

SEM PROGRAMAR, OS CHATBOTS MAIS AVANÇADOS AINDA TÊM LIMITAÇÕES DE QUALIDADE E ALUCINAM COM GRANDES BASES DE DADOS

<https://claude.ai/public/artifacts/2b09851b-ca91-42ae-b191-3c170584047c>

SE O ARQUIVO É APENAS TEXTO OU PDF MAIS SIMPLES OS RESULTADOS PODEM SER MELHORES

SE OS ARQUIVOS SÃO GRANDES PLANILHAS OU SHAPEFILES - O MAIS SEGURO AINDA É PROGRAMAR

O FOCO DESTA AULA NÃO É ENSINAR PROGRAMAÇÃO (TEM APENAS UM EXEMPLO DISSO)

ENTÃO VAMOS MOSTRAR NO QUE FERRAMENTAS DE IA PODEM AJUDAR NA APURAÇÃO INICIAL

## Google AI Studio -

<https://aistudio.google.com/>

### Uma boa configuração para o Google AI Studio

\* **Temperatura 0:** esse parâmetro controla a aleatoriedade das respostas da IA. Com a temperatura definida como 0, o modelo se torna estritamente determinístico, sempre escolhendo a palavra com a maior probabilidade. Isso resulta em respostas consistentes e objetivas, mas pode limitar a criatividade e a naturalidade da linguagem. Para tarefas que exigem alta precisão, como análise de dados, é preferível usar uma temperatura mais baixa.

\* **Thinking mode ativado:** esse recurso, encontrado em modelos como o Gemini do Google e o Claude da Anthropic, permite que a IA mostre seu processo de raciocínio passo a passo. Isso proporciona uma transparência sem precedentes, permitindo que o usuário entenda como o modelo chegou a uma determinada conclusão, o que é fundamental para analisar problemas complexos e melhorar a precisão científica.

\* **Grounding with Google Search ativado:** esse recurso inovador conecta o modelo de IA diretamente à pesquisa do Google, permitindo o acesso a informações em tempo real. Isso melhora significativamente a precisão e a relevância das respostas, além de reduzir as “alucinações” da IA (respostas incorretas ou inventadas). A ferramenta também pode fornecer citações e fontes, o que é muito importante para o trabalho jornalístico.

Entretanto, em outros estágios do jornalismo, essa configuração pode ser limitante. A temperatura 0, por exemplo, pode inibir a criatividade necessária para:

- \* Conceber histórias e ângulos de notícias: uma temperatura um pouco mais alta poderia gerar ideias mais diversificadas e inesperadas.
- \* Escrever textos mais envolventes e com mais nuances: a linguagem gerada na temperatura zero pode soar robótica e repetitiva.
- \* Análise de sentimentos e interpretação de dados não estruturados: tarefas que exigem uma compreensão mais sutil da linguagem poderiam se beneficiar de uma maior flexibilidade do modelo.

### **Recomendações para uso em jornalismo:**

- \* Use a configuração acima para:
  - \* Verificação de fatos e de informações.
  - \* Criar resumos factuais de documentos e eventos.
  - \* Analisar dados estruturados para identificar tendências.
- \* Considere ajustar a "Temperatura" para valores um pouco mais altos (entre 0,3 e 0,7) para:
  - \* Gerar diferentes versões de títulos e chamadas para ação.
  - \* Explorar diferentes estilos de redação para o mesmo tópico.
  - \* Ajudar a criar roteiros para podcasts ou vídeos.

### **Sobre o Deep Research**

Na plataforma mais simples da Gemini (<https://gemini.google.com/>), o modelo 2.5 Pro traz relatórios ainda melhores, análises mais detalhadas e raciocínio aprimorado com o Deep Research. Mas ele só está disponível para o Google AI Pro

O Deep Research é um recurso que transforma a IA em um agente de pesquisa autônomo. Em vez de apenas dar uma resposta rápida, ela realiza um processo de pesquisa mais amplo e complexo.

Como funciona:

- \* Planejamento: A IA cria um plano de pesquisa multiponto com base em sua solicitação.
- \* Pesquisa autônoma: navega na Internet de forma autônoma para encontrar informações relevantes e atualizadas de diversas fontes.
- \* Síntese e relatório: a IA analisa e sintetiza os dados coletados para gerar relatórios detalhados e estruturados, que podem incluir citações.

Tanto o Gemini do Google quanto o ChatGPT da OpenAI oferecem a funcionalidade de “pesquisa profunda” (pode aparecer como Executar Investigação). O objetivo é economizar horas de trabalho manual em tarefas como análise de mercado, pesquisa acadêmica e comparação de produtos.

