

Cloud Native Day São Paulo

Do Prometheus ao **GPT-4** :
Uma nova Era da Observabilidade Inteligente

Ana Luiza Primo, 27

Site Reliability Engineer

Stone CO e Itaú

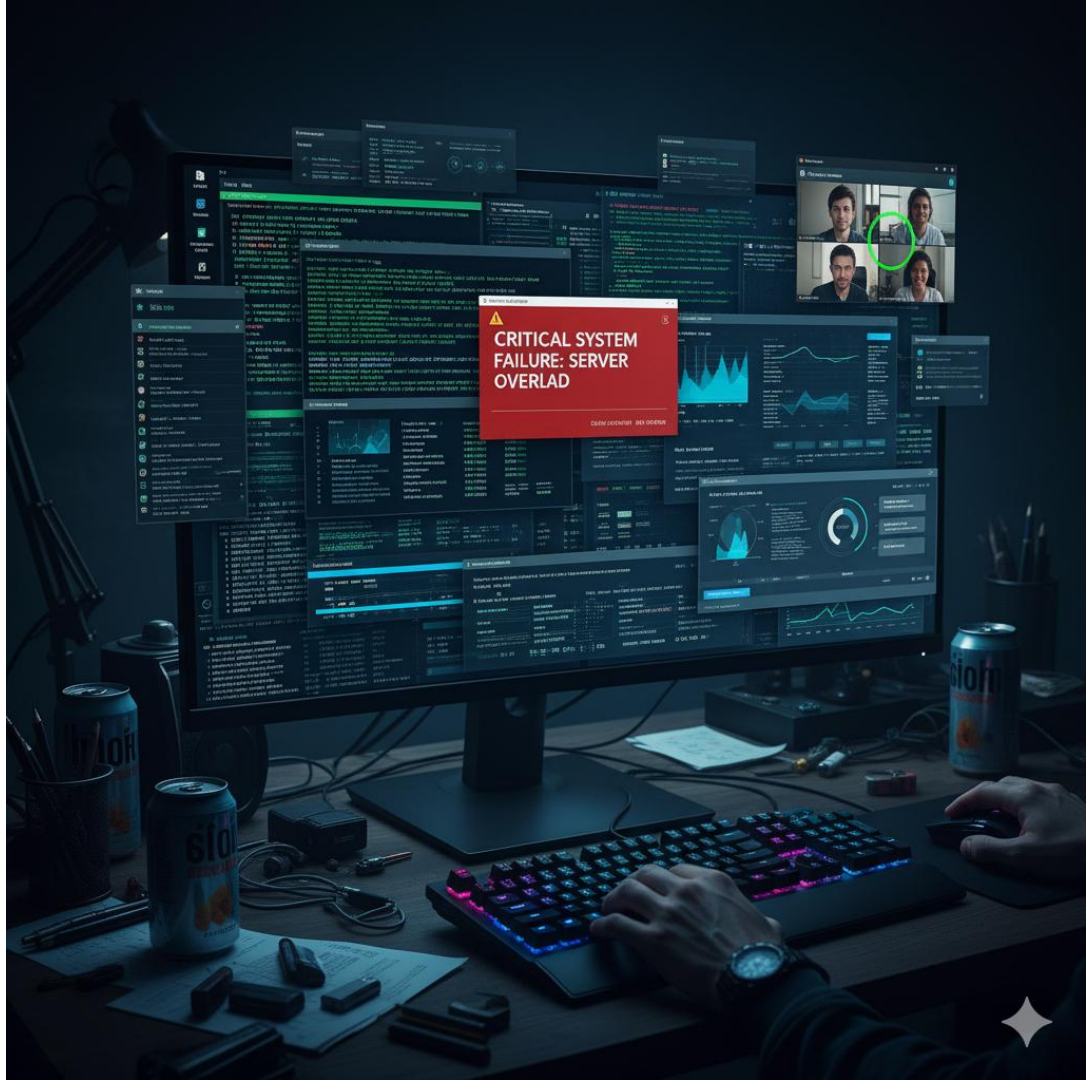
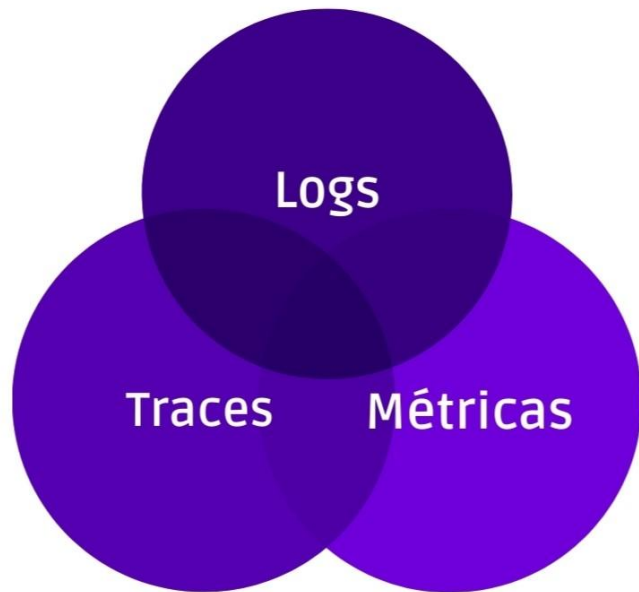
**Project Management – Ohio
University**



