

Introdução à IA empresarial: um guia para iniciantes



Red Hat

Sumário



Introdução

página 3

Capítulo 1

A ascensão da IA

página 6

Capítulo 2

Como escolher o modelo de IA certo

página 9

Capítulo 3

O que você precisa para começar

página 12

Capítulo 4

Adote e escale com a Red Hat

página 16

Mais informações

É hora de avançar na sua jornada de adoção da IA

página 20

Introdução

As organizações estão cada vez mais cientes das oportunidades que a inteligência artificial (IA) oferece em todos os aspectos dos negócios.

Os casos de uso de IA estão se expandindo e ganhando espaço em todos os setores, como engajamento de clientes, suporte, vendas, infraestrutura de TI, processos, desenvolvimento de código e entrega de soluções.

Conforme o IDC, a previsão é que esse mercado ultrapasse os US\$ 423 bilhões até 2027, com uma taxa de crescimento anual composta (CAGR) em cinco anos de 26,9%. Muitas empresas devem concentrar as iniciativas de IA na melhoria da eficiência operacional, da experiência do cliente e da produtividade¹.

Diante dessa rápida evolução, as lideranças se encontram sob pressão para identificar, escolher, criar e entregar soluções de IA que darão uma vantagem competitiva às suas organizações. No entanto, a velocidade da inovação em IA e a capacidade da maioria das organizações de se desenvolver nessa tecnologia caminham em ritmos diferentes. Por isso, em muitos casos, aproveitar todo o valor que a IA tem a oferecer é um verdadeiro desafio que gera mais perguntas do que respostas.



O objetivo deste e-book é responder a muitas das dúvidas sobre a IA no estágio atual, seja para quem está começando na jornada, quer entender mais sobre o impacto futuro dessa tecnologia nos negócios ou precisa descobrir como escalar implementações de IA já realizadas.

Quais são os tipos de IA?

Para aproveitar todo o potencial da IA, você precisa saber tudo sobre essa tecnologia. Estes são os dois principais tipos que as organizações estão adotando:

IA Preditiva: esse tipo de IA ajuda as organizações a identificar padrões e tomar decisões sobre o futuro com base em dados históricos. Os modelos preditivos são úteis em aplicações relacionadas a previsão de demanda, manutenção preditiva e planejamento operacional. A IA Preditiva usa técnicas consagradas de ciência de dados e machine learning (ML), permitindo que ela seja aprimorada conforme aumenta o volume de dados processados.



IA Generativa (gen AI): usando modelos de deep learning, como os transformadores, a **IA generativa** pode criar conteúdo novo, como textos, imagens e códigos. Esse tipo de IA é útil principalmente em aplicações como de chatbots, geração automatizada de conteúdo e ferramentas criativas. Modelos como os transformadores generativos pré-treinados (GPTs) revolucionaram o processamento de linguagem natural e os campos criativos com a produção de textos e imagens semelhante à humana.

Quais são os benefícios das implementações de IA?

Ainda não sabemos tudo o que a IA é capaz de fazer. No entanto, entender como essa tecnologia em rápida evolução já está sendo benéfica para organizações de todos os portes de diversos setores ajuda a determinar onde integrar a IA nos seus negócios.

Pense em como os seguintes aspectos da IA podem beneficiar sua organização:

Volume de dados: com o crescimento exponencial dos dados, as organizações muitas vezes têm dificuldade em gerenciar e extrair insights da imensa quantidade de informações que coletam. A IA pode processar e analisar grandes conjuntos de dados rapidamente para revelar tendências e insights úteis que seriam difíceis de identificar com o esforço manual.

Ineficiência operacional: muitas organizações entendem que processos ineficientes e obstáculos podem prejudicar a produtividade e, por consequência, exigir mais tempo e esforço para resolver os problemas. A automação que usa IA ajuda a simplificar as operações, reduzindo os erros e tornando os processos mais eficientes. Isso pode incluir, por exemplo, aplicações simples como gerar automaticamente atas de reuniões com ações e próximas etapas claras ou acelerar a criação de imagens e vídeos para sites ou redes sociais.

Expectativas dos clientes: os clientes esperam ter experiências personalizadas e descomplicadas. Ao analisar os dados do cliente e oferecer recomendações e interações sob medida, a IA pode melhorar o atendimento e a personalização.

Competitividade no mercado: a inovação contínua é uma obrigação para continuar à frente da concorrência em um mercado que evolui rapidamente. A IA pode ajudar as organizações a se adaptarem mais rápido às mudanças no mercado e manter uma vantagem competitiva. É possível usar a IA Generativa até mesmo para aperfeiçoar a abordagem na hora de colaborar com a liderança ou se preparar para uma reunião importante.

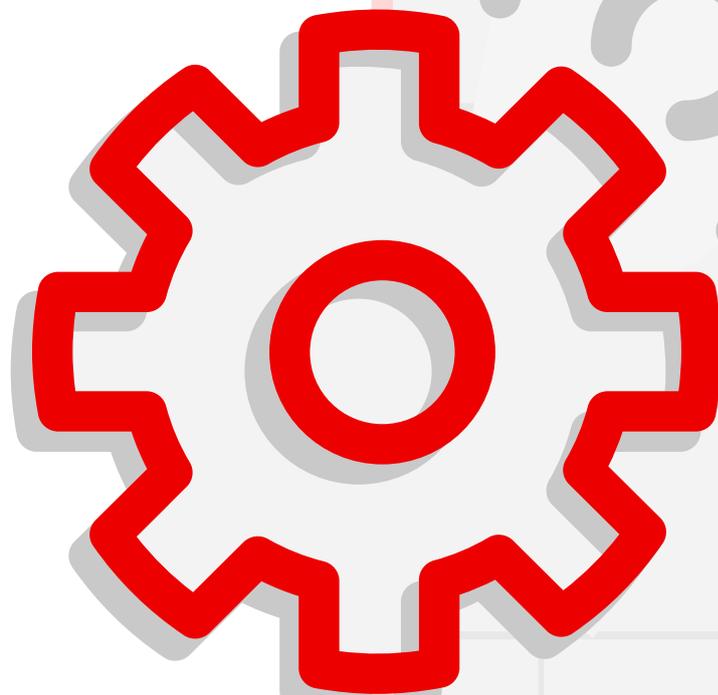
1010
11011



A ascensão da IA

A IA está em evolução há décadas, impulsionando avanços em setores como de saúde, finanças e manufatura.

