

IpeaIA

Projeto Estratégico de Inteligência Artificial

Instituto de Pesquisa Econômica Aplicada — Ipea

2025 – 2026



ipeaIA - Diretrizes



Soberania de dados

Pesquisas sensíveis não podem sair da instituição. Sem dependência de APIs externas.

Produtividade

Automatizar análises repetitivas. Liberar pesquisadores para o que realmente importa.

Acesso democrático

IA de ponta disponível para todos os pesquisadores e técnicos do Ipea.

Dados não estruturados → CSV

Tabulação de processos judiciais, arquivos de áudio e imagens em dados estruturados

Novas formas de estruturar informação

Busca acelerada de informações e interação direta com resultados de pesquisa

Por que usar?



IpealA — Visão Geral

O IpealA é o projeto estratégico do Ipea para incorporar Inteligência Artificial nas atividades de pesquisa e gestão institucional. Engloba infraestrutura de hardware, modelos de linguagem, aplicações e capacitação para pesquisadores e técnicos.



Infraestrutura

Servidores de IA dedicados, on-premise, com hardware de última geração



Modelos de IA

LLMs internos e privados, sem envio de dados sensíveis para terceiros



Aplicações

Ferramentas práticas integradas ao fluxo de trabalho dos pesquisadores



Educação

Cursos e capacitação em IA para pesquisadores e equipe técnica

Infraestrutura de IA

set/2024

Primeiras GPUs

- NVIDIA Quadro RTX 4000 (8 GB)
- Retiradas de workstations do Ipea
- Instaladas no datacenter
- Capacidade para modelos até ~13B parâmetros
- Viabilizou primeiros experimentos com LLMs

jun/2025

POC — Servidor Intel Gaudi 2

- Prova de Conceito (POC). Jul/2025 a jun/2026
- Hardware: Intel Gaudi 2
- 8 × 96 GB VRAM = 768 GB total (tensor par)
- Suporta modelos até ~235B parâm. em BF16, e até ~500B de parâm. em FP8
- vLLM (substitui o Ollama)

mai/2026

2 × Servidores GPU — Nvidia H200

2025: R\$ 2,5m (0,4m EUR)

- 4 × GPUs Nvidia H200 por servidor
- Capacidade para modelos de 70B+ parâmetros
- Infraestrutura permanente e soberana do Ipea
- Suporte a múltiplos projetos simultâneos

29 modelos

testados e avaliados

Hardware: Quadro RTX 4000 (set/24 – jul/25)

Hardware: Intel Gaudi 2 (jul/25 – presente)

Processo: benchmark sistemático (Arena ELO, MMLU, MMLU-Pro) + testes com equipes de pesquisa

