

Insights dos especialistas

IBM Institute for
Business Value

Lidando com o dilema dos dados de IA

A abordagem certa para integração,
governança e ferramentas



Especialistas no assunto



Michael Haydock

IBM Fellow, Vice-presidente e Cientista Chefe da Global Business Services, IBM
mhaydock@us.ibm.com
linkedin.com/in/drmikehaydock

Michael é um especialista líder em inteligência artificial com o uso do deep learning para IBM Global Business Services (GBS). Ele se especializou em otimização numérica e áreas de aprendizado de máquina quantitativo, aplicando essas tecnologias exponenciais ao entendimento do relacionamento com o cliente e à otimização da cadeia de suprimentos para empresas. Michael tem Ph.D. em Pesquisa Operacional pela Universidade Walden, Minneapolis, MN.



Steven Eliuk

Vice-presidente, IA e Automação de Governança, Global Chief Data Office (GCDO), IBM
Steven.eliuk@ibm.com
linkedin.com/in/m3the01

Steven lidera o desenvolvimento de componentes de plataforma e infraestrutura para IA na Cognitive Enterprise Data Platform, o data lake da IBM. Ele também aplica a IA corporativa em processos internos, mantendo a governança, segurança, privacidade e confiança. Suas equipes também desenvolvem, mantêm e operam o conjunto global de Serviços de Privacidade da IBM, incluindo GDPR, CCPA e outros requisitos regulamentares. Antes de ingressar na IBM, Steven liderou o projeto de infraestrutura de computação de alto desempenho (HPC) para inteligência artificial na Samsung Research America. Steven obteve um Ph.D. em Ciência da Computação pela University of Alberta.



Susara van den Heever

Diretora, Cloud and Cognitive Software Expert Labs
svdheever@fr.ibm.com
linkedin.com/in/susaravandenheever

Susara atua como Diretora da Expert Labs EMEA, com foco na criação de valor para o cliente por meio do IBM Cloud and Cognitive Software. Sua paixão é usar dados e IA para tornar o mundo um lugar melhor. Ela trabalha para ajudar pessoas e empresas a serem mais eficientes, economizar tempo, gerenciar melhor os recursos e até mesmo fazer melhores escolhas ambientais. Sua carreira abrangeu muitos setores: desde ajudar um produtor de cerveja a criar uma melhor cadeia de suprimentos, garantir a alocação justa de largura da banda para provedores de telecomunicações, ajudar empresas de energia a criar uma produção sustentável, até tornar as cidades lugares melhores para se viver. Susara tem Ph.D. em Engenharia Química pela Carnegie Mellon University.

Por que cerca de metade dos projetos de IA falham? Em uma palavra, dados.

Principais conclusões

Para obter o máximo valor da IA, sempre comece com o problema de negócios. Em seguida, busque vários tipos de dados, estruturados e não estruturados, internos e externos, qualitativos e quantitativos, para enfrentar o problema e enriquecer a solução.

Incorpore governança robusta e baseada em permissão que estabelece a procedência dos dados para construir confiança nos dados e nos insights de IA.

Planeje os desafios da preparação rigorosa de dados e as complexidades da fusão de fontes de dados díspares. Reutilize dados, automatize processos e adote as ferramentas certas.

Os desafios de dados exclusivos da IA

A inteligência artificial (IA), que não é mais uma novata na área (lugar ocupado pela computação quântica), está sendo rapidamente adotada e incorporada em todos os tipos de negócios e usos sociais. No início da pandemia, 84% de todas as organizações esperavam manter ou aumentar o nível de foco organizacional em IA, com quase 1/3 aumentando seus investimentos em IA como resultado direto da pandemia.¹

Em uma pesquisa recente, a IA juntou-se à Internet das Coisas (IoT) e à computação em nuvem como as três principais tecnologias que os CEOs esperam que ajudem a entregar resultados.² 43% dos profissionais globais de TI relataram que a sua empresa acelerou o desenvolvimento da IA como resultado da pandemia de COVID-19.³

Para as empresas, isso se traduz em uma necessidade premente de desenvolvimento de uma capacidade estratégica de IA (da estratégia ao modelo operacional e ao talento) e integração dessa capacidade nos negócios.