No entanto, a ascensão recente da IA Generativa chamou a atenção devido à sua habilidade de criar textos próximos aos de humanos, imagens realistas e até mesmo códigos de software. Diferentemente da IA tradicional, que automatiza tarefas ou analisa dados, a IA Generativa abre caminho para a solução de problemas criativos e a criação avançada de conteúdo.



Tipos de modelos de IA que estão acelerando a inovação

Os [Large Language Models \(LLMs\)](#) e os modelos de difusão estável são alguns dos modelos de IA responsáveis pelo crescimento expressivo da IA Generativa. Os LLMs, como os GPTs, são pré-treinados com conjuntos de dados enormes. Eles podem entender e gerar linguagem natural, sendo de grande valor para a automação do atendimento ao cliente, a geração de textos de marketing e muitos outros usos. Por outro lado, os modelos de difusão estável criam imagens hiper-realistas, o que impulsiona a inovação nas áreas de entretenimento, marketing e outras.

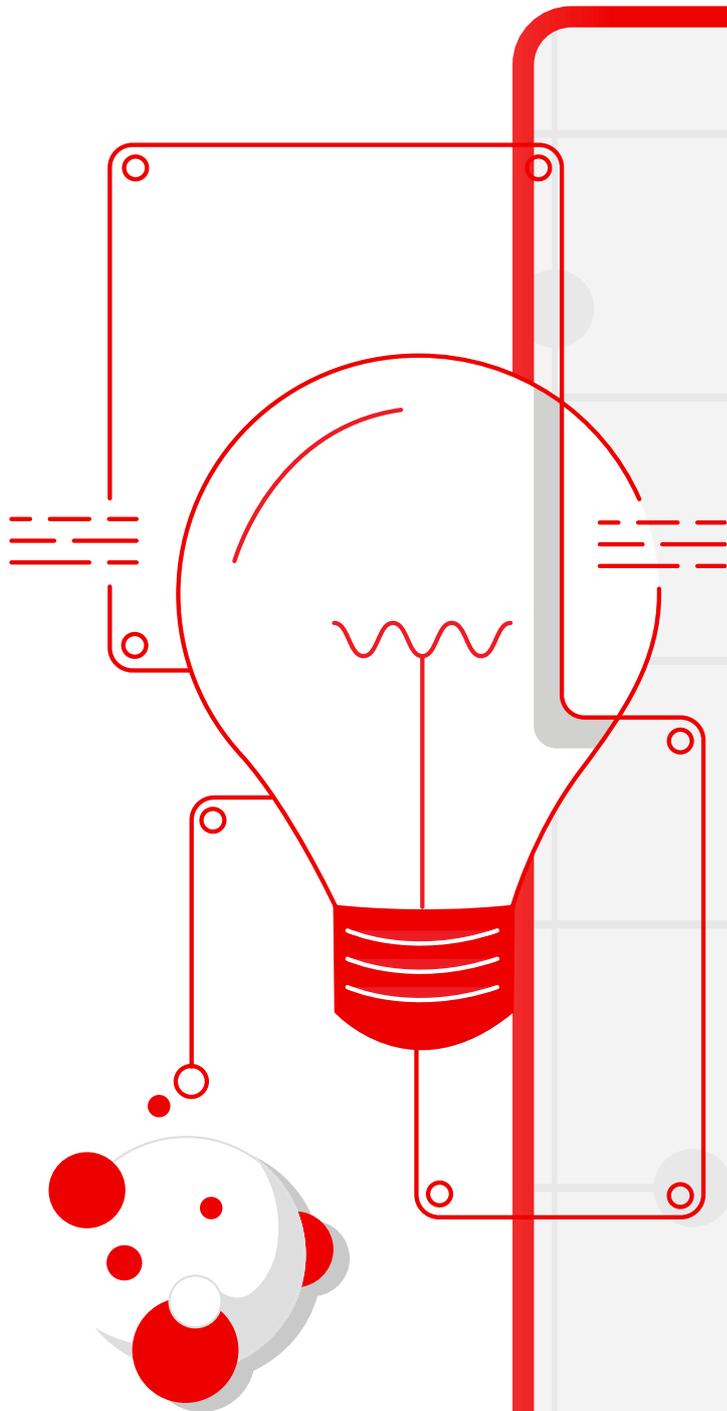
Novas tendências a considerar

As empresas estão investindo cada vez mais em IA multimodal, que combina recursos de processamento de dados, textos e imagens em um único modelo e oferece soluções mais versáteis. É essencial estar à frente dessas tendências para aproveitar o potencial total da IA no cenário empresarial.

Open source: uma base para a inovação em IA

A [estratégia de IA da Red Hat](#) é profundamente ancorada na filosofia open source para ajudar as empresas a avançar no uso da IA Generativa com transparência, confiança e custos reduzidos. As organizações que usam as plataformas de [nuvem híbrida](#) aberta da Red Hat conseguem inovar com liberdade e manter o controle sobre suas soluções de IA.