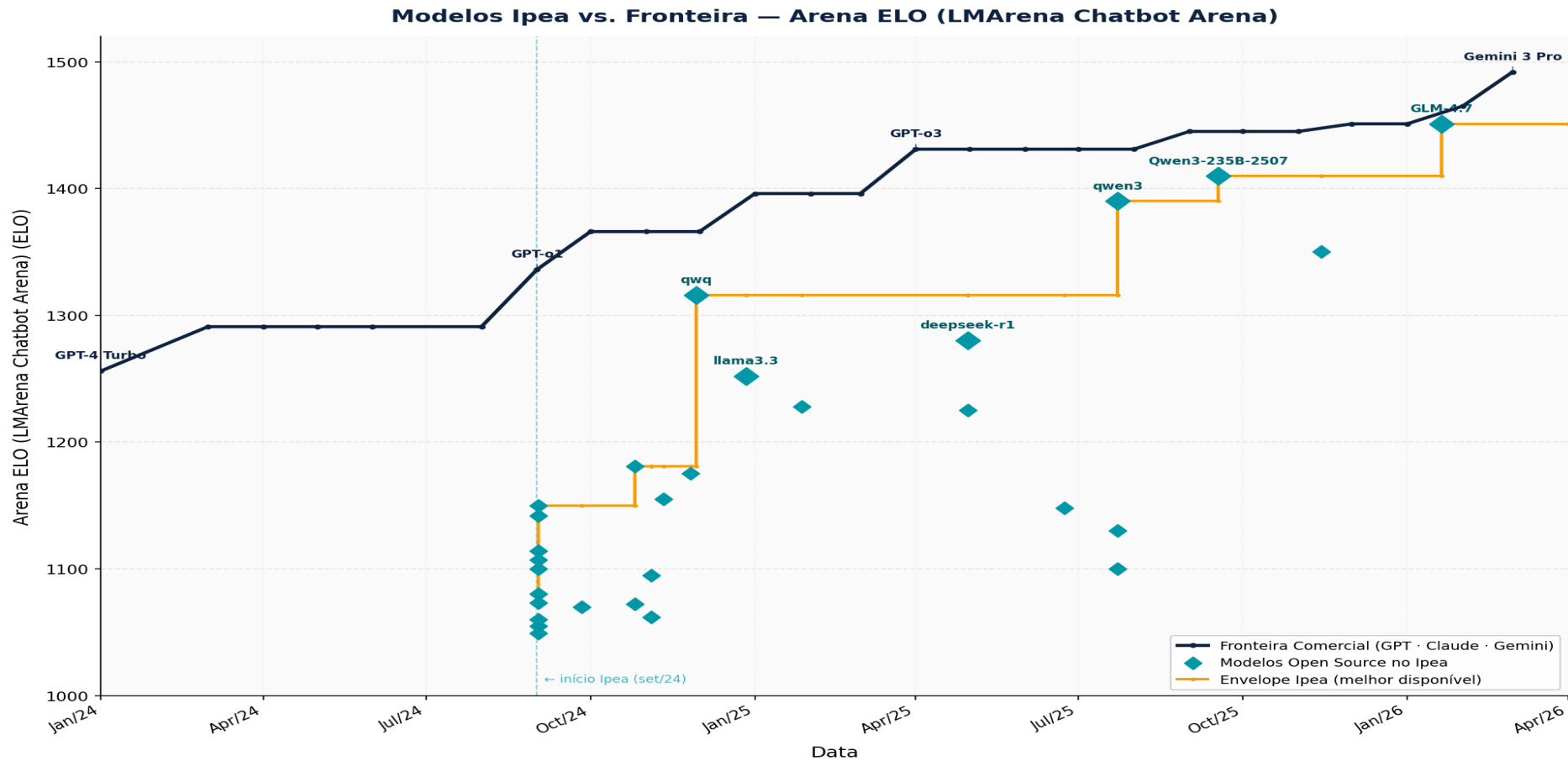
Perspectiva: com H200 (mai/26), esperamos rodar modelos rápidos como Qwen3.5-9B – Qwen3.5-122B, Qwen3.5-397B e GLM-5 755B

- ▶ **Raciocínio:** qwq, deepseek-r1
 - ▶ **Geral:** llama3.3, phi4, gemma3, qwen3, GLM-4.7
 - ▶ **Código:** deepseek-coder-v2, qwen2.5-coder, deepcoder
 - ▶ **Matemática:** qwen2-math, mathstral
 - ▶ **Multimodal:** llama3.2-vision
-

