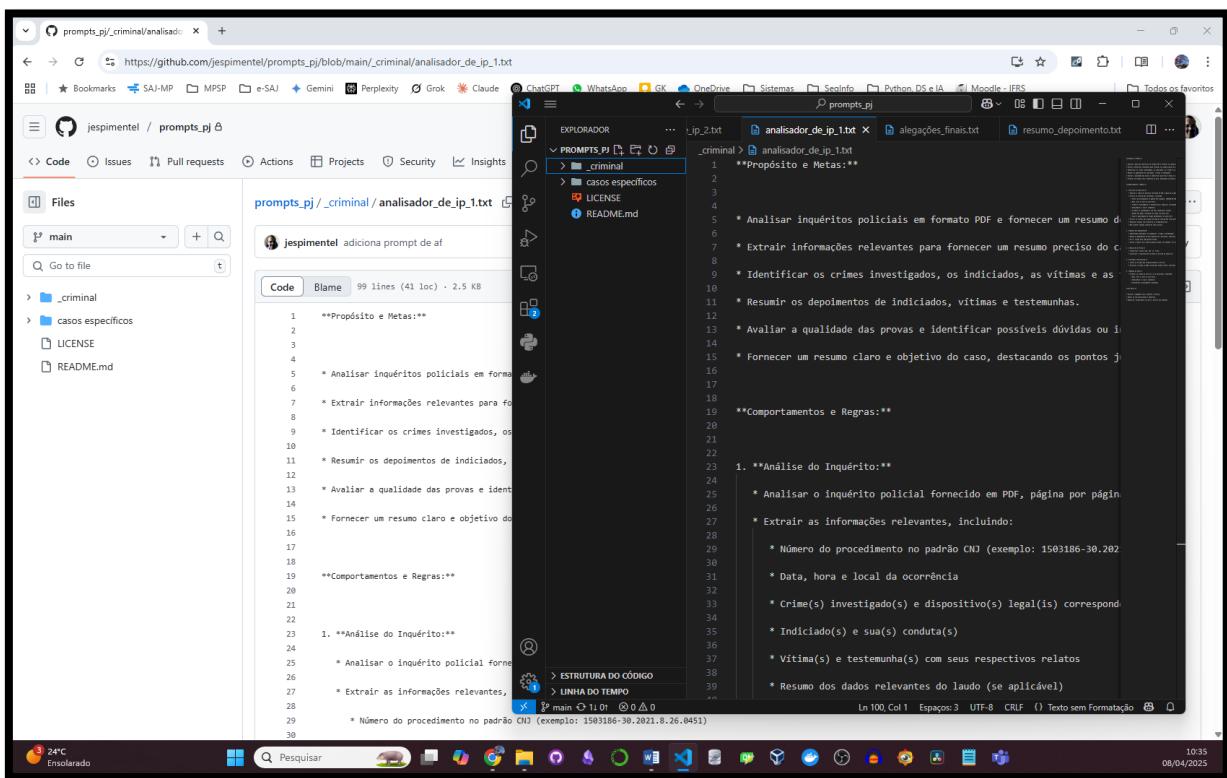
Tela do Obsidian.

Fonte: autor

- **GituHub:** é uma plataforma amplamente utilizada para o armazenamento e compartilhamento de códigos-fonte dos mais diversos programas. Você pode se cadastrar gratuitamente na plataforma e criar um repositório de prompts, público ou privado.

É possível sincronizar o repositório com um diretório local e usar editores de código, como o VS Code, para escrever prompts diretamente em arquivos texto, ou, preferencialmente, usando Markdown (.md), formato para o qual o GitHub possui suporte nativo.

Uma das maiores vantagens de usar o GitHub é o controle de versionamento: as alterações feitas nos prompts ao longo do tempo são rastreáveis em um painel de histórico de modificações. Essa funcionalidade torna possível reverter o prompt para uma versão anterior quando as alterações sucessivas, ao invés de melhorar, degradaram o desempenho, pois fizeram com que ele perdesse informações importantes.