[Mais informações sobre os LLMs e como eles funcionam](#)



Assuma o controle dos LLMs com o open source

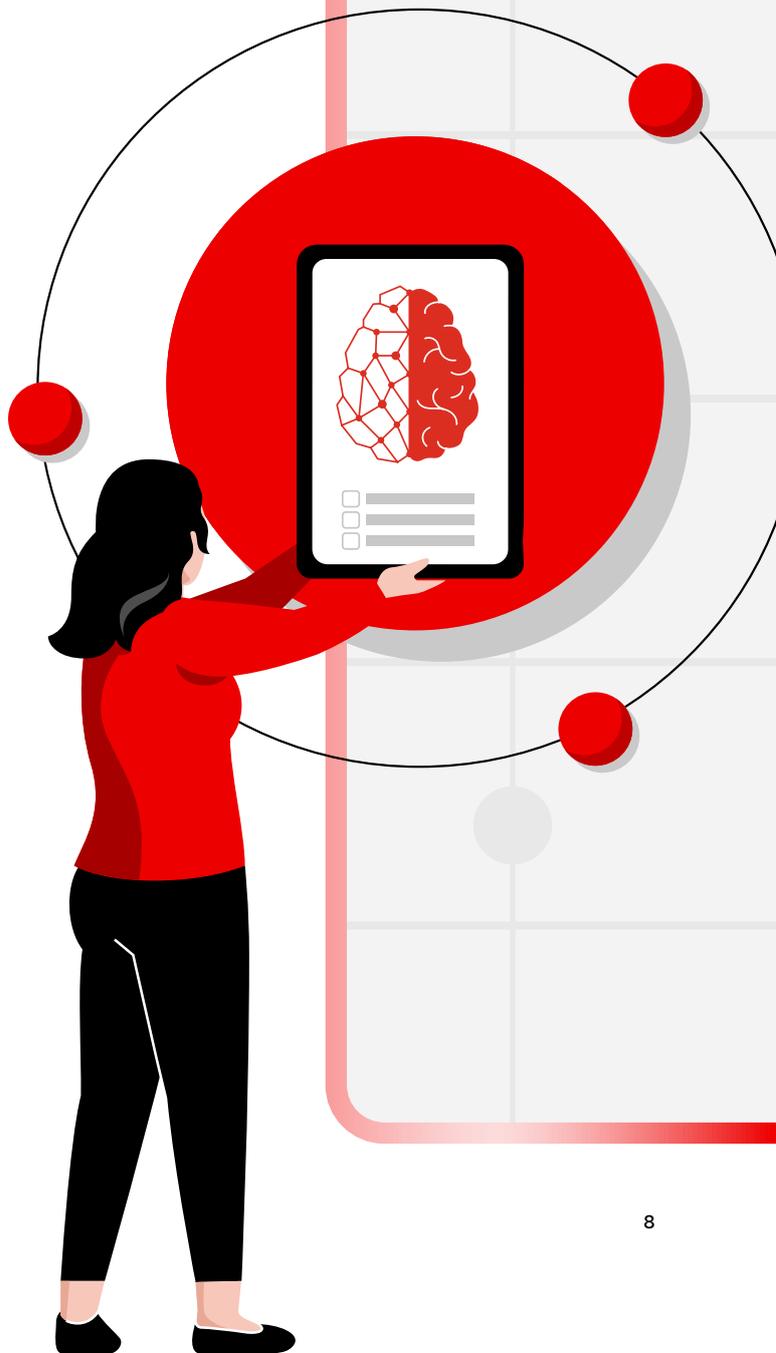
A IA Generativa está transformando quase todos os aspectos dos negócios, da criação de software à comunicação, mas os modelos (LLMs ou de outros tipos) usados nessa tecnologia muitas vezes são rigorosamente controlados pelo provedor do serviço. Isso significa que nem sempre é fácil para uma empresa avaliar tudo o que um serviço de IA Generativa pode fazer sem ter habilidades especializadas e pagar caro (em termos de dinheiro e tempo).

Em caso de falta de transparência nos conjuntos de dados que criaram o modelo ou nos detalhes sobre o uso dos dados, as empresas correm riscos em termos de conteúdo gerado por IA. E se o treinamento de um modelo de geração de código usou código-fonte com direitos autorais? Os códigos que esse modelo gerar também estarão protegidos pelos direitos autorais do código usado no treinamento? Muitas perguntas como essas ainda não foram totalmente respondidas. E, por entender as consequências, as empresas estão buscando soluções em IAs open source.

A abordagem de IA da Red Hat é ancorada na filosofia open source, assim como nosso suporte a modelos open source, como a família de modelos de base **IBM Granite**.

As soluções de IA da Red Hat até mesmo contribuem diretamente com o desenvolvimento de modelos de IA graças ao **InstructLab**, uma solução da comunidade para melhorar os recursos de LLM.

[Explore o InstructLab no GitHub](#)



Como escolher o modelo de IA certo

Modelos de IA diferentes são usados e aplicados a casos de uso variados.

É possível usar modelos de IA Preditiva e Generativa juntos em uma única aplicação ou serviço. Cada um desses modelos terá um custo diferente, mas proporcionará benefícios exclusivos e poderá ajudar as organizações a reduzir o time to market em uma prova de conceitos inicial. Os modelos de segmentação de imagem, de conversão de voz em texto e de reconhecimento de imagens são exemplos comuns e altamente capazes. Mas o importante é avaliar o que é melhor para seu caso de uso.

Os modelos de base, treinados com quantidades imensas de dados, têm capacidades altamente flexíveis. No entanto, por serem muito grandes, eles podem aumentar os custos, as demandas de gerenciamento e a complexidade. Por isso, não são adequados para todos os casos de uso.

Modelos menores e com ajuste fino (que ainda fazem parte da família da IA Generativa) podem ser uma solução mais adequada quando a melhor opção é ajustar um modelo que já existe para atender aos seus requisitos. Na hora de escolher um modelo, você pode escolher uma opção pré-criada. Esse tipo de modelo é prontamente disponibilizado e fácil de integrar aos seus sistemas. Um exemplo comum desse tipo de modelo é o LLM, uma ferramenta avançada e já treinada com grandes quantidades de dados.

