AGENTES DE IA E PROCESSOS JUDICIAIS SIGILOSOS

Estratégias e limitações

.....

José Eduardo de Souza Pimentel
GE Cível da Capital | 30/09/2025

Agenda

- Vivendo o *hype* da IA Gen
- A IA Gen e o sigilo processual
- As estratégias das corporações
- Tudo é contexto na "Engenharia de Prompt"
- **M365 Copilot** e Agentes
- Por onde seguir?

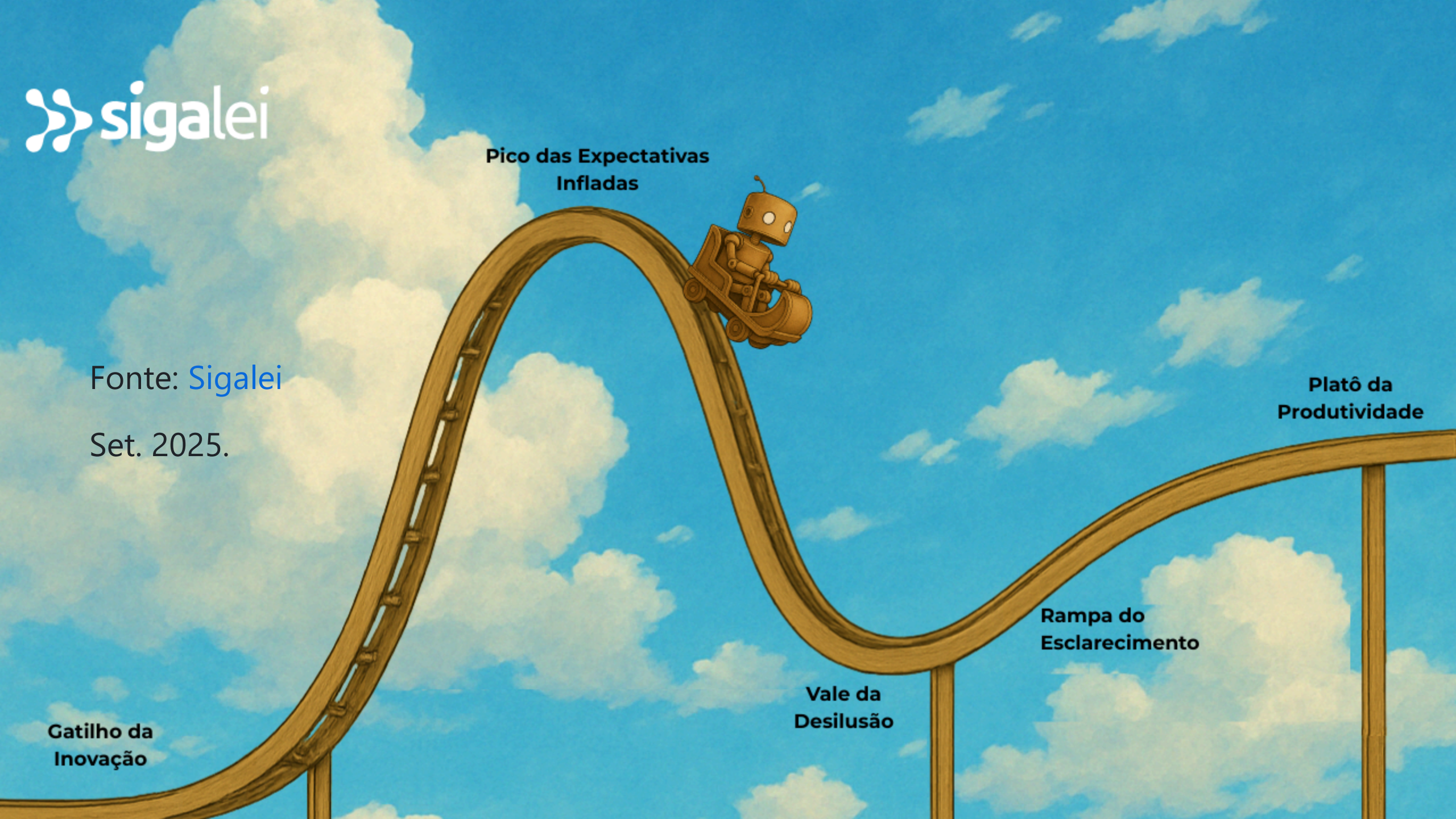
Vivendo o *hype* da IA Gen

OpenAI Imagines Our AI Future

Stages of Artificial Intelligence	
Level 1	Chatbots, AI with conversational language
Level 2	Reasoners, human-level problem solving
Level 3	Agents, systems that can take actions
Level 4	Innovators, AI that can aid in invention
Level 5	Organizations, AI that can do the work of an organization