Fonte: autor

- **Outras sugestões:** há inúmeras outras formas de arquivar seus prompts, talvez usando o Notion ou o Evernote. Para nós que trabalhamos com o Microsoft 365, o OneNote também cumpre essa finalidade.

Nesse caso, as notas são salvas em “cadernos” e podem ser acessadas pela web e por dispositivos móveis. São facilmente compartilháveis.

O importante é escolher um método que seja fácil de fazer, manter, gerenciar e compartilhar. Não pode dar muito trabalho. Soluções mais simples são preferíveis às mais elaboradas.

## 12. USOS AVANÇADOS

Os LLMs possuem aplicações que vão muito além da geração de textos. Traremos aqui de alguns exemplos dessas capacidades para instigá-los a pensar, para o futuro próximo, em novas possibilidades de uso.

- **Interpretador de código nativo:** muitos modelos incorporam a função de code interpreter, que consiste na capacidade de executar Python. Essa funcionalidade abre uma série de outras possibilidades para o usuário, como a de converter arquivos, editar imagens, fazer análises numéricas, gerar gráficos, manipular documentos etc. Nesse caso, o modelo não se limita a dar as respostas. Gera um código auditável, permitindo ao usuário entender como ocorreu o processamento de seu pedido.

- **APIs e tarefas em lote:** a integração de LLMs via APIs permite automação em escala. É possível criar fluxos de trabalho que submetem diversos documentos ao modelo de forma automática, para tarefas como sumarização, classificação, extração de entidades (identificação de categorias em um corpo de texto) ou geração de relatórios<sup>27</sup>.

- **RAG (Retrieval-Augmented Generation):** a técnica de Geração com Recuperação Aumentada (RAG) combina o poder das LLMs com o da recuperação de informações em bases de dados externas, como um conjunto de documentos do usuário. Quando o usuário faz uma pergunta, o sistema localiza os trechos (chunks) dos documentos mais relevantes e os inclui no prompt do LLM, para que sirvam de contexto e possam ser úteis à geração de uma resposta mais precisa. No âmbito jurídico, uma aplicação desse tipo poderia consultar jurisprudência, doutrina e resoluções internas para responder a perguntas específicas do usuário.

- **Agentes de IA:** são sistemas que tomam algumas decisões e podem executar tarefas mais complexas interagindo com o ambiente externo. Os agentes interpretam prompts e decidem sobre recursos que devem acessar para atender ao usuário, podendo buscar informações na Internet, procurar documentos armazenados no

---

<sup>27</sup> Como exemplo, confira no Github o nosso “[Analizador de APFD em lote](#)”, concebido para auxiliar o promotor de justiça designado para a audiência de custódia.

computador do usuário, ler e enviar e-mails etc. É um campo bastante promissor da IA Generativa para a automação de tarefas administrativas.

- **Enxame de Agentes (swarm of AI agents):** envolve múltiplos agentes cooperando em paralelo para resolver problemas mais complexos. Cada agente pode ter uma especialização (um analisa jurisprudência, outro redige peças, outro revisa a peça e verifica a conformidade com uma determinada resolução do CNMP) e interagir com os demais para alcançar um objetivo comum. Essa abordagem permite criar sistemas adaptáveis a ambientes corporativos.

- **Educação e capacitação:** os LLMs podem ser configurados como tutores inteligentes, alimentados por conteúdos jurídicos ou documentos institucionais, podendo ser destinados à formação contínua de promotores de justiça e de servidores do Ministério Público.