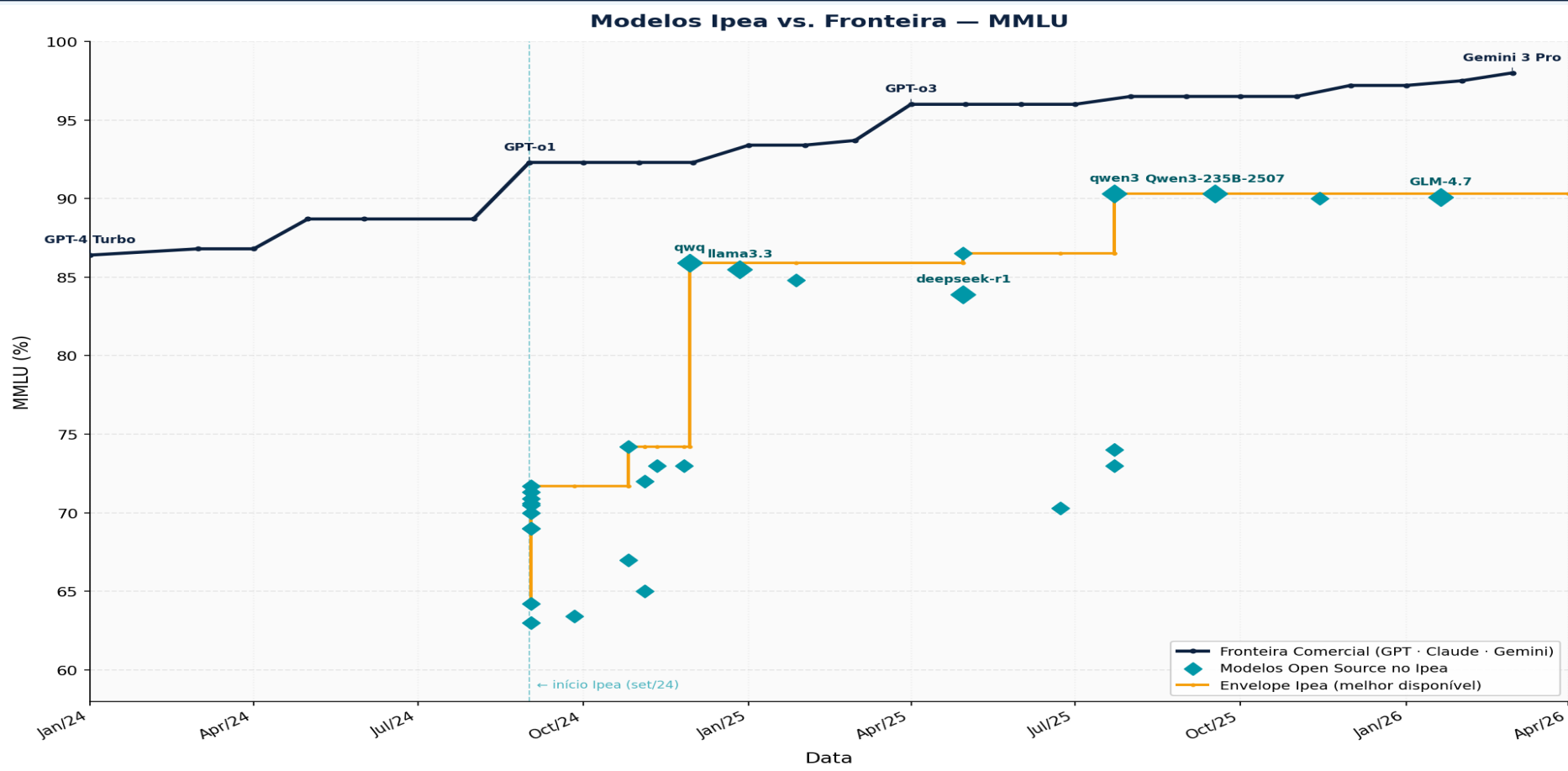
★ **Destaques:** qwq · qwen3 · Qwen3-235B-2507 · GLM-4.7