## **Como simular um Deep Research no Google AI Studio**

Atualmente, não existe um recurso chamado "Deep Research" que possa ser ativado com um único botão ou comando direto no Google AI Studio. Para que o Gemini realize pesquisas profundas no AI Studio, você precisa criar uma instrução detalhada que oriente o modelo pelo processo de pesquisa. Basicamente, você está pedindo ao modelo que simule o comportamento de um investigador.

Aqui está um guia passo a passo e um exemplo de como fazer isso:

1. Seja específico e estruturado em sua pergunta: Não pergunte apenas "tópico de pesquisa X". Dê ao modelo um roteiro claro do que você espera.
2. Solicite um plano de pesquisa: Instrua o modelo a primeiro delinear as etapas que você vai seguir. Isso força o modelo a "pensar" (usando seu "Modo de pensar" implícito) antes de agir.

3) Solicite a execução e a coleta de dados: Peça ao modelo para buscar informações para cada ponto do plano. Como o Gemini no AI Studio tem a capacidade de acessar informações (semelhante ao "Grounding with Google Search"), ele pode procurar dados atualizados.

4. Solicite a síntese e as citações: por fim, peça ao modelo para sintetizar as informações em um relatório coeso e, o mais importante, para citar as fontes das quais as informações foram extraídas.

## Chatgpt no navegador (Plus, Team ou Enterprise)

### Configuração recomendada para tarefas factuais

- **Modelo:** selecione **o3-pro** (ou GPT-4o, se ainda não tiver acesso ao o3) pelo menu “Model” na área de chat. Estes modelos entregam raciocínio mais sólido e reduzem alucinações.
- **Temperatura 0 – 0,2:** garante respostas determinísticas e replicáveis, ideais para checagem de fatos, resumo técnico e extração de entidades. Esse parâmetro só é exposto na API da OpenAI
- **Browsing ativado:** em Ferramentas habilite Buscar na Web
- **Advanced Data Analysis (ADA):** ligue a ferramenta quando quiser subir CSV, PDF ou shapefiles para estatísticas rápidas, gráficos ou checagem de planilhas volumosas diretamente no chat
- **Modo explicativo:** peça “explique seu raciocínio passo a passo, mas exiba apenas a conclusão no final” para evitar respostas prolixas e ainda auditar a cadeia lógica.

### Prompt para “investigação profunda” (equivalente ao Deep Research)

1. **Planejamento** (role: assistant): “Liste um plano de 3-5 passos para responder à questão X e cite as fontes esperadas.”

2. **Execução** (role: user): “Execute o passo 1 agora e traga o resultado com URL.”  
– repetir para cada passo.
3. **Síntese**: “Produza relatório de até 8 parágrafos com citações completas.”

Esta abordagem força o modelo a decompor o problema, a consultar o navegador em cada etapa e a deixar claro o raciocínio antes da síntese final.

O **o3** supera o GPT-4o em profundidade de raciocínio e contexto, mas gpt-3.5-turbo ou 4o-mini são ótimos para tarefas volumosas onde erros pontuais são toleráveis.

## Boas práticas de prompting para jornalistas

1. **Declare o papel**: “Você é um repórter investigativo especializado em finanças públicas, cite sempre as fontes com URL completo.”
2. **Forneça contexto e formato**: inclua dados-chave, links preliminares e peça resultado em Markdown com listas ou tabelas simples.
3. **Itere**: refine após cada resposta; pequenas correções no prompt melhoram coerência.
4. **Peça auditoria**: “Explique eventuais incertezas ou lacunas nos dados e proponha checagens adicionais.”
5. **Controle extensão**: delimite parágrafos ou tokens para evitar respostas longas demais.

## Onde economizar sem perder qualidade

- Execute **raspagens e parsing** com gpt-3.5-turbo, reserve o **o3** para sínteses finais.
- Use gpt-4o-mini local (Hugging Face ou Azure) para questões de confidencialidade.
- Ative **cache de tokens** na Realtime API para baixar custos em consultas repetitivas.

## Solicitar raciocínio passo a passo apenas em modelos rápidos

A instrução "ChainofThought" foi criada para modelos que respondem de forma direta e rápida. Em modelos reflexivos (como **Gemini 2.5 Pro**, **o3**, **gpt-4.1 (flagship)**, **claude4opus**, **mistrallarge**), o raciocínio interno já está implícito, e forçar a explicação dobra o tamanho do texto sem melhorar sua qualidade. Modelos rápidos (**gemini2.5 flash e flash-lite**, **gpt-4o**, **gpt-4.1mini**, **claude3.5 haiku**, **mistralsmall**, **modelos llama3**). Regra geral:

- \* Modelos rápidos → solicitar raciocínio passo a passo;
- \* Modelos reflexivos → solicitar apenas um resumo do raciocínio.