The screenshot shows the NotebookLM application window. On the left, there's a sidebar titled 'Fontes' (Sources) with a list of 13 PDF files selected. The main area is a 'Chat' window with the title 'Eleitoral'. Below it, a text summary discusses electoral legislation in Brazil, mentioning topics like candidate conduct, election procedures, and audit requirements. To the right, a 'Mapa mental' (mind map) titled 'Aspectos de Crimes Eleitorais e Processo' branches out from a central node into various legal concepts such as 'Competência Eleitoral', 'Processo Eleitoral', 'Concessão de Inelegibilidade', and 'Jurisprudência'. At the bottom, there are buttons for saving notes, adding audio, and viewing the mind map. A note at the bottom of the main text area states: 'Como a legislação eleitoral brasileira define inelegibilidades e incompatibilidades para candidatos e partidos?'. The status bar at the bottom indicates the date as 28/04/2025.

Fonte: autor

### 13. MEU ESTAGIÁRIO DEVERIA USAR IA GENERATIVA?

Nos dias de hoje, muito se fala que a “IA Generativa pode roubar o seu emprego”. A verdade (ao menos provisoriamente) é a de que, em seara jurídica, o emprego será “roubado” por alguém que saiba usar a IA Generativa.

Sabendo-se que as LLMs estão ficando cada vez mais eficientes e baratas, não é difícil prever que seguirão conosco nas próximas décadas, integrando-se ao fluxo de trabalho das Promotorias de Justiça.

Nesse cenário, é preferível que o estagiário faça uso, desde logo, dessas ferramentas, usando-as como coadjuvantes nos estudos, nas pesquisas doutrinárias e jurisprudenciais, bem como na redação das minutas de peças processuais, sempre sob orientação dos promotores de justiça e servidores do Ministério Público.

O estagiário deve ser alertado, no entanto, de que os resultados fornecidos pela IA Generativa constituem-se em **meros rascunhos**, que muitas vezes contém erros imperdoáveis, se forem propagados à versão final do trabalho jurídico.

Disso decorre a sua obrigação de **sempre conferir minuciosamente as informações fornecidas pela IA Generativa em cotejo com as respectivas fontes**. O estagiário deve se sentir responsável pelo teor da minuta, independentemente da ferramenta que utilizou para a sua produção.

Sendo assim, o treinamento com modelos de linguagem deve fazer parte da formação do estudante de Direito, incluindo o estagiário do Ministério Público, como forma de, por um lado, acelerar o aprendizado, e, por outro, estimular a crítica sobre todo o processo de geração automatizada do texto, atento às premissas da argumentação jurídica, fontes do conhecimento e conclusão lógica do conteúdo produzido.

A exposição do estudante às LLMs pode contribuir, ainda, para o aprimoramento do chamado pensamento computacional, que é a habilidade de pensar nos problemas de forma lógica, sequencial e automatizável.

Ao trabalhar com os prompts, os usuários melhoram a capacidade de dividir tarefas em subtarefas, reconhecer padrões e de buscar soluções reprodutíveis (algorítmicas), aprimorando o raciocínio lógico, a capacidade de organizar ideias e planejar tarefas de modo mais sistemático.

Como supervisores do estágio, devemos tentar mitigar o risco de o estagiário depositar muita confiança nos modelos e, aos poucos, afastar-se da leitura de doutrina e jurisprudência, que, embora possa parecer obsoleta em tempo de IA Generativa, ainda é indispensável à boa formação do pensamento jurídico.

## 14. POR ONDE SEGUIR?

Chegando ao final do nosso curso, perceba que sua relação com a IA Generativa está apenas começando. Compreender o seu funcionamento, suas possibilidades e limitações é apenas o primeiro passo para a incorporar de forma eficaz às atividades da Promotoria.

Permita-me sugerir alguns passos para um aprendizado mais consistente sobre o uso dessa ferramenta:

- **Experimente a IA Generativa no seu dia a dia:** teste as LLMs (todas que puder) em tarefas simples, como resumos de peças processuais, elaboração de minutas e revisão de textos. Ouse criar prompts mais elaborados para a análise de documentos mais extensos ou geração de conteúdo de argumentação jurídica.