Source: Bloomberg reporting

Bloomberg

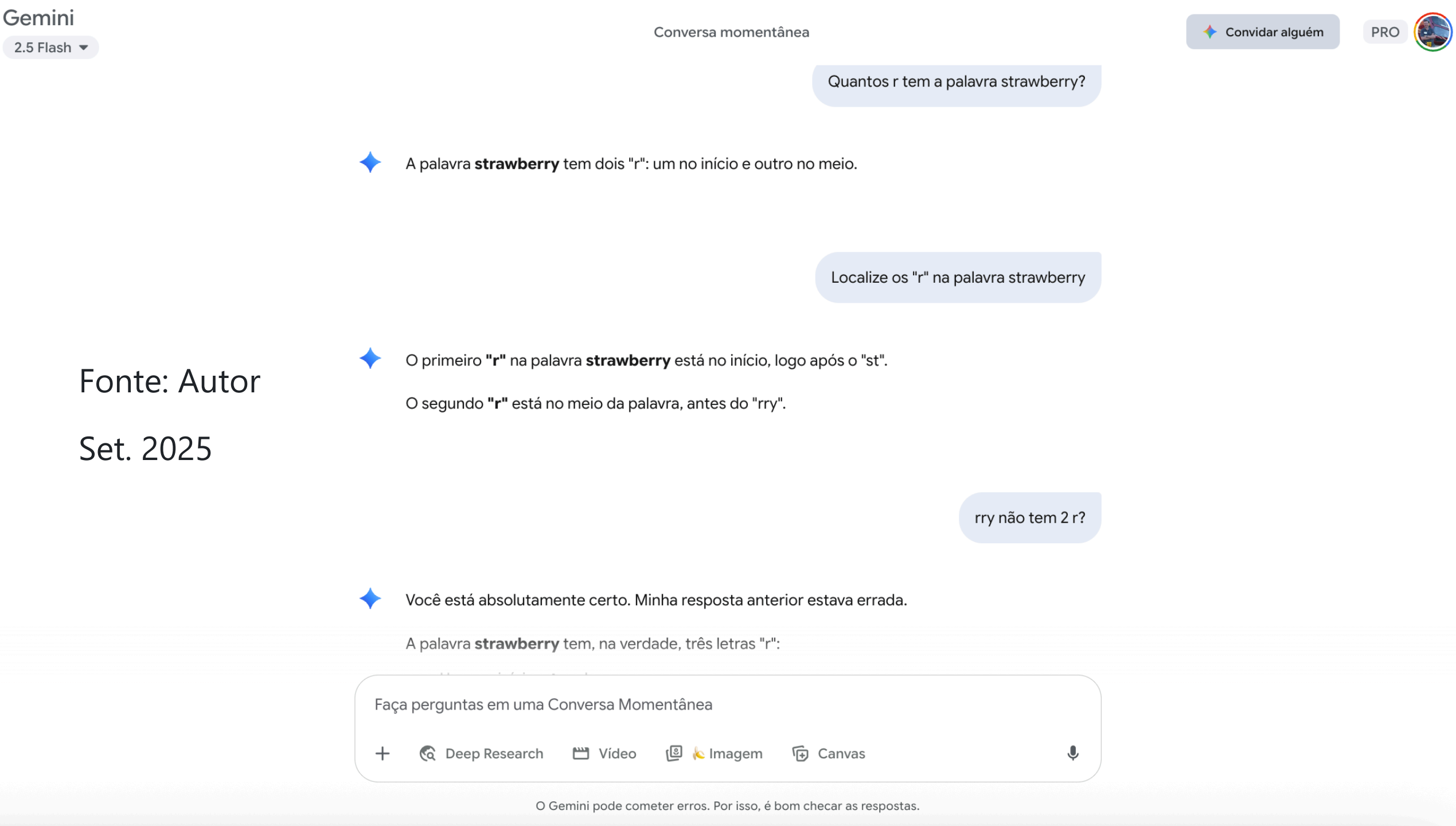


Fonte: Sigalei

Set. 2025.

Visão geral

- Modelos aprendem como as informações se organizam e estão estatisticamente distribuídas no corpus de treinamento
- LLMs: geram textos coerentes e relevantes em resposta aos comandos (**prompts**)
- Não possuem conhecimentos factuais (?)



Fonte: Autor

Set. 2025

Aplicações mais óbvias

- Redação de documentos
- Análise e resumo de texto
- Pesquisa jurídica (Jurisprudência GPT)
- Revisão e correção de texto
- Treinamento

Dá pra fazer também...

