Limites e possibilidades do uso do ChatGPT para cruzamento e análise de dados por jornalistas[1]

Vítor Almeida dos Santos[2]
Universidade Estadual de Ponta Grossa, Ponta Grossa, PR

Resumo: Este resumo expandido busca fomentar a discussão sobre a utilização real de um chatbot, no caso, o ChatGPT para o cruzamento e análise de dados. Para isso, ofereceu-se um caso real do seu uso, incluindo os comandos utilizados. A partir desse exemplo, indicamos uma discussão sobre a inserção dessa ferramenta na rotina jornalística, ainda que mais próxima do Jornalismo de Dados, e da necessidade de refletir sobre a falta de proteção do chatbot em relação aos dados fornecidos ao sistema de inteligência artificial. Para este último movimento, analisamos a política de privacidade da OpenAi, empresa dona do ChatGPT.

Palavras-chave: ChatGPT, análise de dados, proteção de dados.

1. Introdução

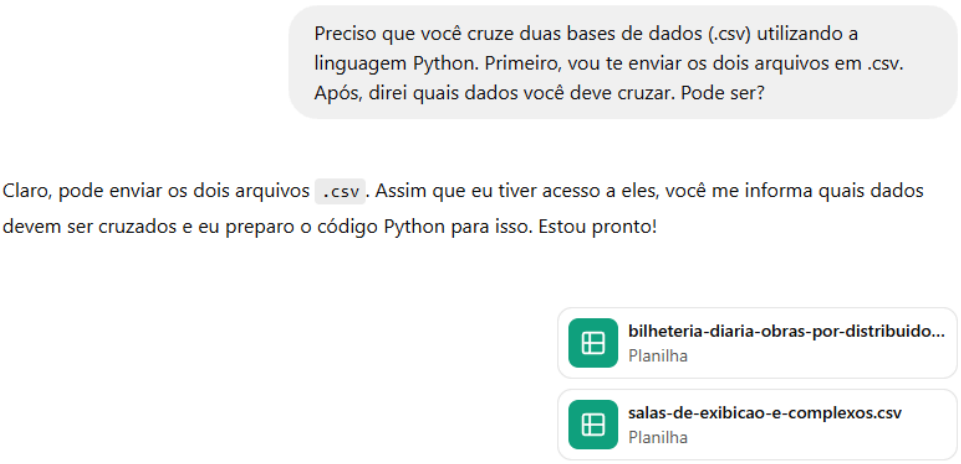
Este resumo expandido apresenta uma forma de aplicação do ChatGPT no cruzamento e análise de dados, para a elaboração de uma pauta. A partir disso, discute-se, ainda que inicialmente, como o ChatGPT pode ser utilizado no processo de produção jornalística, também é abordada a falta de segurança na proteção dos dados enviados ao chatbot.

Essa pesquisa se motiva em uma tentativa pessoal de cruzar duas bases de dados da Agência Nacional do Cinema (ANCINE)[3] para uma possibilidade de pauta. No caso, pedimos ao ChatGPT que cruzasse as salas de exibição que tiveram início de funcionamento em 2024, com o público do mês de novembro dessas salas do filme "Ainda estou aqui", para ver se a abertura de novas salas, através do fundo setorial, se

[1] Resumo expandido apresentado no 28º Seminário de Inverno, realizado entre 2 e 6 de junho de 2025, na Universidade Estadual de Ponta Grossa (UEPG).
[2] Mestrando no Programa de Pós-Graduação em Jornalismo da Universidade Estadual de Ponta Grossa. Bolsista CNPQ-CAPES, e-mail: vitoralmeida.jor@gmail.com
[3] Dados disponíveis no endereço: https://www.gov.br/ancine/pt-br/oca/dados-abertos. Acesso em 18 de maio de 2025.

traduziu em público expressivo para um filme de grande repercussão para o cinema nacional. Para isso, importamos duas planilhas em formato .csv, uma com as salas de exibição e complexos, outra com a bilheteria diária das obras por distribuidora, referente a novembro de 2024. Pedimos que esse cruzamento fosse feito utilizando a linguagem de programação Python, pois teríamos condições de avaliar o retorno dos nossos comandos (prompts)[4]. Um fator de impacto no retorno desses prompts é que utilizamos a versão gratuita do ChatGPT, que impossibilitou que o cruzamento fosse feito dentro da plataforma. No nosso caso, a IA indicou que havíamos chegado ao limite de prompts, nos oferecendo a migração para o plano pago. Também nos foi possibilitada a descrição dos códigos em Python para que pudéssemos executar em um IDLE (ambiente onde se executa os comandos de uma linguagem de programação), específico. Portanto, executamos externamente. Por experiência prévia, utilizamos a IDLE gratuita Jupyter Notebook.

