



- Oficina -

MLOps: Do Colab à produção

Da Teoria à Prática na Implantação de Modelos de Machine Learning

Guilherme Henrique dos Reis



guilherme_reis@discente.ufg.br



/in/guilhermehreis

**~50% dos modelos de
Inteligência Artificial
NÃO ENTRAM EM PRODUÇÃO.**

- Gartner

Alguns Desafios

Apenas **treinar** um modelo **não é suficiente!**

Monitorar os Modelos em produção pode ser mais **complexo** do que desenvolver os Modelos

Difícil integração com código legado

Grandes volumes de **dados**

Dados de **treinamento** vs dados **reais**

"Paradoxo dos Pilotos de IA: lançar pilotos é fácil, mas implementá-los em escala é difícil."

• - Chirag Dekate, Analista VP na Gartner

Processo de Desenvolvimento de Machine Learning

Fase 1: Pesquisa e Desenvolvimento

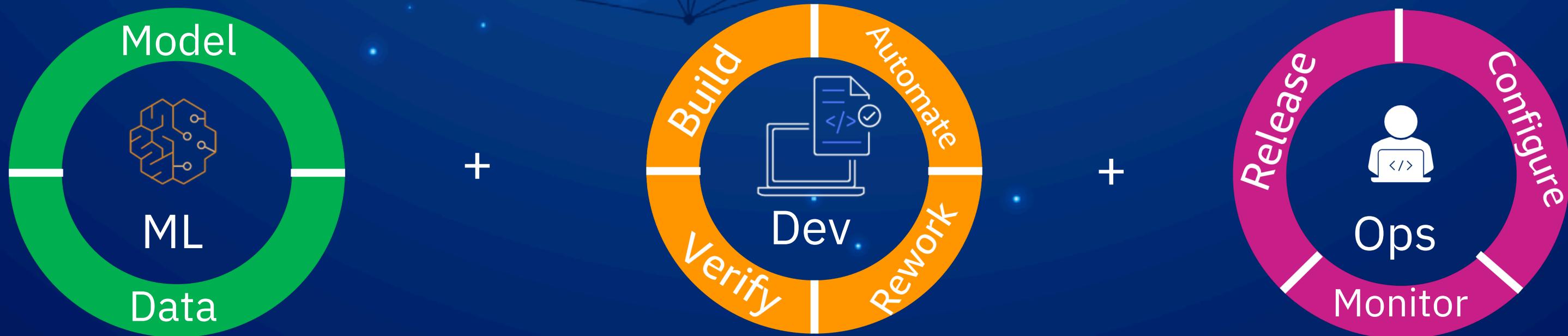


Fase 2: Operacionalização



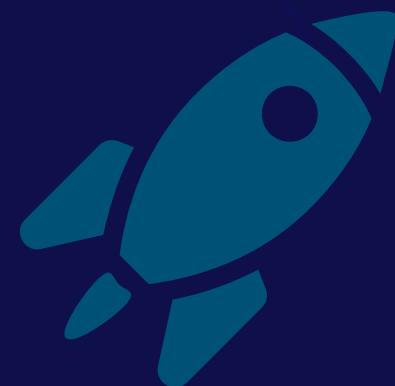
MLOps – O que?

ML + Dev + Ops = MLOps



Colaborativo e experimental por natureza | Automatizar o máximo possível |
Melhoria contínua de modelos de ML | Padronizar e Escalar

MLOps – Por que?



Agilidade

- Entregas rápidas e continuas
- Fácil de modificar
- Fácil de corrigir

Experimentos

- Experimentos rápidos e controlados
- Rápida integração de experimentos positivos

Escalabilidade

- Modularidade
- Códigos padronizados
- Custos de operação otimizados

Velocidade

- Tempo de implantação reduzido
- Planejamento mais fácil
- Controle de expectativas