No entanto, se você tiver requisitos específicos de negócios, preocupações quanto à privacidade dos dados ou o desejo de ter maior controle sobre o comportamento do modelo, talvez seja necessário criar e hospedar um modelo personalizado.

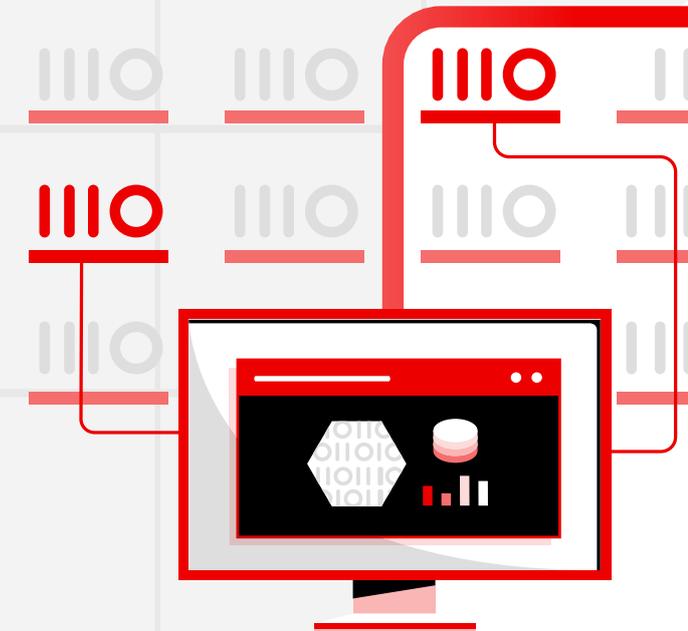
Criação e ajuste de modelos

Criar um modelo de IA do zero pode ser uma tarefa gigantesca. Você precisará coletar e preparar grandes conjuntos de dados para o caso de uso empresarial que sua organização quer solucionar. Depois, precisará escolher e treinar um algoritmo apropriado com seus dados. Como esse processo exige uma capacidade computacional significativa e considerável expertise, é um esforço que demanda muito tempo e recursos. Ao criar um modelo de base tradicional, você tem uma solução personalizada, mas esse nem sempre é o caminho mais eficiente.

Por outro lado, ajustar um modelo de base envolve adaptar um modelo pré-treinado aos requisitos específicos da organização. Uma abordagem comum é transferir o aprendizado, isto é, retreinar um modelo já treinado com um grande conjunto de dados usando um conjunto menor e específico para uma área. Seguindo esse método, o modelo é treinado de novo usando o conhecimento geral que aprendeu no treinamento inicial, mas se adaptando às particularidades dos seus dados específicos.

Ajuste fino do modelo

Outra abordagem é fazer o ajuste fino dos parâmetros do modelo para melhorar o desempenho em uma tarefa específica. Os parâmetros são as variáveis que podem ser estimadas ao adequar os dados ao modelo. Às vezes, o ajuste fino envolve alterar a taxa de aprendizado, modificar a arquitetura do modelo ou treinar algumas camadas mais do que outras. Essas técnicas ajudam a aprimorar o conhecimento do modelo para ele ser mais eficaz em um determinado caso de uso.



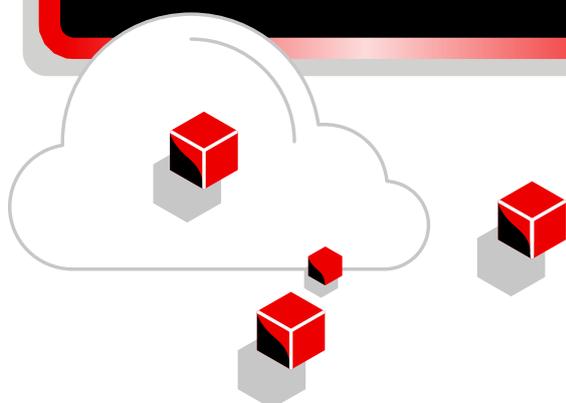
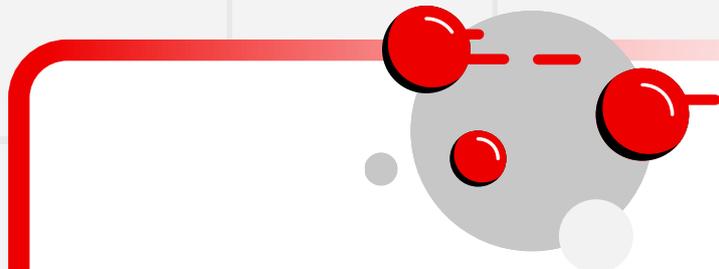
O [InstructLab](#) adota uma abordagem de ajuste fino, com o objetivo de reduzir o conhecimento prévio de IA e simplificar a inclusão de conhecimento empresarial aos modelos de base do Granite.

Alternativas ao ajuste fino de modelos

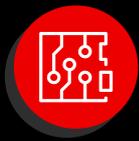
Os pesquisadores vêm buscando formas de ajustar modelos de base com mais rapidez e eficiência. Uma técnica bastante comum é a [Geração Aumentada por Recuperação \(RAG\)](#), usada para recuperar fatos de uma fonte externa na qual conhecimento adicional (contexto) foi codificado.

A RAG usa um ou mais bancos de dados (vetoriais) externos para dar mais contexto à pergunta feita ao modelo de IA Generativa. Outra abordagem em crescimento é a dos sistemas de IA Agente. Eles combinam vários agentes de IA Generativa para consultar o conhecimento de sistemas de outras entidades, como bancos de dados internos, intranets de empresas ou a internet, para alimentar os modelos de IA Generativa com as informações mais atualizadas e exatas.