No comando inicial informamos a linguagem de programação a ser utilizada, o Python. Além dos formatos dos arquivos que utilizamos para o cruzamento. Ver Figura 1:



[4] Os prompts utilizados e os retornos podem ser vistos acessando https://chatgpt.com/share/682b87df-bfa0-8005-907a-5c253f9302bc

Figura 1 – Prompt 1 e retorno 1
Fonte: Seminário de Inverno (2025)

Após o envio dos arquivos, o ChatGPT nos retornou confirmando o recebimento e logo perguntou sobre qual tipo de cruzamento escolheríamos. Neste momento, descrevemos as etapas do nosso cruzamento, indicando colunas específicas do arquivo e a ordem, enumerada, de cada passo que o ChatGPT deve seguir. Ver Figura 2:

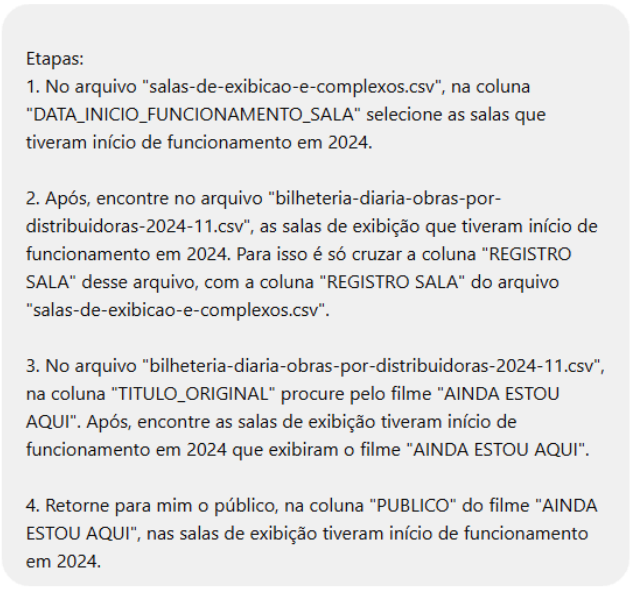


Figura 2 – Prompt 3 e retorno 3
Fonte: Seminário de Inverno (2025)

O ChatGPT nos retornou um erro, mas automaticamente recomeçou a fazer o cruzamento. Como foi mencionado na introdução, a versão gratuita não nos permitiu executar o código em Python internamente, nos oferecendo apenas a possibilidade de copiar o código para executar em outro ambiente. Então pedimos o código em Python. Ver Figura 3:

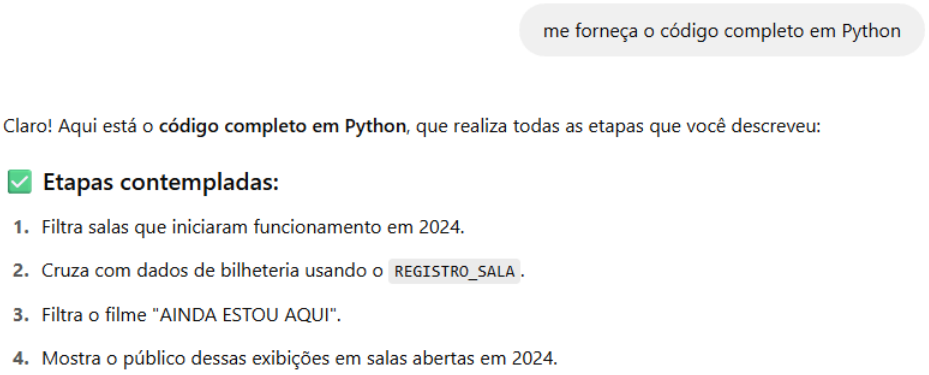


Figura 3 – Pedimos o código em Python
Fonte: Seminário de Inverno (2025)