## + Combate ao Trabalho em Condições Análogas às de Escravo

<https://www.gov.br/trabalho-e-emprego/pt-br/assuntos/inspecao-do-trabalho/areas-de-atuacao/combate-ao-trabalho-escravo-e-analogo-ao-de-escravo>

### Pinpoint -

[https://journaliststudio.google.com/pinpoint/search?collection=8867e14b8c58373d&utm\\_source=collection\\_publish\\_link](https://journaliststudio.google.com/pinpoint/search?collection=8867e14b8c58373d&utm_source=collection_publish_link)

Exemplo de perguntas:

How many employers from the state of Minas Gerais (MG) are listed in the document as having subjected workers to conditions analogous to slavery?

What is the CNAE code for the establishment of ADELICIO BARBOSA DE ARAUJO (ID 10)?

What was the date the Cadastro de Empregadores was last updated?

The CNPJ 00.284.169/0001-10 has which address at the Receita Federal do Brasil?

**Exemplo de prompt para PDF ou CSV (mude para o que for usar) anexado do Cadastro de Empregadores que tenham submetido trabalhadores a condições análogas à de escravo (Lista Suja do Trabalho Escravo):**

Tarefa-macro

Você é um jornalista investigativo. Pesquise no contexto local e na internet informações confiáveis e produza um relatório claro que cubra os cinco pontos abaixo com evidências verificáveis. Cite a fonte ao fim de cada parágrafo (URL completo) e destaque possíveis limitações ou lacunas.

O arquivo a ser investigado em anexo é o Cadastro de Empregadores que tenham submetido trabalhadores a condições análogas à de escravo (Lista Suja do Trabalho Escravo) no Brasil, dados de 1/jul/2025

Fonte:

<https://www.gov.br/trabalho-e-emprego/pt-br/assuntos/inspecao-do-trabalho/areas-de-atuacao/combate-ao-trabalho-escravo-e-analogo-ao-de-escravo>

O que produzir:

1 - Dicionário de dados

Para cada coluna do arquivo cadastro\_de\_empregadores.csv:

Nome técnico e tradução em português comum.

Tipo de dado (ex.: texto, inteiro, data).

Exemplos de valores.

Observações sobre qualidade (nulos, formatação, ambiguidades).

2 - Interesse público e tema central

Explique por que a sociedade deve se importar com esse conjunto (relevância social, econômica, ambiental ou de direitos).

Contextualize com leis, políticas públicas ou decisões judiciais pertinentes (citar artigos e parágrafos).

3 - Cruzamentos recomendados, se forem possíveis

Liste no máximo 10 bases externas que contenham chaves comuns (CNPJ, CPF, CEP, coordenadas, CNAE, código IBGE, processo judicial etc.).

Descreva, para cada base, o endpoint ou caminho de download, o formato (CSV/XLSX/API/GeoJSON) e o script ou consulta API sugerida para ligar ao arquivo principal.

Outras fontes correlatas

Portais de transparência estaduais ou municipais.

Dados satelitais (ex.: PRODES, MapBiomas).

Registros judiciais (PJe, e-TRT, CNJ DataJud).

Listas de certificação (ex.: orgânicos, exportadores, ESG) ou selos de compliance.

Relatórios de ONGs especializadas.

4 - Sugestões adicionais, se forem possíveis

Indicadores ou métricas que ainda não existam na base mas podem ser calculados.

Possíveis entrevistas com vítimas, órgãos de fiscalização ou especialistas.

Visualizações ideais (mapa, linha do tempo, treemap, sankey).

Sinais de alerta para inconsistências (valores fora de faixa, duplicidades, datas impossíveis).

Estratégia de versionamento de scripts/notebooks para auditoria posterior.

Regras de qualidade:

Use linguagem objetiva, sem adjetivos opinativos.

Não conclua sem antes verificar no mínimo duas fontes independentes para cada afirmação factual.

Se algo não for encontrado, descreva o passo de busca realizado e marque como “não identificado”.

