

Coloque seus modelos de IA em operação: quatro elementos essenciais

Na corrida para adotar a inteligência artificial (IA) nos negócios, algumas organizações acreditam que basta implantar um modelo de IA para alcançar a linha de chegada. Na verdade, o uso de operações de machine learning (MLOps) no modelo é um processo contínuo. Os quatro elementos a seguir são os principais componentes para colocar seu modelo em operação após a implantação, idealmente usando uma plataforma de desenvolvimento de IA.

1 Monitoramento

O monitoramento eficaz ajuda a garantir que o modelo de IA continue tendo o desempenho esperado em diferentes circunstâncias. Isso é essencial para identificar e adaptar os desvios nos dados com rapidez, mantendo a confiabilidade.

Estas são as métricas comumente rastreadas nos modelos de IA operacionalizados:

- ▶ **Precisão e desempenho.** É a parcela de previsões corretas do modelo.
- ▶ **Qualidade dos dados.** Inclui testes de completude, consistência e distribuição dos dados, entre outros.
- ▶ **Utilização de recursos.** É o monitoramento do uso da CPU/GPU e da memória pelo modelo.
- ▶ **Latência e taxa de transferência.** Referem-se ao tempo que o modelo leva para gerar previsões.
- ▶ **Desvio do modelo.** Indica o desvio em relação às previsões de linha de base e o compara à precisão do modelo.

Após a coleta e seleção das métricas, a organização deve criar um framework para responder aos valores adversos de modo eficaz. Isso inclui os seguintes componentes:

- ▶ **Determinação das métricas.** Identifique as principais métricas a serem rastreadas.
- ▶ **Coleta e armazenamento.** Os processos definidos precisam ser iniciados para capturar e armazenar os dados.
- ▶ **Alertas em tempo real.** É preciso configurar o sistema para alertar as partes relevantes ou realizar uma ação, como retreinamento de pipelines, quando uma métrica monitorada sai do padrão.

2 Manutenção

Implementar a manutenção proativa dos modelos de IA é parte essencial da operacionalização deles. Isso engloba atualizar os modelos, corrigir os erros e adaptar-se às mudanças nos ambientes.

O processo de manutenção também inclui:

- ▶ **Realizar atualizações do modelo.** Defina um cronograma de testes regulares dos modelos de IA para realizar o retreinamento conforme o necessário.
- ▶ **Priorizar a correção de erros usando um processo estruturado.** Defina um cronograma para identificar e corrigir os erros no conjunto de dados do modelo.
- ▶ **Planejar o gerenciamento do ciclo de vida.** Crie pipelines de ciência de dados repetíveis para o treinamento e validação de modelos e integre-os aos pipelines de DevOps para distribuir os modelos na sua empresa.
- ▶ **Usar a automação.** Com processos de implantação automatizados, os modelos são mais consistentes, reduzindo o risco de erros, inconsistências e time to market. É importante escolher uma plataforma de automação que funcione bem com a sua plataforma de desenvolvimento de IA.
- ▶ **Adotar um controle de versões rígido.** Quando os modelos são treinados e retreinados constantemente, deve haver o controle rígido das mudanças nos dados e modelos. Consultar o conjunto de dados incorreto ou usar o modelo errado pode prejudicar toda a aplicação.

3 Retreinamento do modelo

Neste processo, o modelo é reajustado de modo proativo para aumentar a precisão.

Muitas vezes, o processo de MLOps envolve o treinamento automático de um modelo seguindo um cronograma ou como resposta a um evento.

O objetivo do retreinamento é fazer com que o modelo de IA sempre entregue os resultados mais precisos para suas necessidades empresariais específicas, reduzindo ou revertendo o desvio do modelo.

Neste processo, os cientistas de dados realizam análises detalhadas para estabelecer as métricas de desempenho e degradação, identificando os ajustes necessários e aplicando verificações, como testes A/B. Os sistemas de monitoramento de modelos com análises e alertas em tempo real permitem iniciar o retreinamento automaticamente.

Para estabelecer um processo de retreinamento estável, escolha uma plataforma que permita criar pipelines de ciência de dados repetíveis. Deve ser possível integrá-los aos pipelines de DevOps na sua organização.

4 Governança

A governança envolve estabelecer diretrizes éticas e práticas de segurança, além de aderir aos requisitos regulatórios. Esse processo não é opcional.

É importante garantir que os modelos atendam aos limites éticos, regulatórios e de segurança definidos em sua criação. Além disso, a implantação repetida e o retreinamento não devem causar desvios fora desses limites.

Os principais componentes da governança são:

Governança de dados, que analisa a origem, coleta, precisão e atualidade dos dados, além de verificar se eles expõem informações pessoais ou privadas.

Governança dos processos, que diz respeito à formalização dos processos de MLOps. Em geral, essa etapa é mais relevante nos setores com requisitos regulatórios rígidos, como finanças.

Para aderir à governança, é essencial:

- ▶ O uso de documentações de modelo e relatórios regulares e abrangentes com controle de versão.
- ▶ Auditorias dos sistemas de IA e ML programadas ou automatizadas.
- ▶ Um sistema abrangente para a documentação dos dados.
- ▶ O gerenciamento rígido dos metadados de IA e ML.
- ▶ Um processo para validar os modelos de IA.

Para configurar um processo de MLOps e colocar os modelos de IA em operação, é preciso ter uma plataforma de desenvolvimento de IA projetada com base nos fluxos de trabalho de MLOps. A Red Hat® Consulting ajuda a identificar as necessidades empresariais da sua organização.

Leia mais

Descubra como a colaboração da [Red Hat com a MLOps Foundation](#) orientou o projeto da solução e agende uma discovery session gratuita.

Veja como

Aprenda a desenvolver e implantar aplicações de inteligência artificial e machine learning no Red Hat OpenShift® AI com [nosso curso robusto](#). O treinamento pode ser presencial ou virtual, dependendo das suas necessidades.



Sobre a Red Hat

A Red Hat ajuda os clientes a definir padrões entre diferentes ambientes e a desenvolver aplicações nativas em nuvem, além de integrar, automatizar, proteger e gerenciar ambientes complexos com serviços de consultoria, treinamento e suporte [premiados](#).

f facebook.com/redhatinc
X @redhatbr
in linkedin.com/company/red-hat-brasil

AMÉRICA LATINA
+54 11 4329 7300
latammktg@redhat.com

BRASIL
+55 11 3629 6000
marketing-br@redhat.com