

Ciclo de Deploy: Lead Time de Features Pré vs Pós-IA

Dias necessários desde o conceito técnico até o deploy em produção.

MATRIZ DE ALOCAÇÃO ESTRATÉGICA DE RECURSOS

CUSTO VS VELOCIDADE

 **Raciocínio de Alta Precisão**

Elite

Auditorias, código crítico, decisões de crédito e planejamento estratégico.

US\$ 3.00 – US\$ 4.50 / 1M tokens

 **Soberania & On-Premise**

Zero Risk

Dados sigilosos, defesa, saúde privada e conformidade estrita com LGPD.

Custo Fixo de GPU Interna

 **Automação em Massa**

Sweet Spot

Triagem de dados, suporte ao cliente, extrações estruturadas e e-mails.

US\$ 0.10 – US\$ 0.55 / 1M tokens

 **Real-Time & Voz Interativa**

< 150ms

Agentes de voz bidirecionais, busca instantânea e automações de tela.

Sub-150ms TTFT

DESTAQUE DE MERCADO

De 21 dias para 4 dias

Aceleração de 5x no time-to-market de novas funcionalidades

TI

DIRETRIZ DE ENGENHARIA

Adoção de Model Routers inteligentes reduz a fatura de computação em até 70% sem sacrificar acurácia.

CANAL TELEGRAM

t.me/oquevemai

Fontes: DORA State of DevOps Report 2026

oquevem.ai · Intelligence Hub

Ciclo de Deploy: Lead Time de Features Pré vs Pós-IA

Dias necessários desde o conceito técnico até o deploy em produção.

MATRIZ DE ALOCAÇÃO ESTRATÉGICA DE RECURSOS

CUSTO

VS

VELOCIDADE

Raciocínio de Alta Precisão

Elite

Auditorias, código crítico, decisões de crédito e planejamento estratégico.

US\$ 3.00 – US\$ 4.50 / 1M tokens

Soberania & On-Premise

Zero Risk

Dados sigilosos, defesa, saúde privada e conformidade estrita com LGPD.

Custo Fixo de GPU Interna

Automação em Massa

Sweet Spot

Triagem de dados, suporte ao cliente, extrações estruturadas e e-mails.

US\$ 0.10 – US\$ 0.55 / 1M tokens

Real-Time & Voz Interativa

< 150ms

Agentes de voz bidirecionais, busca instantânea e automações de tela.

Sub-150ms TTFT

DESTAQUE DE MERCADO

De 21 dias para 4 dias

TI

DIRETRIZ DE ENGENHARIA

Adoção de Model Routers inteligentes reduz a fatura de computação em até 70% sem sacrificar acurácia.

CANAL TELEGRAM

t.me/oquevemai

Fontes: DORA State of DevOps Report 2026

oquevem.ai - Intelligence Hub