O código recebido foi este. Ver Figura 4:

```python
import pandas as pd

# Carregar os arquivos CSV
salas_df = pd.read_csv("salas-de-exibicao-e-complexos.csv", sep=";", encoding="latin1")
bilheteria_df = pd.read_csv("bilheteria-diaria-obras-por-distribuidoras-2024-11.csv", sep=";", encoding="utf-8")

# Converter a coluna de datas para o tipo datetime
salas_df["DATA_INICIO_FUNCIONAMENTO_SALA"] = pd.to_datetime(
    salas_df["DATA_INICIO_FUNCIONAMENTO_SALA"], format="%d/%m/%Y", errors="coerce"
)

# Etapa 1: Filtrar salas que começaram a funcionar em 2024
salas_2024 = salas_df[salas_df["DATA_INICIO_FUNCIONAMENTO_SALA"].dt.year == 2024]

# Etapa 2: Cruzar com dados de bilheteria usando "REGISTRO_SALA"
salas_2024_ids = salas_2024["REGISTRO_SALA"].dropna().unique()
bilheteria_salas_2024 = bilheteria_df[bilheteria_df["REGISTRO_SALA"].isin(salas_2024_ids)]

# Etapa 3: Filtrar pelo filme "AINDA ESTOU AQUI"
filme_filtrado = bilheteria_salas_2024[bilheteria_salas_2024["TITULO_ORIGINAL"].str.upper() == "AINDA ESTOU AQUI"]

# Etapa 4: Exibir o público do filme nessas salas
resultado = filme_filtrado[["NOME_SALA", "DATA_EXIBICAO", "PUBLICO"]]

# Mostrar o resultado
print("\n🎬 Público do filme 'AINDA ESTOU AQUI' em salas inauguradas em 2024:\n")
print(resultado)

# Se quiser, salvar em CSV
resultado.to_csv("publico_ain_da_estou_aqui_2024.csv", index=False)
```

Figura 4 – Código recebido em Python para executar externamente

Fonte: Seminário de Inverno (2025)

Esse procedimento demonstrado acima serve para pensarmos que o jornalista precisa reconhecer dois fatores que permeiam a utilização desse chatbot para esse fim: a autoria no processo de produção jornalística e a proteção dos dados informados ao ChatGPT.

Quando utilizamos o chatbot para construir um produto jornalístico, de quem é a autoria? (Laia; Bravin; Guimarães; Morais, 2024). Ainda que no trabalho desses autores tenha-se dedicado a outras possibilidades de uso de chatbots, como resumo de textos, roteiro de perguntas para entrevistas, buscas por fontes, adaptação de formatos, reescrita e tradução de discursos, por exemplo, a pergunta nos parece também contemplar jornalistas que utilizam ou pretendem utilizar o ChatGPT para cruzar dados.

Na pesquisa de campo feita nas redações jornalísticas do estado do Piauí, Berti (2024) observou o uso de sistemas de inteligência artificial no processo jornalístico. Por mais que nessa pesquisa não seja explicitado o que caracteriza, de fato, o uso de I.A em cada etapa de produção, o autor (2024) destacou que as pautas, "nos seus momentos iniciais" (Berti, 2024, p.142), recebem auxílio desses sistemas. Já a apuração é descrita como a etapa mais humanizada na redação, ainda que Berti (2024, p. 142) mencione para a possibilidade do uso de inteligência artificial em uma "apuração para Jornalismo de Dados".

Mesmo que o uso de chatbots, ou sistemas de inteligência artificial, a depender da abordagem da pesquisa, estejam na rotina do jornalismo (Laia; Bravin; Guimarães; Morais, 2024), entendemos que, ao recortar a utilização dessa ferramenta para uma análise de dados, acabamos por direcionar seu uso dentro de parâmetros atrelados ao Jornalismo de Dados, onde os recursos computacionais são entendidos como uma ampliação da expertise jornalística (Träsel, 2014).

> É possível encontrar uma agulha num palheiro sem a ajuda de nenhuma ferramenta, caso o sujeito disponha de tempo, mas ninguém poderá negar que o uso de uma peneira ou um soprador para agilizar o processo produza um resultado diferente – uma agulha. Da mesma

> forma, as ferramentas de computação agilizam o trabalho de correlacionar dados brutos, sem prejuízo para o resultado final em termos jornalísticos. Ou, antes, qualquer prejuízo ao resultado final é tão atribuível ao uso de computadores quanto uma notícia mal redigida é atribuível ao fato de ter sido digitada numa máquina de escrever. (Träsel, 2014, p.103).

Ainda que possamos entender que o jornalista tenha capacidade técnica de avaliar a cada resposta oferecida por um chatbot, no nosso caso, o ChatGPT, precisamos entender que há um componente silenciosamente inerente ao uso do sistema da OpenAI para análise e cruzamento de dados: a proteção dos dados.