Ainda assim, muitos projetos de IA estagnam, mesmo depois de uma fase promissora de prova de conceito (PoC). 90% das empresas têm dificuldade em escalar a IA em suas empresas.⁴ Portanto, não é surpreendente que cerca de metade dos projetos de IA falhem.⁵

Por quê? Em uma palavra, dados. Mais da metade dos responsáveis pela estratégia de IA reconhece que não sabe com clareza quais são suas necessidades de dados de IA.⁶ E 39% dos profissionais de TI relatam que analisar dados para criar e escalar IA confiável é a parte mais difícil da jornada de IA de sua organização, enquanto 32% dizem que a complexidade e os silos de dados são a principal barreira para a adoção de IA.⁷

Não surpreendentemente, mais da metade das organizações cita os dados como culpados pela paralisação dos projetos de IA: o principal fator citado (58%) são os problemas de qualidade dos dados, seguido pela falta de dados bem organizados (45%), e problemas de governança de dados (40%).⁸ Na verdade, mesmo as melhores práticas gerais para recursos de dados avançados podem ser insuficientes para IA. Embora a IA possa ser apenas um caso de uso para dados (ainda que um importante), as lições aprendidas precisam ser estendidas e expandidas no domínio da IA.

A IA tem várias considerações de dados específicos que fazem diferença na prática:

- *Escala de IA*: com IA, em geral, quanto mais dados disponíveis, melhor será a qualidade e a precisão dos resultados. Portanto, o volume de dados necessário para IA pode estar muito além até mesmo para algumas análises avançadas.
- *Velocidade de IA*: os dados devem estar atualizados para atingir a velocidade de resposta necessária para alguns insights de IA e para as melhores previsões. Às vezes, eles até podem estar em tempo real ou muito próximo disso.
- *Variabilidade de dados*: os resultados de IA geralmente são melhores com dados que podem incluir contexto, não apenas com um volume maior de dados. No entanto, o resultado de um modelo de IA pode ser mudado por meio da manipulação dos dados (de forma maliciosa ou não intencional), ou mesmo apenas em decorrência de um “desvio de dados” generalizado, portanto, certificar-se de que todos os dados sejam tratados adequadamente torna-se essencial.
- *Qualidade dos dados*: como a IA é altamente sensível aos dados, é fundamental que os dados correspondam à realidade subjacente. Em alguns casos, a IA observa atentamente os picos que outros métodos analíticos tradicionais podem considerar anômalos, portanto, a precisão é essencial.
- *Perspectiva humana*: a maneira como os humanos consideram os dados, incluindo as distorções de nossas próprias experiências e as turvas “caixas pretas” de nossos cérebros, geralmente afeta o modo como eles são usados. Os dados têm significado no contexto e, portanto, devem ser vistos e entendidos dessa forma. Sem o contexto adequado, eles podem ser mal utilizados ou mal interpretados involuntariamente.

Muitas vezes, porém, as empresas são sobrecarregadas por essas complexidades e acabam presas em abordagens bem-intencionadas, mas inadequadas, para resolver os desafios de dados. Para atender às necessidades de dados exclusivas da IA, uma organização precisa primeiro ter uma visão clara dos problemas de negócios que está tentando resolver e, em seguida, de uma abordagem pragmática para resolvê-los.

Como identificar o problema de negócios

Alguns especialistas em dados com treinamento clássico ainda têm dificuldades com IA e acabam se concentrando demasiadamente nos detalhes da ciência e da engenharia de dados, em *como* fazer IA. Sem compreensão e consideração suficientes dos maiores problemas dos negócios, o *motivo* pelo qual a empresa está investindo em IA, as provas de conceitos e os projetos de pesquisa podem proliferar sem beneficiar os negócios. Além disso, os cientistas e engenheiros de dados geralmente cedem à adoção de “soluções de big data” para problemas de negócios que às vezes podem ser resolvidos com “soluções de small data” de alta qualidade, precisas e até mesmo *obtidas qualitativamente*.