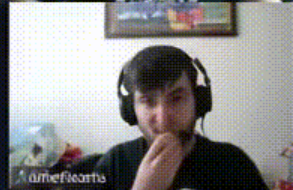
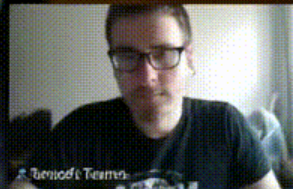
A woman with long dark hair, wearing a red lace-trimmed top, stands in a city street. She is holding a white rectangular sign in front of her chest. The background is a blurred city street with buildings, trees, and a dark car parked on the side. The lighting is bright, suggesting daytime.

**Tudo nesta
apresentação foi gerado
com IA**

O Que é Observabilidade?



War Room



Clase
trabajadora

Inteligencia
artificial



Mercado

AIops Tradicional (Baseada em ML): Focada em detecção de anomalias, correlação de eventos e redução de ruído.

GenAI (Baseada em LLMs): A nova onda, focada em interação em linguagem natural, sumarização de incidentes e geração de remediação.

"Co-pilotos" de Observabilidade (Aplicações de GenAI)



Bits AI SRE: atua como um "colega de equipe" de IA no on-call.

Funcionalidades Chave:

- **Sumarização de Incidentes:** Cria resumos em tempo real de alertas e mudanças, explicando o que está acontecendo.
- **Análise de Causa Raiz (RCA):** Sugere causas prováveis para um incidente correlacionando métricas, traces e logs.
- **Geração de Post-mortem:** Ajuda a rascunhar post-mortems com base nos dados do incidente (algo que alinha com seu interesse em SRE).
- **Geração de Correções:** Pode sugerir correções de código ou etapas de remediação.

"Co-pilotos" de Observabilidade (Aplicações de GenAI)



new relic®



Instrument and onboard.

- ✓ What does Apdex mean?
- ✓ Identify alert coverage gaps.
- ✓ Help me integrate with AWS.
- ✓ What are the most important metrics for infra monitoring?



Find and fix issues.

- ✓ Explain this stack trace.
- ✓ Did this app stop sending data?
- ✓ Analyze this error log.
- ✓ How do I fix this error?



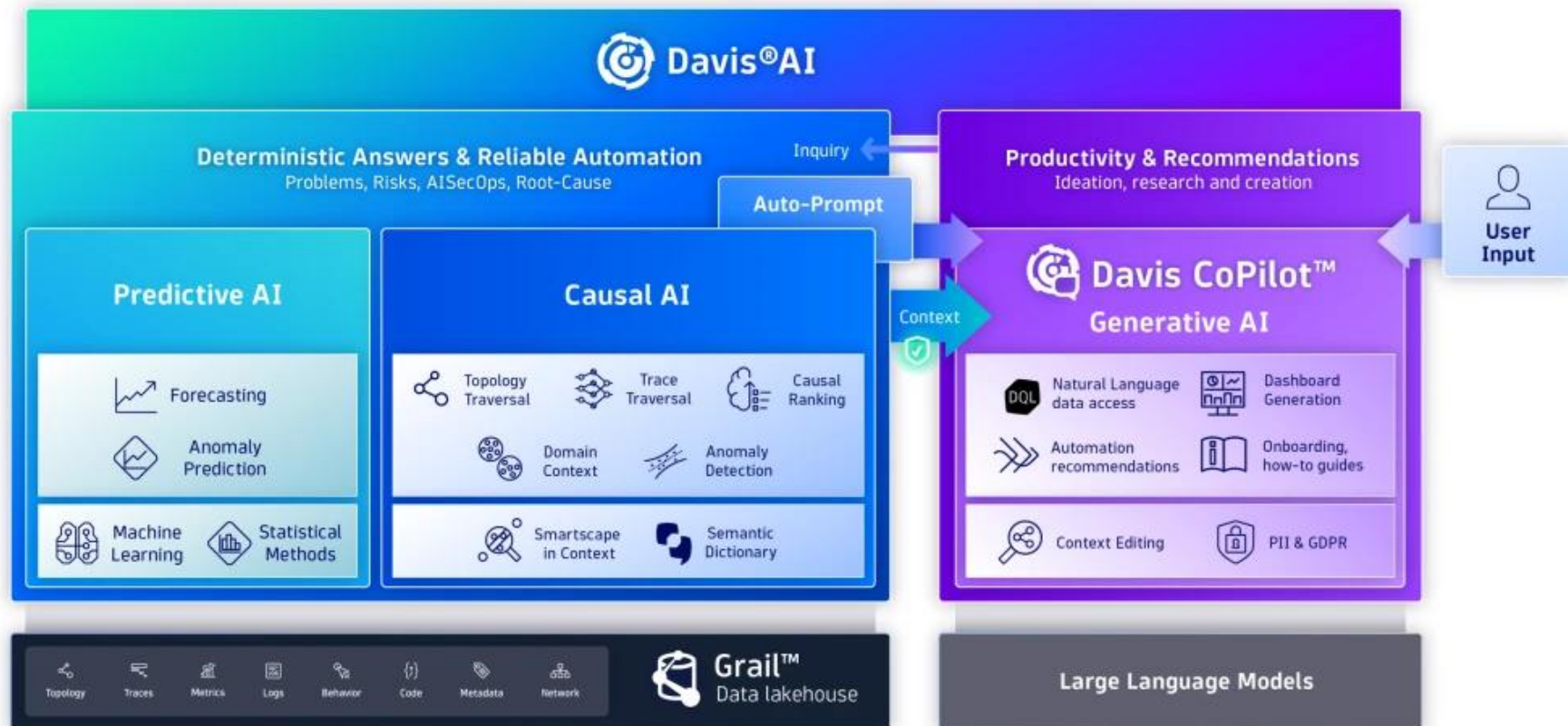
Get insights fast.