- Obtenção de hashes e manipulação de arquivos
- Obtenção de frames e tratamento de imagens
- Análise de dados
- Infográficos, linhas do tempo, esquemas
- Brainstorm
- Apps ("vibe coding")
- Etc... (**é um "canivete suíço"**)

Privacidade

- Conhecer termos de uso e opções *opt-in* ou *opt-out* dos provedores

Alucinação

- Modelos não possuem compreensão real dos assuntos
- Respostas plausíveis (baseadas em padrões probabilísticos)
- "Não sabem o que não sabem"

Sugestões:

- Conferir as informações geradas
- Não utilizar como única fonte de pesquisa
- Estar cientes das limitações da tecnologia

A IA Gen e o sigilo processual

Regulamentação

- PCA nº 0000416-89.2023.2.00.0000 (j. 21/06/2024)
- Resolução nº 615/2025 - CNJ
  - assinatura ou cadastro privado (se não tiver solução corporativa)
  - neste caso, proibido em dados sigilosos ou protegidos por segredo de justiça
- Considerações:
  - solução de produtividade pessoal
  - assimetria com Defensores/Advogados

Aviso nº 009/2025 - CGMP

- utilização **prioritária** do Copilot
- **anonimização** dos dados ao se utilizar ferramentas de IA não contratadas
- abstenção do compartilhamento de **dados sensíveis ou sigilosos** em "plataformas abertas"
- **revisão criteriosa de todo conteúdo gerado**

As estratégias das corporações

Ferramentas "oficiais"

- Judiciário: **ApoIA** (integrada ao PD e Codex)
  - Chave da API (tribunal custeia | privado)
  - Não contempla: criminal e proc. sigilosos
- MPSP: **Tilene**
  - RAG e APIs (GPT-4o e GPT4.1)
  - Prompts: pré-configurado | livre | pós-upload
  - Casos de uso (?)

Create filters to allow or block specific types of content

- Basic information
- Input filter
- **Output filter**
- Deployment (optional)
- Review

Set output filter

What are these categories

Content will be annotated by each categories and blocked according to the threshold. For Violent content, Hate content, Sexual content, and Self-harm content category, adjust the threshold to block harmful content with equal or higher severity levels.

Category	Media	Action	Threshold
Violence	Text Image	Annotate and block	Medium Allow Low / Block Medium and High
Hate	Text Image	Annotate and block	Medium Allow Low / Block Medium and High
Sexual	Text Image	Annotate and block	Medium Allow Low / Block Medium and High
Self-harm	Text Image	Annotate and block	Medium Allow Low / Block Medium and High
Protected material for text	Text	Annotate and block	Protected material will be blocked

Fonte: Microsoft

Set. 2025

Tudo é contexto na "Engenharia de Prompt"

Noções de Engenharia de prompt

- **Contexto**: fornece informações situacionais que ajudam o modelo a compreender melhor o cenário sobre o qual ele aplicará as instruções.
- **Dados de entrada**: informação ou arquivo fornecido à IA para processamento.
- **Persona**: define o papel do modelo (exemplo: "Você é um promotor de justiça.")
- **Tarefas**: define as tarefas que o modelo deve executar (exemplos: "analise", "compare", "liste", "reescreva", “resuma de forma estruturada”).
- **Formato de saída**: 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(s) exemplo(s) fornecido(s)").

Dicas para a elaboração de bons prompts:

- Comece simples
- Divida tarefas complexas
- Dê um papel ao modelo
- Adicione contexto relevante
- Use instruções claras, específicas e diretas
- Forneça exemplos
- Diga o que não fazer
- Utilize tags (<tag></tag>) e/ou Markdown (#, ** e -)
- Converse com o modelo (iteração)

"Vazamento" do System Prompt da Anthropic

- System Prompt do Claude 4 (22/05/2025)

Lições aprendidas

- O prompt pode ser grande (15k+)
- Seções estruturadas por XML: <externa><interna></interna></externa>
- Emprego de "intenções declarativas" (capacidades e limitações)
- Uso de lógica condicional (muitos "if")
- Repetição das instruções importantes
- Ênfases com maiúsculas e um pouco de Markdown (#, ** e -)
- Restrições: "DO NOT", "Do not" e "don't"
- Exemplos: bons e maus

M365 Copilot e Agentes

Visão geral

- Especificidades
- Capacidades
- Pros e Cons
- "Lost in the middle"

Sugestões de uso

- Corretor de peças (prompt básico)
- Relatórios processuais (few-shot prompting)
- Peças processuais (carga de exemplos)
- "Esquematizador" de processos (resultado formatado)

Dicas de implementação

- Casos mais simples
- Criação de agentes "especializados"
- Nova configuração de peças:
  - fatos
  - fundamentos jurídicos
  - identificação dos casos reaproveitáveis (recuperação)
- BD de prompts
  - Institucional (?)
  - Pessoal (GitHub, Google Keep, Obsidian, etc.)
- Compartilhamento de agentes
- Aprendizado constante (sempre há novidades...)

Por onde seguir?

- A IA GENERATIVA NA PROMOTORIA (Pimentel)
- Prompting Guide 101 (Google)
- What Is ChatGPT Doing … and Why Does It Work? (Stephen Wolfram)
- Acervo de vídeos da ESMP

Slides da palestra

https://jespimentel.github.io/ge_civel_2025/

Repositório dos prompts deste GE:

https://github.com/jespimentel/ge_civel_2025/tree/main/prompts

Obrigado!

linkedin.com/in/jespimentel

jespimentel.blogspot.com

github.com/jespimentel

pimentel@mpsp.mp.br