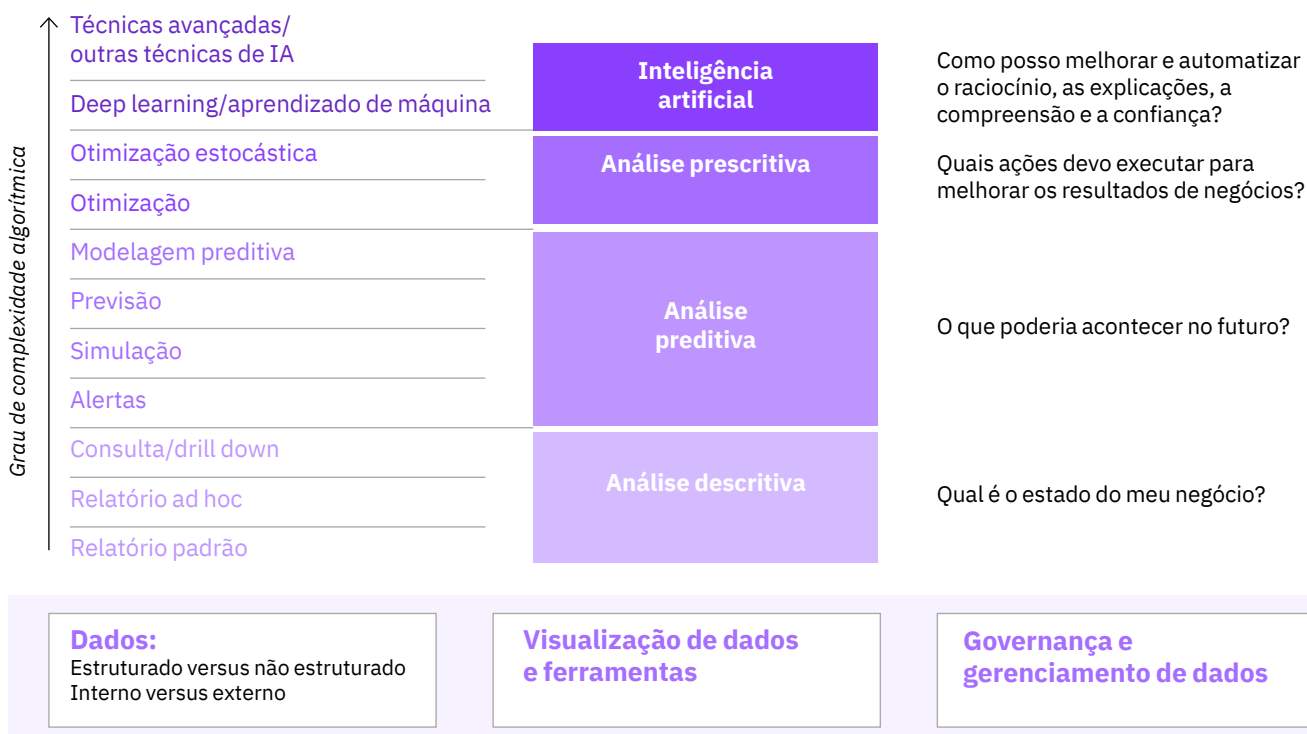
Primeiro, as equipes precisam fazer duas perguntas básicas: qual problema de negócios estamos tentando resolver? E qual a melhor maneira de resolvê-lo? Às vezes, a IA mais sofisticada pode não ser a melhor resposta (consulte a Figura 1). Isso também pode ajudar a esclarecer quando e como as intervenções do fluxo de trabalho de negócios devem ocorrer com base nos insights que os dados revelam e nas ações que a IA recomenda ou executa.

Gere resultados rapidamente para manter a empresa interessada e focada.

Figura 1

Necessidades correspondentes

Selecionando a ferramenta certa do AI-analytics continuum.



Uma abordagem pragmática para gerar valor

Dada a importância dos dados para a IA e os muitos desafios em obter, integrar, preparar e gerenciá-los adequadamente, um impulso comum para muitas organizações é lançar um único projeto para organizar todos os seus dados. Frequentemente, isso envolve colocá-los em um grande data lake e tentar resolver todos os problemas de dados da organização de uma vez por todas. Claro, isso raramente funciona.

Essa ambição é impraticável devido ao seu escopo confuso, que impede um caminho claro para um retorno sobre investimento (ROI) razoável. Mas, ainda mais importante, os dados e as necessidades de negócios mudam muito rapidamente para tornar aconselhável uma tarefa tão grande.

Em vez disso, as organizações que constroem um conjunto sólido de recursos de IA aprenderão a pensar pragmaticamente (consulte “The Weather Company: Lições aprendidas com o negócio de previsões climáticas”). Quais dados estão prontamente disponíveis, acessíveis e limpos? Uma vitória rápida que mostre resultados de negócios manterá a empresa interessada e focada. Usar dados com problemas de qualidade ou disponibilidade é arriscado, pois pode transformar uma potencial vitória rápida em um exercício de limpeza de dados longo e improdutivo.

Organizações de sucesso terão como alvo casos de uso orientados a negócios com um foco de curto ou médio prazo, mas manterão um foco de longo prazo na entrega contínua de valor de IA.