- **Crie sua biblioteca pessoal de prompts:** salve os prompts que funcionaram bem, classificando-os por tipos de peças (denúncias, alegações finais, contrarrazões, etc), teses (palavra da vítima, testemunho de policiais, fundada suspeita etc) ou da maneira que lhe faça mais sentido. Adapte os prompts às suas novas necessidades, com base na experiência acumulada. Descarte os que não funcionaram. Compartilhe prompts e os bons resultados com os colegas. Em comunidade, as forças se multiplicam e o aprendizado é potencializado.

- **Fique atento à regulamentação:** acompanhe as orientações do CNMP e da Corregedoria-Geral do Ministério Público sobre o uso das LLMs e restrições na Promotoria de Justiça. Contribua com o debate institucional sobre o emprego da IA no sistema de Justiça.

- **Estude:** participe de eventos e capacitações sobre o uso da IA Generativa no Ministério Público. O CNMP e a ESMP têm feito um excelente trabalho nessa área, ressaltando a importância de que seja usada com responsabilidade e com cuidados concernentes à privacidade.

- **Aprenda o básico da programação com Python:** o Python é uma linguagem de programação feita para não programadores, muito poderosa e de fácil aprendizado. É muito usada nos campos da IA e das automações. Saber Python amplifica, em

muito, o que se pode fazer com os LLMs. No meu artigo “[Por que promotores de justiça deveriam conhecer o Python?](#)”, publicado na Medium em setembro de 2023, dou algumas dicas sobre a iniciação com a linguagem.

- **Seja crítico:** tenha sempre uma postura crítica diante das respostas da IA. Revise todos os resultados. Você é o responsável pela peça processual ou manifestação que assina e jamais poderá culpar a IA, mera ferramenta auxiliar do processo intelectual de elaboração do documento, por um erro cometido. O profissional, membro ou servidor do Ministério Público, é o protagonista desse fluxo de trabalho.

## 15. CONCLUSÕES

Ao longo deste texto, abordamos diversos aspectos da Inteligência Artificial Generativa, focando em sua aplicação prática como ferramenta de produtividade pessoal no cotidiano da Promotoria de Justiça, em contraste com soluções puramente corporativas. Superada a fase inicial de curiosidade que marcou nossas primeiras interações com os chats há pouco mais de dois anos, vislumbramos hoje a possibilidade concreta de os modelos de linguagem se consolidarem como um recurso valioso no exercício de nossas funções.

De fato, a capacidade das LLMs de processar e analisar grandes volumes de texto, resumir autos judiciais e documentos complexos, auxiliar na pesquisa doutrinária e jurisprudencial e gerar minutas de peças processuais libera tempo precioso de membros e servidores da Instituição. Esse tempo pode ser redirecionado para atividades que demandam interação com o público, análise crítica, pensamento estratégico, empatia e julgamento humano.

Contudo, a incorporação dessa nova tecnologia à nossa rotina suscita preocupações legítimas com a privacidade e a segurança dos dados, especialmente porque lidamos com informações sensíveis e sigilosas. Igualmente relevante, no atual estágio de desenvolvimento dessas ferramentas, é a questão das "alucinações", assim entendida a propensão dos modelos a gerar informações incorretas, ainda que com aparência plausível. Essa anomalia reforça a necessidade imperativa de verificação humana rigorosa de todo conteúdo produzido pela IA e a importância de reiterar, por meio de condutas individuais e normatização, que a responsabilidade final pelo resultado e pelas decisões tomadas sempre recai sobre o profissional.

Observamos também que a eficácia da IA Generativa está diretamente ligada à habilidade do usuário em interagir com ela. A engenharia de prompt, a capacidade de formular instruções claras e contextualizadas e a disposição para iterar e refinar os comandos foram apresentados como fundamentais para extrair os melhores resultados.

A criação de bibliotecas de prompts e o compartilhamento de conhecimento entre colegas também foram destacados como práticas valiosas para acelerar a

adaptação e maximizar os benefícios dessas ferramentas, fomentando uma cultura de colaboração e aprendizado contínuo no âmbito do Ministério Público de São Paulo.