Explique seu raciocínio e onde foram retiradas as informações

## + Sigmine - Processos minerários

<https://dados.gov.br/dados/conjuntos-dados/sistema-de-informacoes-geograficas-da-mineracao-sigmine>

**Arquivo de Metadados**

**Processos minerários ativos - Brasil**

+ Cadastro Mineiro

<https://dados.gov.br/dados/conjuntos-dados/sistema-de-cadastro-mineiro>

**Microdados SCM**

<https://app.anm.gov.br/dadosabertos/SCM/microdados/>



Ver dados de um processo -

<https://sistemas.anm.gov.br/scm/extra/site/admin/dadosprocesso.aspx>

Pesquisar um processo -

<https://sistemas.anm.gov.br/scm/extra/site/admin/pesquisarprocessos.aspx>

Script em Python -

<https://colab.research.google.com/drive/1sPQQ5Nb18j9ypNZgXyxEXgBd-4ykgfNz?usp=sharing>

Veja uma análise completa de como LLMs e IA Generativa analisam shapefiles hoje -

<https://claude.ai/public/artifacts/afecb42d-c7c1-49b2-9d31-e2596e553824>

## Prompt de análise

Mas mesmo assim, por exemplo, você pode usar um modelo raciocínio/reflexão avançada como o **o3 da OpenAI** no ChatGPT para pelo menos fazer uma análise inicial de dados de shapefile de Processos minerários ativos - Brasil, para ter uma ideia bem inicial dos dados e/ou receber um script em Python para testar

## Para criar um código

Você é um jornalista de dados.

Contexto: Estou investigando mineração no Piauí usando o shapefile BRASIL.dbf, BRASIL.shp e outros arquivos que forem importantes, que contém polígonos de Processos minerários ativos - Brasil. Poligonais dos processos minerários ativos (incluindo áreas em disponibilidade) mantidos pela Agência Nacional de Mineração.

Extensão geográfica: Brasil. Formato: Shapefile. Fonte:

<https://dados.gov.br/dados/conjuntos-dados/sistema-de-informacoes-geograficas-da-mineracao-sigmine>

Campos relevantes: AREA\_HA (área em ha), UF (estado), FASE (fase do processo), DSProcesso (código do processo de mineração).

Tarefas:

Ler o shapefile com Python/GeoPandas e reprojetar para EPSG:5880 (Equal-Area para Brasil).

Calcular:

- número total de feições;
- área total (km<sup>2</sup>) e mediana;

- top-10 maiores áreas com DSProcesso e UF;
- distribuição por UF e por FASE (tabela).

Gerar explicação dos resultados em até 8 parágrafos, linguagem acessível, voz ativa.

Indicar até três pontos que exigem checagem adicional.

Mostrar o código usado (comentado) e mencionar fontes das funções.

Saída: texto + tabelas inline. Não omitir advertências sobre limitações do CRS ou qualidade dos dados.

### **Para análise sem código:**

Você é um assistente de IA especializado em dados espaciais para jornalistas leigos.

Contexto: investigo mineração no estado do Piauí com o shapefile do Sigmine (processos minerários ativos). Arquivos fornecidos: BRASIL.shp, BRASIL.dbf, BRASIL.prj, BRASIL.cpg. Atualização: 1/jul/2025.

Objetivos que preciso entender:

Quantos processos existem e qual a área total e mediana em km²?

Quais os dez maiores processos (DSProcesso + UF) e suas áreas?

Como se distribuem por UF e por fase (FASE)?

O que quero que você faça passo a passo:

abra o shapefile, explique rapidamente o que é cada arquivo auxiliar;

reprojete para EPSG 5880 antes de medir áreas, explicando por que isso é preciso;

calcule as estatísticas acima e apresente-as em tabelas fáceis de ler;

descreva os resultados em linguagem acessível, em até 8 parágrafos, voz ativa;

crie 3 listas numeradas: (a) pontos que exigem checagem extra; (b) implicações públicas ou políticas; (c) perguntas de acompanhamento;

destaque limitações do formato shapefile e do dado do sigmine;

cite toda informação numérica indicando “sigmine [data]”.

Formato de saída: texto corrido + tabelas inline; cada termo técnico novo deve ser definido em no máximo 20 palavras. Sem código; apenas narrativa e tabelas.

Orientações extras:

- se o arquivo for grande ou a conexão instável, peça à IA para amostrar 10.000 linhas antes da análise completa; isso economiza tempo sem perder noção de escala .

- adicione a instrução “explique cada etapa de forma que um aluno do 8.º ano entenda” para prevenir respostas excessivamente técnicas .

- depois de receber o resumo, peça mapas ou gráficos se precisar de visualização adicional — mencione explicitamente “use cores contrastantes e legenda clara”.