Dito isso, muitas das áreas de foco generalizadas para dados não são novas; algumas têm até décadas de idade, mas readquirem urgência, importância e oportunidade para melhorias na era da IA:

The Weather Company: Lições aprendidas com o negócio de previsões climáticas

A The Weather Company (TWC) foi construída para colocar dados (“toneladas” deles) para funcionar. Mais especificamente, aplicando IA e modelos de todos os tipos para gerar previsões e insights. A empresa tem trabalhado com IA por duas décadas e é líder global em aplicá-la em grande escala. Ao longo do caminho, a empresa aprendeu muitas lições sobre como atender aos requisitos de dados peculiares da IA.

Um dos primeiros é não subestimar a tarefa em mãos. Pode parecer simples mesclar dados meteorológicos e de vendas, mas não é. O insight obtido com a experiência é necessário ao extrair dados, saber como realizar as junções de dados necessárias e compreender as implicações de como os resultados serão executados e entregues. Por exemplo, os modelos precisarão ser executados centralmente, na nuvem ou “na borda” no dispositivo de um usuário final? Eles serão treinados on-line e, portanto, atualizados continuamente, ou off-line e atualizados periodicamente? As respostas têm implicações sobre como lidar com os dados.

Na verdade, uma grande lição da TWC é que as empresas precisam planejar de forma realista (antecipar o custo, os recursos e o tempo necessários) para a preparação e a integração dos dados. A experiência mostra que 80% do tempo de uma equipe pode ser gasto na preparação de dados, relegando o trabalho realmente prazeroso de executar modelos, ajustá-los e questionar os resultados para uma fração muito menor de seu tempo. A automação pode ajudar nisso.

Considere apenas uma área de trabalho do TWC: previsão de influências sazonais e impactos de alergias. Os desafios exclusivos de mesclar todos os dados necessários significa construir pipelines de dados customizados e serviços, até mesmo criar um data lake especial para modelagem. Isso requer profundos conhecimentos de dados, assim como ser capaz de responder a uma pergunta aparentemente simples: podemos dividir esses dados da maneira que quisermos ou precisarmos? Felizmente, ao enfrentar isso pela primeira vez, o TWC já tinha engenheiros e cientistas de dados capacitados para a tarefa.

Igualmente importante? A equipe foi apoiada por executivos que entenderam, aceitaram o projeto e apoiaram o investimento de tempo e recursos necessários para cumprir os objetivos e alcançar os resultados desejados.

1. *Integração.* Construir uma capacidade que utiliza os dados para IA de todas as partes de uma organização e além dela.
2. *Governança.* Aplicar as abordagens mais recentes de governança ao gerenciar dados para IA para construir confiança nos insights que surgem.
3. *Ferramentas.* Fornecer as ferramentas necessárias para as equipes que precisam delas.

Com o grau certo de perspicácia empresarial e pragmatismo, as empresas podem progredir na geração contínua da “força” dos dados institucionais para lidar com algumas das características únicas da IA.

O imperativo da integração

A pandemia de COVID-19 demonstrou claramente como a renúncia padrão de que “o desempenho passado não é indicativo de resultados futuros” se aplica aos dados e aos insights extraídos deles. As empresas que confiaram na análise dos dados dos anos anteriores e nos modelos históricos acharam suas previsões inúteis.⁹ Para sobreviver, elas tiveram que começar a processar dados de curto prazo, por exemplo, vendas recentes e dados meteorológicos, até mesmo preços do petróleo, e atualizar os modelos de IA para previsões mais relevantes. Na verdade, a IA rapidamente se tornou ainda mais popular devido à sua capacidade de processar dados voláteis de curto prazo para aumentar os dados existentes, se adaptar rapidamente e produzir previsões relevantes.

Muitos desses dados de curto prazo rapidamente atualizáveis se originam fora da empresa, portanto, uma abordagem robusta para integração de dados públicos e de terceiros é um recurso cada vez mais necessário para a IA.

A capacidade de traçar um caminho a partir de um evento importante em curso até seus efeitos subsequentes, como desemprego ou redução de renda, gera insights para o gerenciamento.

Se as fontes de dados que fornecem os níveis corretos de granularidade forem efetivamente usadas, os insights derivados delas poderão ser altamente localizados e, portanto, verdadeiramente úteis. A capacidade de traçar um caminho a partir de um grande evento em curso, como a pandemia, até seus efeitos posteriores, como o desemprego, a redução de renda e os consequentes efeitos nos padrões de compra, gera insights para o gerenciamento.