Para este resumo expandido, analisou-se a Política de Privacidade do ChatGPT atualizada em novembro de 2024. Esse documento registra que a OpenAi, empresa dona do ChatGPT, coleta o "conteúdo de usuário" [...] incluindo os seus prompts e outros conteúdos que você faz upload, tais como arquivos, imagens e áudio, dependendo das funcionalidades que você utiliza" (OpenAI, online, 2024).

A OpenAI também registra que pode divulgar nossos dados nos seguintes casos: "fornecedores e prestadores de serviços"; "transferências comerciais", "Autoridades Governamentais ou Outros Terceiros"; "Afiliados", "Administradores de Contas Corporativas", "Outros Usuários e Terceiros com quem Você Interage e Compartilha Informações" (OpenAI, online, 2024). Em "Seus direitos", a empresa também informa que é possível enviar uma solicitação de privacidade[5], ainda que não fique claro qual a cobertura da proteção desses dados.

Ao final dessa exploração inicial, duas capacidades dos jornalistas aparecem como horizonte esperado. Primeiro, é preciso entender o que se está pedindo ao chatbot, a linguagem de programação e as etapas necessárias para a obtenção de um resultado exitosos. A capacidade de "promptar", no caso demonstrado, não se sobrepõem a capacidade jornalística de reconhecimento dos critérios para uma possível pauta, corroborando com o entendimento que uso o ferramental de recursos computacionais

[5] É possível solicitar pelo endereço: https://privacy.openai.com/policies. Acesso em 18 de maio de 2025.

ampliam a capacidade do jornalista. Por outro lado, que dados estão sendo enviados ao ChatGPT? Sem esse tipo de reflexão prévia, o jornalista fica à mercê de uma política de privacidade que não protege a informação, muito menos as fontes, desrespeitando inclusive o que prevê o artigo 5° do código de ética dos jornalistas, que nos garante o sigilo de fontes (Fenaj, 2007).

Referências

Agência Nacional do Cinema. Dados Abertos. Disponível em: https://www.gov.br/ancine/pt-br/oca/dados-abertos. Acesso em: 19 maio 2025.

ANCINE, Agência Gov | Via. Governo Federal financia novas salas de cinema e modernização do circuito exibidor. 2024. Disponível em: https://agenciagov.ebc.com.br/noticias/202411/numero-recorde-de-salas-de-cinema-no-brasil#:~:text=Apoio%20e%20indu%C3%A7%C3%A3o%20do%20FSA&text=Al%C3%A9m%20da%20moderniza%C3%A7%C3%A3o%20de%20outras,em%20todo%20o%20territ%C3%B3rio%20nacional.. Acesso em: 19 maio 2025.

BERTI, Orlando Maurício de Carvalho. Jornalismo e inteligência artificial: desafios nas redações. Revista Pauta Geral: Estudos em Jornalismo, Ponta Grossa, v. 11, p. 133-148, dez. 2024.

FENAJ. Código de Ética dos Jornalistas Brasileiros. 2007. Disponível em: https://fenaj.org.br/codigo-de-etica-dos-jornalistas-brasileiros/. Acesso em: 19 maio 2025.

LAIA, Evandro J. M.; BRAVIN, Adriana; GUIMARÃES, Lara Linhalis; MORAIS, Marina Magalhães de. Inteligência artificial e jornalismo: alianças possíveis para traduções além do humano. **Galáxia**, São Paulo, v. 49, p. 1-22, out. 2024.

OPENAI. Política de Privacidade. 2024. Disponível em: https://openai.com/pt-BR/policies/privacy-policy/. Acesso em: 19 mar. 2025.

OPENAI. OpenAI Privacy Portal. 2024. Disponível em: https://privacy.openai.com/policies. Acesso em: 19 maio 2025.

PYTHON. IDLE — Python editor and shell. Disponível em: https://docs.python.org/3/library/idle.html. Acesso em: 19 maio 2025.

TRÄSEL, Marcelo Ruschel. ENTREVISTANDO PLANILHAS: estudo das crenças e do ethos de um grupo de profissionais de jornalismo guiado por dados no brasil. 2014. 315 f. Tese (Doutorado) - Curso de Programa de Pós-Graduação em Comunicação Social, Pontifícia Universidade Católica do Rio Grande do Sul., Porto Alegre, 2014.