O último exemplo é o de [ajuste de prompt](#). Nessa técnica, os modelos de IA recebem pistas ou prompts de front-end, como palavras extras ou números gerados por IA, que os guiam até a decisão desejada. O resultado de uma consulta usando a técnica RAG e o ajuste de prompt serve como contexto adicional para o prompt. Combinando o ajuste fino de um LLM com as técnicas de RAG e ajuste de prompt, organizações com poucos dados podem personalizar um modelo de base para uma tarefa específica.

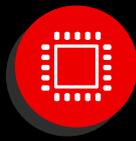


A infraestrutura que sustenta o modelo de IA é tão importante quanto o próprio modelo. Tarefas distintas exigem tipos diferentes de hardware.



Unidade central de processamento (CPU)

Processador tradicional que realiza tarefas de computação em geral. Embora versátil, nem sempre é tão eficiente para lidar com cargas de trabalho de IA em grande escala.



Unidade de processamento gráfico (GPU)

Processador especializado para lidar com tarefas de processamento paralelo. É ideal para o treinamento de modelos de deep learning que precisam processar grandes volumes de dados simultaneamente.



Unidade de processamento neural (NPU)

Um novo tipo de processador projetado especificamente para tarefas de IA. Trabalha com determinados tipos de modelos com mais eficiência e rapidez.

O papel da nuvem híbrida na adoção da IA em empresas

Os ambientes de nuvem híbrida são fundamentais para a adoção da IA. A nuvem híbrida combina infraestrutura on-premises e recursos de nuvem pública e privada, oferecendo mais flexibilidade no método e no local da implantação e do gerenciamento de cargas de trabalho de IA. Por exemplo, é possível treinar os modelos de IA com GPUs potentes baseadas na nuvem e os implantar on-premises ou na nuvem privada por motivos de segurança ou conformidade. Portanto, um dos principais pontos a considerar ao utilizar essa abordagem é a consistência das ferramentas e da plataforma escolhida.

A abordagem de nuvem híbrida aberta da Red Hat ajuda as organizações a integrarem a IA em diferentes ambientes, melhorando a consistência, a escalabilidade e a flexibilidade. Com essa abordagem, você pode gerenciar cargas de trabalho de IA em vários ambientes de nuvem (on-premises ou na edge da rede), otimizar o posicionamento de dados e promover uma migração fluida de dados. Tudo isso facilita a adoção da IA em escala empresarial.

Compreendendo os modelos de IA, os dados e a infraestrutura, você pode lidar melhor com as complexidades da adoção da IA e aproveitar o potencial dela ao máximo.

O que você precisa para começar

Assim como na adoção de qualquer tecnologia nova, a IA traz alguns desafios que as organizações precisam superar para terem sucesso.

Considere os pontos abaixo ao avaliar se a sua organização está pronta para adotar a IA e identificar o que precisa ser aprimorado para acelerar a adoção.



Avalie a qualidade e a disponibilidade dos dados.

Para implementar a IA, é essencial ter acesso a dados relevantes de alta qualidade. A qualidade dos dados é um fator crítico no treinamento de modelos de IA precisos. Por isso, é importante avaliar se, além de relevantes, seus dados estão completos e corretos.

Avalie a infraestrutura tecnológica. Determine se sua infraestrutura atual aguenta as cargas de trabalho de IA. Isso inclui avaliar se você tem ao seu dispor recursos de armazenamento, computação de alto desempenho, automação e rede.

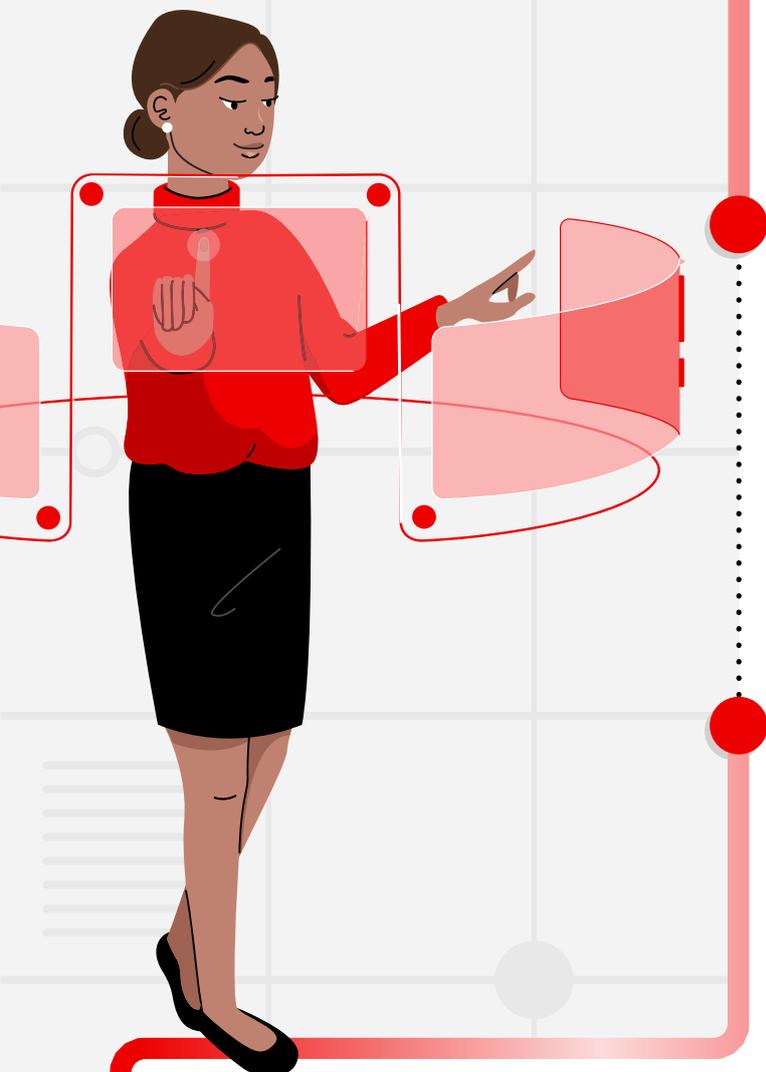
Identifique onde as habilidades são necessárias.