- ✓ Summarize recent performance issues.
- ✓ Explain this dashboard.
- ✓ How many active users did we have last week?
- ✓ How did the last deployment go?



Translate and collaborate.

- ✓ Explain this NRQL query.
- ✓ Rewrite this query to include more error types.
- ✓ Create an exec summary based on query results.
- ✓ Explain this dashboard in Spanish.





TS runner-service.ts × runner-service.test.ts × TS runner-service.ts ×

```
1 import { type Runner } from '@prisma/client';
2 import prisma from '$lib/data/prisma';
3
4 export interface RunnerService {
5   getById(id: number): Promise<Runner | null>;
6 }
7
8 const runnerService: RunnerService = {
9   getById: async (id: number) => {
10     return await prisma.runner.findUnique({
11       where: { id },
12     });
13   }
14 };
15
16 export default runnerService;
```

CHAT



Edit with Copilot

Agent mode

Start your editing session by defining a set of files that you want to work with. Then ask Copilot for the changes you want to make.

Copilot is powered by AI, so mistakes are possible. Review output carefully before use.

 or type # to attach context

 Add Context...  service.md ×

Edit files in your workspace in agent mode



Agent ▾ Claude 3.5 Sonnet ▾



A stylized, pixelated illustration of Superman flying towards the viewer over a city skyline. He is wearing his classic blue suit with a red and yellow 'S' shield, a red cape, and red boots. His expression is serious. The background shows a city with various skyscrapers under a bright, hazy sky with soft clouds. The overall aesthetic is reminiscent of early 3D computer graphics or a low-resolution digital art style.

IA Generativa

Para o SRE



A Estrutura Ideal de um Prompt: (Os 5 Pilares)

Pense no seu prompt como a construção de uma instrução em camadas. Quanto mais camadas relevantes você adicionar, mais preciso será o resultado.



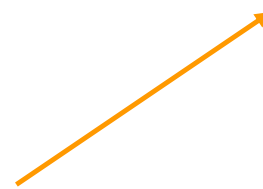
Prometheus-GEMINI Alerter



app.py



bridge.py



e como melhorar...

Fine-Tuning: É como treinar intensivamente o aluno por meses. Ele internaliza o conhecimento e muda sua forma de pensar. O processo é caro, demorado, mas o resultado é um aluno fundamentalmente mais inteligente sobre aquele tópico.

RAG (Retrieval-Augmented Generation): É como dar ao aluno um livro de consulta (uma "cola") para usar durante a prova. O aluno não muda sua inteligência base, mas ele tem acesso instantâneo à informação correta no momento exato em que precisa dela. É mais rápido, barato e fácil de atualizar.

Evolução de papéis

Curador de Dados: Garantir que o **OpenTelemetry** envie dados de alta qualidade. Lixo entra, lixo sai. A IA precisa de 'comida' boa.

Treinador de IA: Configurar as ferramentas de AIOps e, o mais importante, dar o **contexto de negócio** (nossos runbooks, nossa arquitetura) para os LLMs.

Orquestrador de Automação: Usar a IA como uma alavanca para construir sistemas que se auto-reparam (*self-healing*), em vez de apenas apagar incêndios.



OPERADOR DE TECLADO



ARQUITETO/ORQUESTRADOR

O futuro da confiabilidade não é o humano *versus* a máquina.

É o SRE, amplificado pela máquina.

Obrigada!