Modelos com maior ELO Arena e cobertura de benchmarks

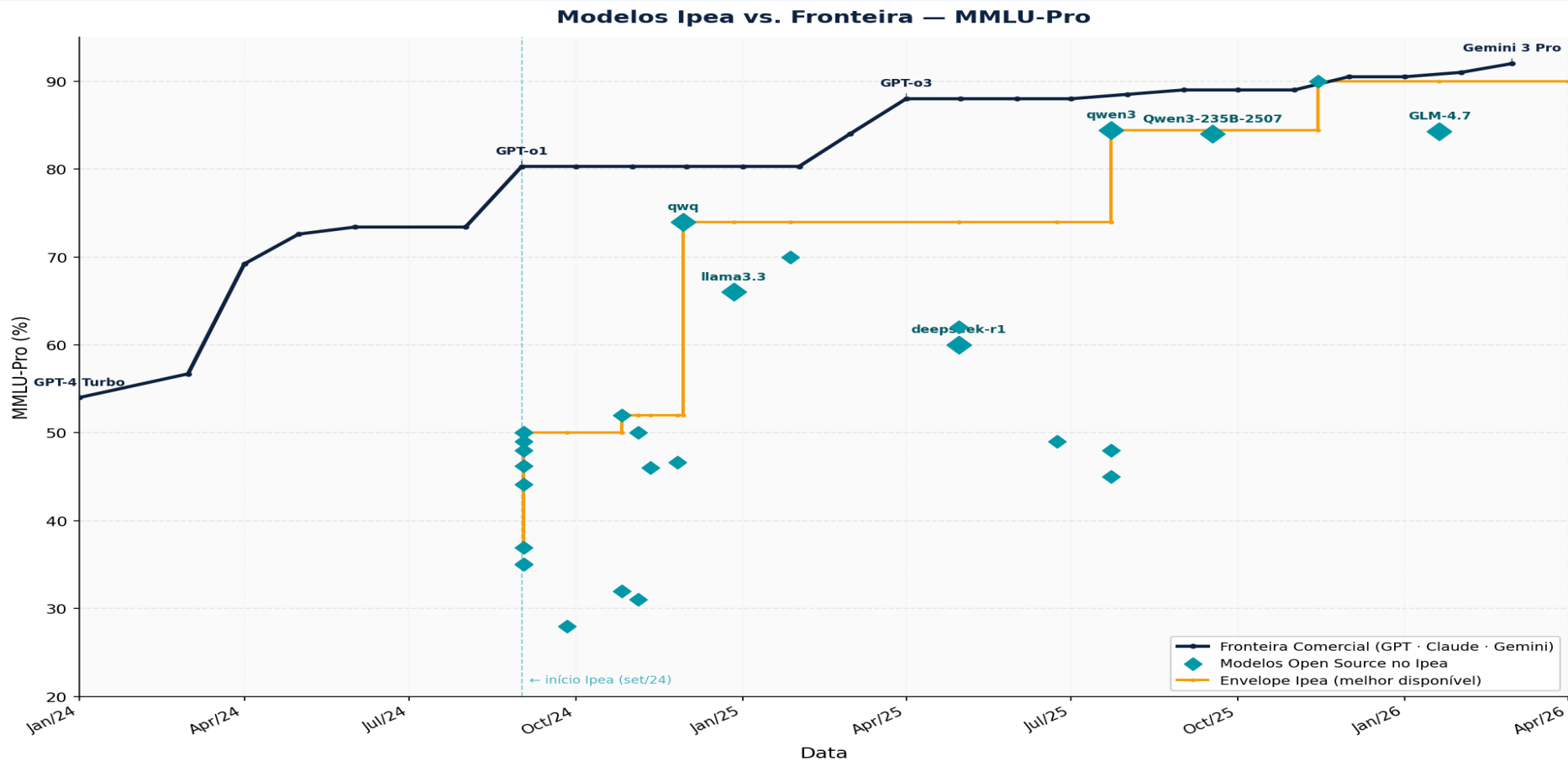
Desempenho — Arena ELO (LMArena)



Desempenho — MMLU (Multitask Language Understanding)



Desempenho — MMLU-Pro



Aplicações Desenvolvidas

- **ipeaGPT**

- Chatbot LLM interno para assistência em pesquisa e tarefas de texto
- Disponível no servidor do sigilo

- **API: LLM-Pesquisa**

Chamadas recorrentes

- **ipeaLEG**

Busca semântica e análise de legislação e normas jurídicas

- **OCR**

Conversão de documentos digitalizados e imagens em texto estruturado

- **Positron IDE Assistant**

Assistente de código integrado ao Positron IDE via modelo local

- **ipeaPUB**

RAG para busca semântica no banco de dados de publicações e relatório do Ipea.