Avalie se existem profissionais com expertise em IA na sua organização. Pode ser necessário avaliar as competências atuais e identificar se há necessidade de treinamento ou habilidades especializadas.

Analise o alinhamento estratégico. Verifique se as iniciativas de IA estão alinhadas às metas e estratégias da empresa. Os projetos de IA devem alavancar os objetivos estratégicos gerais da empresa e gerar valor de negócio mensurável.

Como iniciar a implementação da IA

A velocidade e a escala da adoção da IA pela organização depende de muitos fatores. No entanto, começar devagar e expandir aos poucos é uma boa abordagem para quase todos os projetos de modernização tecnológica.



Confira as oito etapas para sua organização iniciar e avançar na jornada de adoção da IA:

1

Avaliar habilidades e metas

Comece com uma avaliação da capacidade, infraestrutura e metas estratégicas atuais da organização. Determine se a IA se alinha aos objetivos gerais e identifique onde essa tecnologia pode agregar valor. Essa avaliação inicial ajudará a traçar um caminho claro para a jornada de adoção da IA.

2

Identificar casos de uso e equipes de IA

Identifique o que a IA pode resolver na sua organização. Forme uma equipe exclusiva para a aplicação de IA com membros de várias funções, como desenvolvedores, experts de diferentes áreas, cientistas de dados e especialistas em TI. Ela ficará encarregada de liderar a iniciativa. Um caso de uso bem definido serve como guia para direcionar os recursos e esforços de adoção da IA.

3

Selecionar o modelo

Escolha um modelo de IA apropriado, com base no caso de uso identificado. Seja um LLM para IA Generativa ou um modelo preditivo para análise de dados, é necessário que as capacidades do modelo estejam alinhadas com as metas do caso de uso. Considere se o modelo é complexo, escalável e compatível com seus sistemas.

4

Realizar ciclos de teste e validação

Defina critérios claros para o sucesso da implementação da IA, como métricas de desempenho, taxas de acerto e objetivos empresariais. Estipule ciclos de teste e validação para avaliar continuamente a eficácia do modelo. O feedback regular desses ciclos ajudará a fazer o ajuste fino do modelo e manterá a jornada da IA no caminho certo.

5

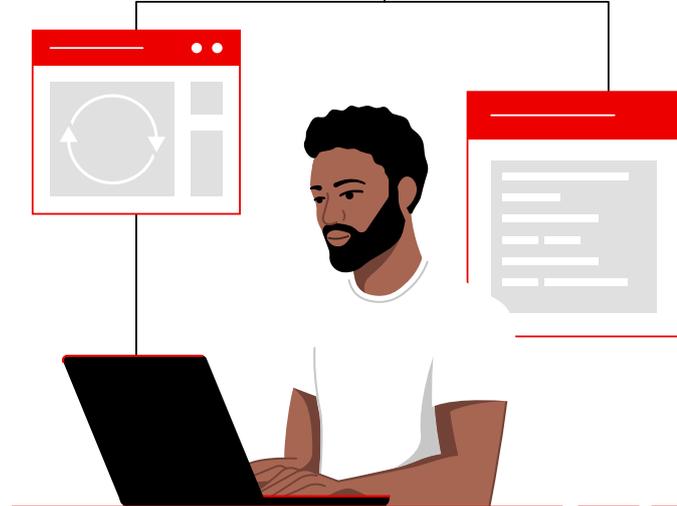
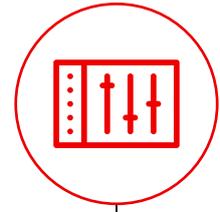
Ajustar o modelo

Personalize o modelo escolhido usando dados da sua organização. O processo de ajuste requer alimentar o modelo com dados relevantes para melhorar a precisão e a adequação a um caso de uso empresarial específico. Fazer o ajuste fino é uma garantia de que o modelo está adaptado ao contexto e às necessidades da sua organização.

6

Treinar com dados sintéticos

Considere usar dados sintéticos para treinar e aprimorar ainda mais seus modelos de IA. Essa abordagem, baseada em métodos como usar um LLM "professor" e outro "aluno", permite gerar dados de treinamento de alta qualidade quando os dados reais são escassos ou confidenciais. O treinamento com dados sintéticos também ajuda a deixar o modelo mais robusto e eficiente, sem comprometer a privacidade.



7

Monitorar desvios

O monitoramento de desvios permite fazer o acompanhamento geral com base no contexto. Em vez do gerenciamento estruturado das configurações, o monitoramento de desvios rastreia as mudanças no conteúdo no sistema de arquivos local. Isso ajuda a detectar e corrigir desvios ou declínios na precisão do modelo ao longo do tempo. Com um monitoramento contínuo, é possível conservar a eficácia e a relevância do modelo conforme as condições mudam.

8

Contratar ajuda especializada

Se sua equipe interna ainda estiver desenvolvendo a expertise em IA, contratar experts externos, como os da Red Hat® Consulting, pode ser uma grande vantagem. Os experts da Red Hat ajudam nos aspectos complexos da adoção da IA oferecendo insights úteis e acesso a treinamentos. A Red Hat Consulting pode ajudar você a acelerar a jornada da IA e aumentar suas chances de sucesso.



A adoção da IA depende da colaboração

Formar uma equipe com diversas capacidades é essencial para o sucesso dos projetos de IA Generativa²

- **Líderes empresariais:** representam as pessoas que usam ou são afetadas pela solução.
- **Especialistas em IA:** ajustam, mantêm e atualizam os modelos de IA Generativa.
- **Cientistas de dados:** fazem o processamento inicial e alimentam os modelos com dados de treinamento sem vieses e corretos.
- **Executivos de ética e conformidade:** garantem que as iniciativas de IA Generativa cumpram as normas.
- **Especialistas em operações de TI:** integram as soluções à infraestrutura e aplicam políticas de segurança.
- **Equipes de desenvolvimento e comunidades:** devem ser envolvidas no início para colaborar, criar, compartilhar e melhorar as ferramentas open source, os frameworks e as práticas recomendadas para a adoção da IA. Isso garantirá que o uso da IA esteja alinhado ao valor de negócio.