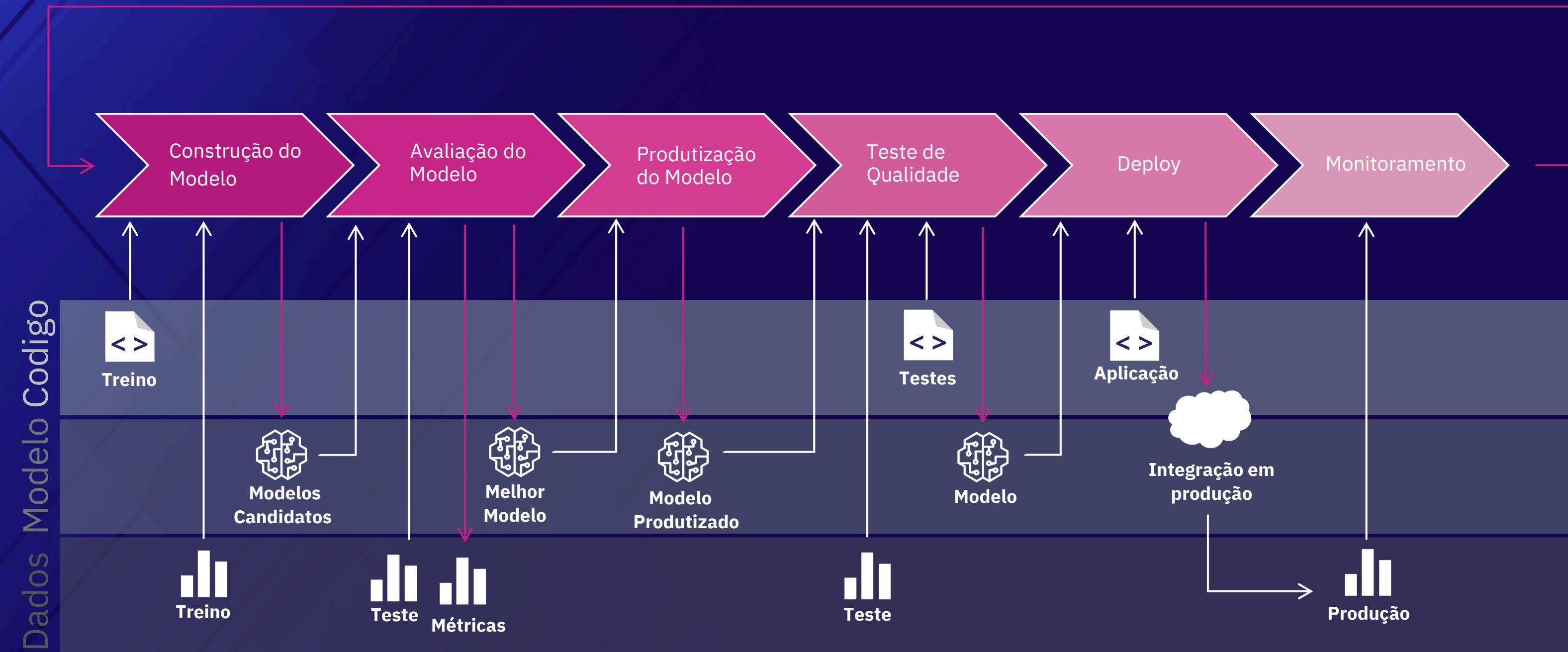
Colaboração

- Colaboração entre as equipes melhora a qualidade do produto

MLOps – Funções



MLOps – Ciclo de vida



Mão na massa!

Objetivo: Colocar um Grande Modelo de Linguagem (LLM) em produção na AWS.

- Etapa 01: Treinar o modelo
- Etapa 02: Produtizar o modelo
- Etapa 03: Fazer Deploy do modelo em nuvem (AWS)

O que vamos usar:

Etapa 01:

- Modelo Llama 3.1 8b
- Hugging Face

Etapa 02:

- Docker
- Fast API

Etapa 03:

- Amazon Web Services (AWS)
- Gradio

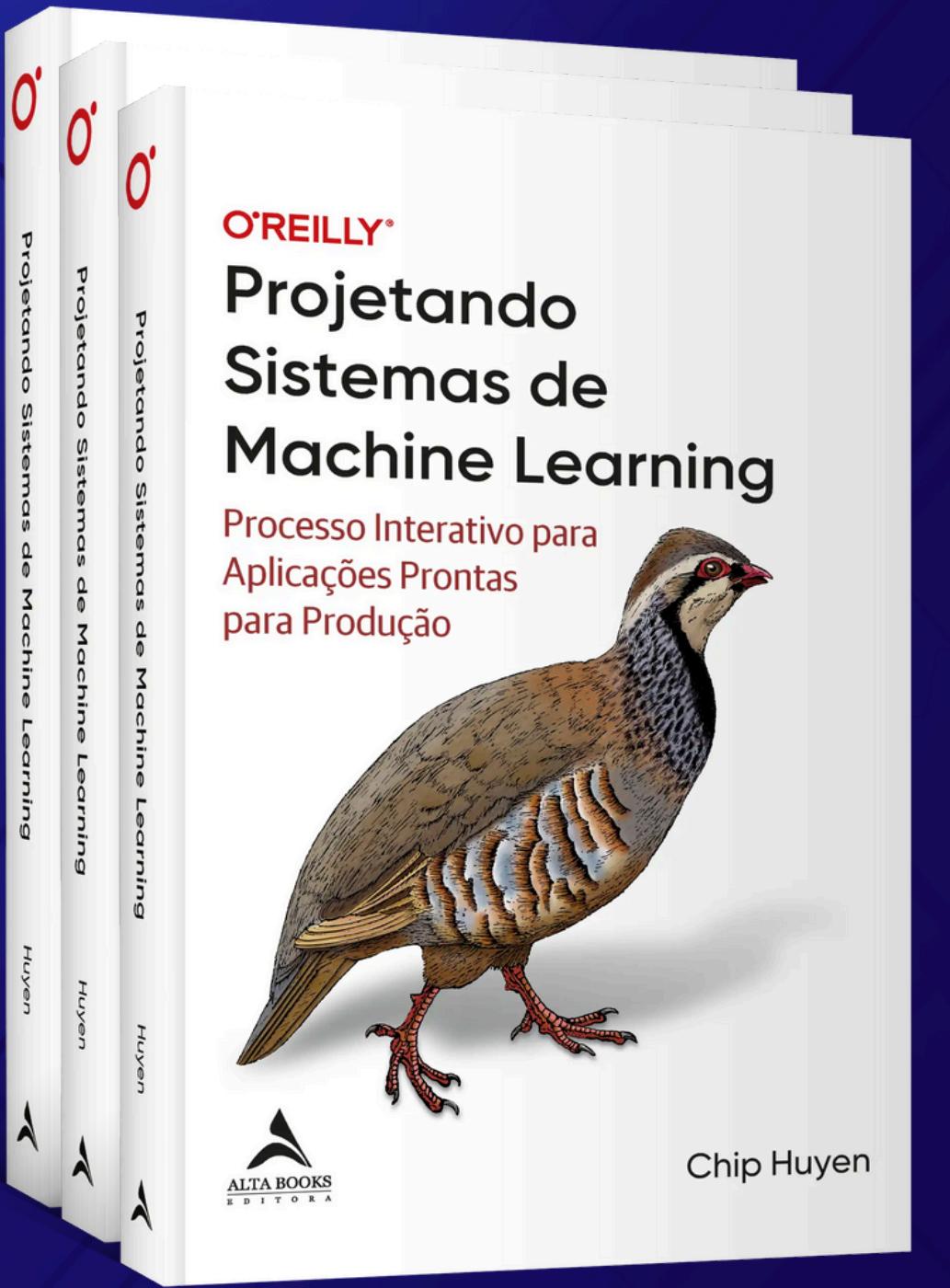
Mão na massa!

Recapitulando...



reisguilherme/OficinaMLOps: Material da oficina de MLOps do Conecta CEIA 2024 (github.com)

Para participar do sorteio:





Obrigado!



Guilherme Henrique dos Reis



guilherme_reis@discente.ufg.br



/in/guilhermehreis