- **ipeaREV**

Cadeira de agentes de revisão de texto seguindo os padrões do Ipea

- **openCode**

Assistente de programação baseado em modelos open-source

ipeaGPT

Open WebUI



Qwen/Qwen3-235B-A22B-Instruct-2507 ▾ +

Set as default



New Chat

Search


Notes

Workspace

Models

Folders

Chats

 Qwen/Qwen3-235B-A22B-Instruc...

Qwen3 é a mais recente geração de modelos de linguagem de grande porte da série Qwen, oferecendo um conjunto abrangente de modelos densos e de mistura de especialistas...

How can I help you today?



↪ Suggested

Grammar check

rewrite it for better readability

Tell me a fun fact

about the Roman Empire

Give me ideas

for what to do with my kids' art

API: LLM-Pesquisa

```
URL = "https://ipeagpt.ipea.gov.br/api/v1/chat/completions"
```

```
AUTH = TOKEN_IPEA_GPT
```

```
import requests
```

```
def classifica_processo_qwen(processo_json):
```

```
    prompt = f"""
```

```
    Classifique o seguinte processo judicial em termos de favorabilidade e se é uma decisão final:
```

```
    Identifique primeiramente os campos do JSON que correspondem a decisões e utilize-os para determinar se a decisão é favorável ao autor, ao réu ou neutra e se é liminar ou definitiva.
```

```
    Processo: "{processo_json}"
```

```
    Responda em formato JSON com os campos "favorable_to" (valores possíveis: "autor", "réu", "neutro") e "is_final" (valores possíveis: true, false).
```

```
    """
```

```
try:
```

```
    payload = {
```

```
        "model": "Qwen/Qwen3-235B-A22B-Instruct-2507",
```

```
        "messages": [
```

```
            {
```

```
                "role": "system",
```

```
                "content": "Você é um assistente jurídico especializado em análise de processos judiciais. Os processos estão em português, e cada documento se encontra em formato JSON, sinalizando o tipo de documento e seu conteúdo.
```

```
            },
```

```
            {"role": "user", "content": prompt},
```

```
        ],
```

```
        "max_tokens": 10000,
```

```
        "temperature": 0.0,
```

```
    }
```

```
    headers = {"Authorization": AUTH, "Content-Type": "application/json"}
```

```
    response = requests.post(URL, json=payload, headers=headers)
```

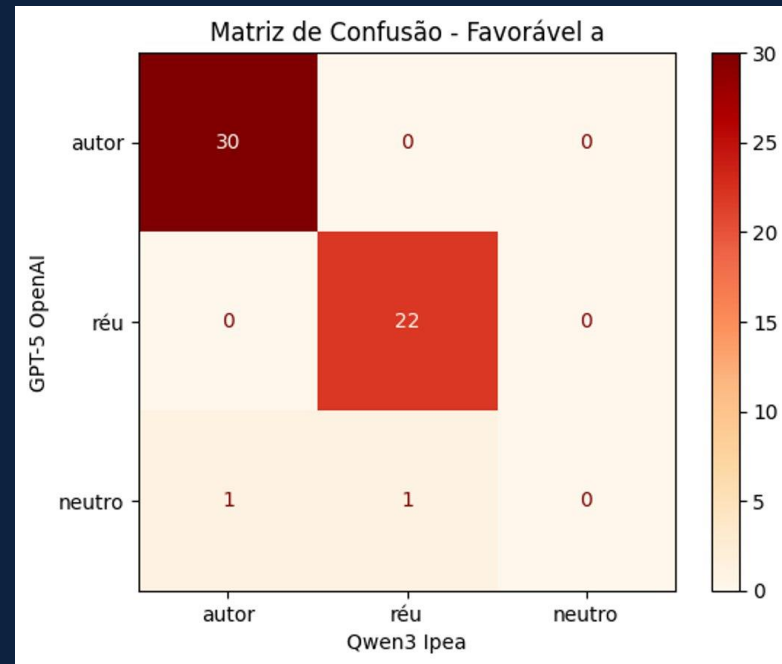
```
    resposta_texto = response.json()["choices"][0]["message"]["content"]
```

```
try:
```

```
    resposta_json = json.loads(resposta_texto)
```