Em suma, o emprego da IA Generativa representa uma oportunidade significativa para aumentar a produtividade e a qualidade do serviço prestado à sociedade, desde que adotado com responsabilidade, criticidade e um compromisso constante com o aprendizado e a adaptação.

## REFERÊNCIAS

### Legislação e atos normativos

BRASIL. Conselho Nacional de Justiça. **Resolução nº 332, de 21 de agosto de 2020.** Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências. [Atualizada em 18 de fevereiro de 2025, conforme mencionado no texto].

BRASIL. **Lei nº 8.906, de 4 de julho de 1994.** Dispõe sobre o Estatuto da Advocacia e a Ordem dos Advogados do Brasil (OAB). Diário Oficial da União, Brasília, DF, 5 jul. 1994. [Referenciada indiretamente via Art. 7º, § 2º].

BRASIL. **Lei nº 10.406, de 10 de janeiro de 2002.** Institui o Código Civil. Diário Oficial da União, Brasília, DF, 11 jan. 2002. [Mencionado Art. 189 do CPC, que remete ao Código de Processo Civil].

BRASIL. **Lei nº 11.343, de 23 de agosto de 2006.** Institui o Sistema Nacional de Políticas Públicas sobre Drogas - Sisnad; prescreve medidas para prevenção do uso indevido, atenção e reinserção social de usuários e dependentes de drogas; estabelece normas para repressão à produção não autorizada e ao tráfico ilícito de drogas; define crimes e dá outras providências. Diário Oficial da União, Brasília, DF, 24 ago. 2006. [Referenciada Arts. 33, 40, 56].

BRASIL. **Lei nº 13.105, de 16 de março de 2015.** Código de Processo Civil. Diário Oficial da União, Brasília, DF, 17 mar. 2015. [Referenciado Art. 189].

BRASIL. **Decreto-Lei nº 2.848, de 7 de dezembro de 1940.** Código Penal. Diário Oficial da União, Rio de Janeiro, 31 dez. 1940. [Referenciados Arts. 97, 133, 155, 157, 171, 216-A, 234-A].

MINISTÉRIO PÚBLICO DO ESTADO DE SÃO PAULO. Procuradoria-Geral de Justiça. **Resolução nº 1.618/2023-PGJ-CPJ-CGMP.** [Norma interna sobre ANPP mencionada como exemplo].

### Jurisprudência

BRASIL. Conselho Nacional de Justiça. **Procedimento de Controle Administrativo nº 0000416-89.2023.2.00.0000**. Plenário. Julgado em 21 jun. 2024.

BRASIL. Superior Tribunal de Justiça. **Habeas Corpus nº 418.529/SP**. Relator: Ministro Nefi Cordeiro. Sexta Turma. Julgado em 17 abr. 2018. DJe, 27 abr. 2018.

BRASIL. Superior Tribunal de Justiça. **Habeas Corpus nº 434.544/RJ**. Relator: Ministro Joel Ilan Paciornik. Quinta Turma. Julgado em 15 mar. 2018. DJe, 3 abr. 2018.

BRASIL. Superior Tribunal de Justiça. **Habeas Corpus nº 436.168/RJ**. Relator: Ministro Ribeiro Dantas. Quinta Turma. Julgado em 22 mar. 2018. DJe, 2 abr. 2018.

BRASIL. Superior Tribunal de Justiça. **Habeas Corpus nº 868.720/SP**. Relatora: Ministra Daniela Teixeira. Quinta Turma. Julgado em 26 nov. 2024. DJe, 11 dez. 2024.

### **Artigos e Publicações online**

PIMENTEL, José Eduardo de Souza. Por que promotores de justiça deveriam conhecer o Python? **Medium**, set. 2023. Disponível em: <https://medium.com/@pimentel.jes/por-que-promotores-de-justica-deveriam-conhecer-o-python-26c9c7860f43>. Acesso em: 28 abr. 2025.