² Kearney. "[Standing up tiger teams to tackle generative AI complexity](#)", 15 de novembro de 2023.

Adote e escale com a Red Hat

O Red Hat AI oferece confiança, flexibilidade de escolha e consistência em toda a nuvem híbrida para acelerar a adoção empresarial da IA.

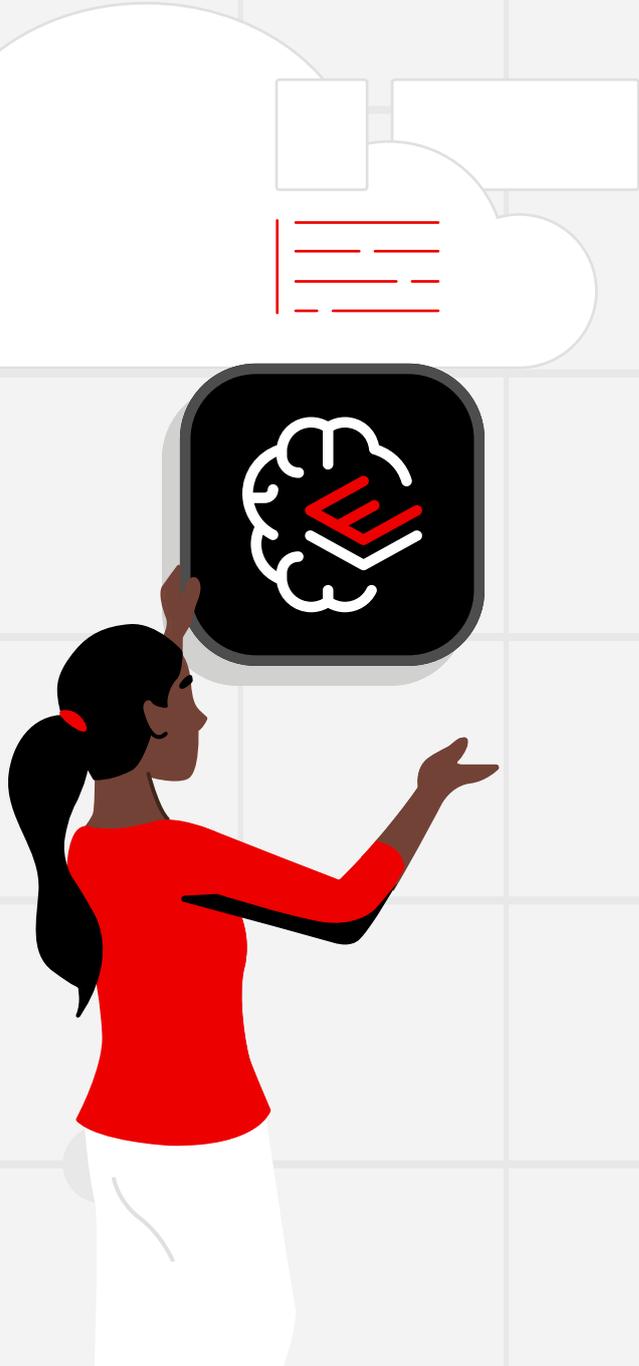
O portfólio do Red Hat AI inclui o Red Hat Enterprise Linux® AI para ambientes de servidores Linux individuais e o Red Hat OpenShift® AI para ambientes distribuídos de plataforma Kubernetes. Além disso, ele fornece recursos integrados de operações de machine learning. As duas soluções contam com modelos e tecnologias open source, ajudando as organizações a se manterem à frente da inovação em IA, acelerar o ritmo de descoberta e democratizar o acesso a ferramentas e tecnologias de ponta.

A Red Hat também conta com um extenso ecossistema de parceiros para potencializar ainda mais suas capacidades em IA. Por exemplo, a NVIDIA, uma das grandes empresas que oferecem soluções de inteligência artificial e famosa por popularizar as GPUs, mantém uma parceria com a Red Hat para o aproveitamento total dessa tecnologia com uma plataforma empresarial otimizada de ponta a ponta para cargas de trabalho de IA. A NVIDIA ajuda os clientes empresariais na adoção da computação acelerada por GPU para aplicações de alto desempenho e IA.

"A Red Hat e a NVIDIA têm uma longa história de estreita colaboração, e o Red Hat Enterprise Linux AI é a prova do nosso compromisso em oferecer software e computação de todo o stack a desenvolvedores e pesquisadores que estão impulsionando a próxima onda de aplicações e tecnologias de IA."³

Justin Boitano, vice-presidente de soluções empresariais na NVIDIA

³ Press release da Red Hat. "[Red Hat entrega inovação acessível de IA generativa open source com Red Hat Enterprise Linux AI](#)", 7 de maio de 2024.



Red Hat Enterprise Linux AI em detalhes

O Red Hat Enterprise Linux AI é formado por quatro componentes básicos:

1

Modelos Granite open source

O Red Hat Enterprise Linux AI inclui modelos Granite open source com suporte total da Red Hat. Esses modelos flexíveis ajudam a criar modelos de linguagem personalizados e usá-los publicamente ou em privado.

2

Alinhamento de modelos do InstructLab

O InstructLab é um projeto open source conduzido pela Red Hat e a IBM. Ele ajuda a personalizar modelos de IA com conhecimento específico e gerar dados sintéticos para treinamento. Por ser uma ferramenta de linha de comando integrável a repositórios git, os usuários podem adicionar habilidades e treinar modelos com facilidade.