Aplicação: classificar processos judiciais

Decisão favorável a:	BPC Deficiente	BPC Idoso	Total
Autor	22	9	31
Inss	18	5	23
Total	40	14	54



ipeaLEG: RAG legislação

- Interface web
- Consulta em linguagem natural
- Respostas estruturadas com indicação de fonte
- Links diretos para os documentos



The screenshot displays the IpeaLEG web interface. At the top left, the logo for 'ipea Instituto de Pesquisa Econômica Aplicada' is visible. At the top right, the email address 'OIA.rodrigo.prates@ipea.gov.br' is shown. The main heading is 'IpeaLEG' with a magnifying glass icon, and below it, the text 'Assistente de Normas e Portarias'. A search prompt 'Digite sua pergunta:' is followed by a text input field containing the question: 'Qual é o fluxo de publicações no IPEA? Cada diretoria tem seus procedimentos próprios?'. A blue 'Enviar' button is located at the bottom left of the input field.

ipeaLEG: RAG legislação

- Cada parágrafo possui referência ao documento de origem
- Links permitem abrir o documento na página correspondente
- Destaque automático do trecho relevante
- Transparência e verificabilidade da informação

Tempo total da resposta: 44.78 s

Resposta

O processo de publicação de documentos técnicos e pesquisas no IPEA envolve a produção de artigos inéditos sobre temas diversos, com o objetivo de fomentar o debate acadêmico e subsidiar políticas públicas. Os principais produtos editoriais, como Textos para Discussão (TDs) e Documentos de Pesquisa (DPs), podem ser posteriormente publicados em revistas acadêmicas ou livros, e seus resultados não refletem necessariamente a posição institucional do IPEA. Esses documentos são acompanhados de um resumo executivo (executive summary) e têm como finalidade compartilhar resultados de pesquisas realizadas no âmbito do Instituto [BOLETIM DE PESSOAL_8_20221230.pdf](#) (no_171 - Status: VIGENTE).

Quanto à padronização dos procedimentos, os documentos recuperados não detalham se todas as diretorias seguem fluxos idênticos ou se há autonomia para procedimentos internos distintos. No entanto, a existência de normas institucionais, como portarias e boletins, indica a busca por certa uniformidade e regulamentação dos processos editoriais.

Sobre a regulamentação, a produção científica elaborada em projetos vinculados ao IPEA deve ser disponibilizada para disseminação institucional. A propriedade intelectual dessas produções pertence ao IPEA e, em casos de co-gestão, também à instituição parceira, resguardando o crédito do autor. O IPEA é responsável por disseminar toda a produção científica resultante dessas atividades [PORTARIA_492_20101230.pdf](#) (no_61 - Status: VIGENTE).

Em resumo, a produção e publicação de documentos técnicos no IPEA seguem diretrizes institucionais que visam padronizar e garantir a divulgação dos resultados, embora detalhes específicos sobre fluxos internos de cada diretoria não estejam explicitados nos trechos apresentados.

Ver texto / copiar resposta

Documentos citados

Clique em um documento para abrir o Assistente do Documento e fazer perguntas específicas sobre ele.

- BOLETIM%20DE%20PESSOAL_8_20221230.pdf
- PORTARIA_492_20101230.pdf

Exportar Log

Educação — Capacitação em IA

nov/2025

Curso Básico de LLMs para Pesquisa

Público: equipe de apoio à pesquisa / Conteúdo: uso prático de LLMs, prompting, ferramentas internas (ipeaGPT) / Carga: 8 horas / Participantes: 26 servidores / Avaliação: feedback positivo de 96%

Concluído ✓

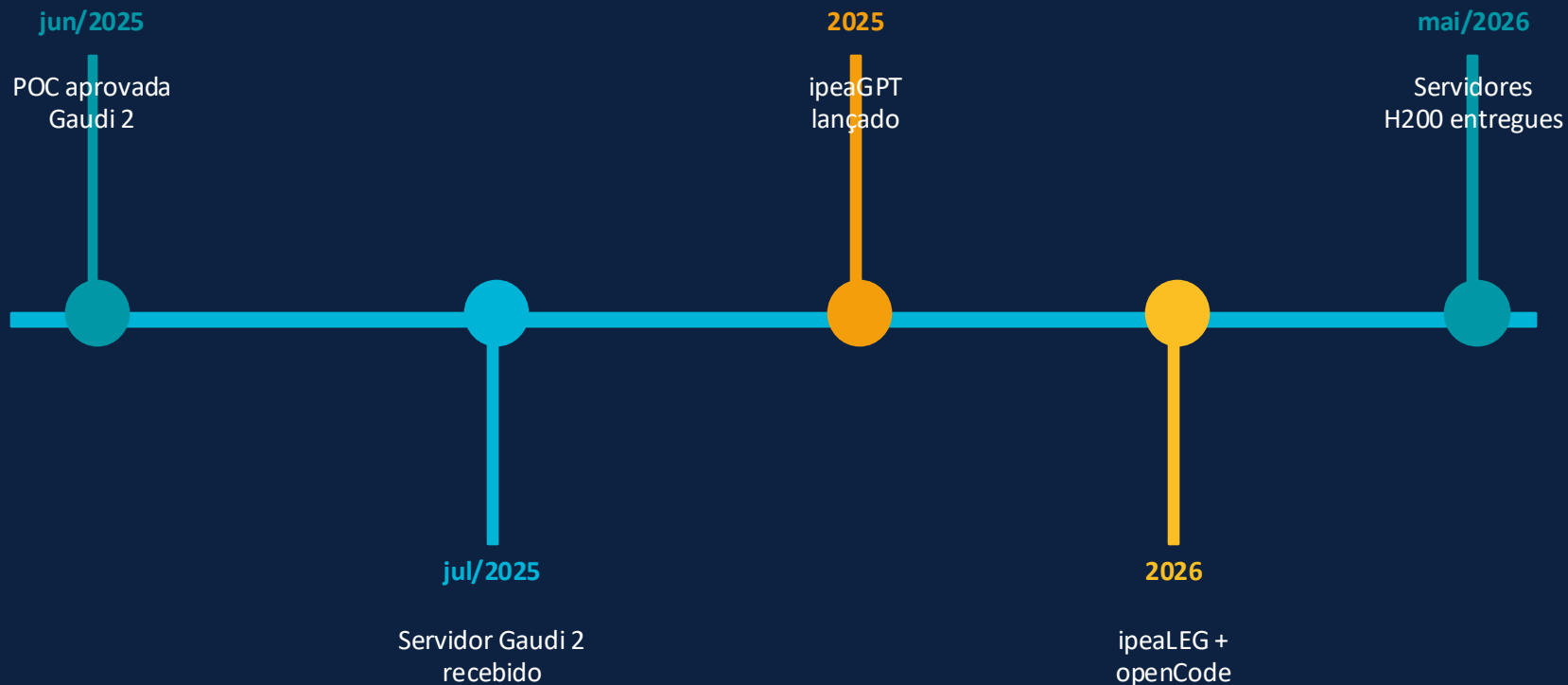
ago/2026

Curso de IA para Pesquisa

Público: pesquisadores do Ipea / Conteúdo: fundamentos de IA, uso de modelos internos, utilização de agentes, aplicações em pesquisa econômica, ética em IA / Carga: 30 horas / Perspectiva: capacitar 30 pesquisadores para uso avançado de IA em publicações e análises

Previsto

Linha do Tempo — IpeaIA



jun/2026 — encerramento do período POC Gaudi 2

IpeaIA

IA Soberana para Pesquisa e Políticas Públicas

- 1 servidor POC (Gaudi 2) — em operação
- 2 servidores definitivos (H200) — mai/2026
- 7 aplicações: ipeaGPT · ipeaLEG · openCode · ipeaPUB · OCR · Positron · ipeaGPT Remoto
- Soberania de dados e privacidade garantidas