3

Red Hat Enterprise Linux inicializável otimizado para os modelos Granite e o InstructLab

Os modelos Granite e as ferramentas do InstructLab são executados em uma imagem especializada do Red Hat Enterprise Linux, otimizada para a IA e compatível com praticamente todos os tipos de hardware e ambientes de nuvem. Essa configuração viabiliza desempenho eficiente com GPUs avançadas, necessárias para agilizar o treinamento e a implantação do modelo.

4

Suporte empresarial e indenização

As subscrições do Red Hat Enterprise Linux AI incluem suporte empresarial, ciclo de vida completo da solução, começando pelo software e modelo Granite 7B, e indenização da Red Hat por violação de propriedade intelectual.

O Red Hat Enterprise Linux AI ajuda a transformar projetos de aplicações de IA Generativa em realidade

Para organizações iniciantes em IA Generativa, o Red Hat Enterprise Linux AI oferece LLMs e modelos de linguagem de código prontos para o uso em um ambiente de servidor único para desenvolvimento e inferência.

Isso significa que há um ambiente unificado com modelos e ferramentas. Assim, é fácil começar a usar a IA Generativa e personalizar modelos com seus dados empresariais, sem precisar de uma expertise extensa ou uma infraestrutura próprias para IA.

Com suporte e indenização totais oferecidos pela Red Hat, o Red Hat Enterprise Linux AI reduz os riscos. Essa solução também inclui uma abordagem simplificada à IA Generativa, pois é mais acessível a desenvolvedores e experts de diferentes áreas. Com ela, é possível colaborar e acelerar os resultados empresariais.

Por que escolher o Red Hat Enterprise Linux AI?

LLMs para empresas

Os LLMs IBM Granite sob a licença open source Apache-2.0 estão incluídos na solução, com suporte e indenização por propriedade intelectual completos oferecidos pela Red Hat.

Colaboração da comunidade

O InstructLab simplifica a experimentação e o ajuste do alinhamento dos modelos de IA Generativa.

Escalabilidade nativa em nuvem

O image mode do Red Hat Enterprise Linux permite o gerenciamento da plataforma de IA como imagem de container, otimizando sua abordagem para escalar.

Aceleração e ferramentas de IA

Resultados mais rápidos com aceleradores de hardware open source e funcionalidades de deep learning otimizadas.

Escale com o Red Hat OpenShift AI

O Red Hat OpenShift AI fornece uma plataforma de MLOps integrada para criar, treinar, ajustar, implantar e monitorar aplicações habilitadas por IA e modelos de base e preditivos em escala em todos os ambientes de nuvem híbrida.

O Red Hat OpenShift AI é baseado no Red Hat OpenShift para oferecer uma experiência uniforme, otimizada e automatizada para atender às exigências de carga de trabalho e desempenho em projetos de inteligência artificial e machine learning. As práticas de MLOps podem ajudar as organizações a responder com agilidade às inovações de IA e a colocar em produção aplicações de IA mais rapidamente.

Pratique no sandbox do Red Hat OpenShift AI

Com componentes selecionados do Open Data Hub e outros projetos open source, o Red Hat OpenShift AI oferece a cientistas de dados e desenvolvedores uma poderosa plataforma de inteligência artificial e machine learning, híbrida e aberta, para obter insights por meio de dados e criar aplicações com IA.

[Faça a avaliação gratuita no sandbox do desenvolvedor >](#)

Por que escolher o Red Hat OpenShift AI?

Escalabilidade na disponibilização de modelos

Os modelos podem ser usados para a integração com aplicações inteligentes on-premises, na nuvem pública ou na edge. Eles podem ser recriados, reimplantados e monitorados com base nas alterações feitas no notebook de origem.

Menos tempo de gerenciamento da infraestrutura de IA

Conceda acesso sob demanda aos recursos. Assim, suas equipes podem se concentrar em explorar dados e criar aplicações que agreguem valor para sua organização.

Ferramentas de inteligência artificial e machine learning (IA/ML) testadas e compatíveis

A Red Hat utiliza a plataforma de aplicações Red Hat OpenShift para monitorar, integrar, testar e oferecer suporte às ferramentas de inteligência artificial e machine learning e à disponibilização de modelos.

Flexibilidade em toda a nuvem híbrida

Disponibilizado no Red Hat OpenShift como um software autogerenciado ou um serviço em nuvem totalmente gerenciado, o Red Hat OpenShift AI oferece uma plataforma segura e flexível na qual você pode escolher onde desenvolver e implantar seus modelos: on-premises, na nuvem pública ou até mesmo na edge.

Operações com nossas práticas recomendadas

A Red Hat Consulting oferece serviços que permitem instalar, configurar e utilizar todo o potencial do Red Hat OpenShift AI. A Red Hat Consulting oferece suporte e orientação tanto para uma experiência piloto com o Red Hat OpenShift AI quanto para a criação da sua base de MLOps.



A Red Hat oferece um portfólio completo de tecnologias, expertise comprovada e parcerias estratégicas para ajudar você a alcançar suas metas com a IA Generativa. Tenha uma base para o desenvolvimento e a implantação de modelos e aplicações de IA Generativa, além de serviços e treinamento para acelerar a adoção.

Mais informações

É hora de avançar na sua jornada de adoção da IA

Acelere a adoção da IA com a estratégia de nuvem híbrida aberta da Red Hat. Você ganha flexibilidade para executar suas aplicações de IA onde forem necessárias.

Dê início aos seus projetos de inteligência artificial e machine learning com os serviços de expertise, consultoria e treinamento da Red Hat para ajudar sua organização a chegar longe com a IA.

Conheça os serviços de inteligência artificial e machine learning: red.ht/aiml-consulting

Agende uma discovery session gratuita: red.ht/consultoria

Mais informações sobre as soluções do Red Hat AI



[Leia mais sobre o Red Hat Enterprise Linux AI](#)



[Descubra como escalar com o Red Hat OpenShift AI](#)