Um exemplo de abordagem que oferece suporte a esse tipo de localização é o IBM COVID-19 Lockdown Index. Ele permite um entendimento, atualizado diariamente com base nos dados mais recentes, do estado atual do impacto da COVID-19 na atividade econômica, município por município, nos EUA. As empresas podem usar este índice em harmonia com seus próprios dados para criar previsões úteis e acionáveis e, em seguida, planejar e ajustar adequadamente (consulte “Perspectiva: integrar diferentes fontes de dados para gerar insights locais da COVID-19”).

Confiança no controle e na procedência

O uso de dados externos ou de terceiros ilustra especialmente a necessidade das organizações conhecerem bem seus dados. Quem é o dono dos dados? Quem pode aprovar seu uso e por quanto tempo? A comercialização de ativos de IA com base em dados de terceiros pode levantar questões ainda mais complexas. Se um ativo de IA foi treinado em dados aos quais uma empresa não tem mais acesso ou aprovação para usar, o que fazer? E embora essas questões obviamente se apliquem a dados externos, problemas semelhantes podem ocorrer com dados internos.

Mais da metade das organizações tem dificuldade com a integração de dados em geral, enquanto 1/3 delas não confia em seus recursos para conectar várias fontes de dados.¹¹

Também, há o problema dos dados regulamentados. As fronteiras dos países podem afetar a disponibilidade de dados e como eles podem ser usados. A qualidade dos dados também pode variar de país para país, o que significa que os dados geográficos podem não ser consistentes ou utilizáveis da mesma forma. Por exemplo, em alguns países, os dados de transações de cartão de crédito são detalhadamente capturados devido aos microchips e leitores de cartão avançados; em outros, que não possuem esse hardware, as máquinas de cartão de crédito coletam menos informações sobre as transações.

A abordagem certa à governança de dados pode ajudar com esses problemas, incluindo conhecer e ser capaz de rastrear a proveniência dos dados. Em sua forma mais simples, a governança de dados refere-se às normas, princípios e regras aplicadas ao gerenciamento de diferentes tipos de dados. A governança de dados pode ser aplicada em um nível muito localizado, dentro de uma organização, para lidar com os dados de maneira adequada e manter sua integridade e utilidade em cada estágio de seu ciclo de vida. Ela também pode ser aplicada à cooperação entre organizações, entre ecossistemas e até mesmo entre países: a maneira como os dados serão compartilhados.

Talvez mais importante do que a tecnologia certa sejam as regras e a cultura certas para tornar o acesso aos dados fácil e bem gerenciado. Uma cultura de dados bem familiarizada com os termos e condições padrão e com a maneira como a organização lida com os dados terá muito mais probabilidade de compartilhar dados e fazer bom uso deles.

Perspectiva: integre fontes de dados diferentes para gerar insights locais da COVID-19

Saber que uma grande tempestade está caindo sobre a metade oriental de um país é interessante de um modo geral, mas saber que ela atingirá seu estado ou sua cidade, ou passará por você, é bem melhor. Insights localizados podem ser poderosos.

O IBM COVID-19 Lockdown Index quantifica o grau de ruptura estabelecido atualmente em cada localidade. Ele combina informações longitudinais com atualizações diárias em tempo real sobre a progressão da doença, estatísticas de leitos hospitalares e infecções, restrições da comunidade local e medidas gerais de volatilidade do mercado dos EUA. Em seguida, projeta quando o pico de densidade da doença pode ser alcançado, município por município, e a forma distinta da curva descendente em cada local, atribuindo, então, uma pontuação de risco e identificando o ritmo em que um lockdown local pode aumentar.

Sua interface fácil e simples ilumina visualmente o grau de lockdown em cada município dos EUA e como esse lockdown provavelmente evoluirá nas próximas quatro semanas (consulte a Figura 2).

O índice usa dados disponíveis publicamente, incluindo: medidas epidemiológicas de requisitos de demanda de hospitais estaduais; número de casos COVID-19 município por município; uso de leito hospitalar; uso de leito em unidade de terapia intensiva e ventiladores em uso. O sistema também analisa as notícias locais em busca de atualizações sobre as datas de abertura/fechamento das escolas e outras restrições potenciais da comunidade.

Outra iniciativa semelhante é a The Emergent Alliance, uma colaboração sem fins lucrativos da IBM, Rolls-Royce, Microsoft e dezenas de empresas globais. A aliança está trabalhando para trazer fotos regionais precisas e atualizadas dos casos de Covid-19 para ajudar as autoridades locais a montar uma resposta mais eficaz aos surtos de coronavírus.

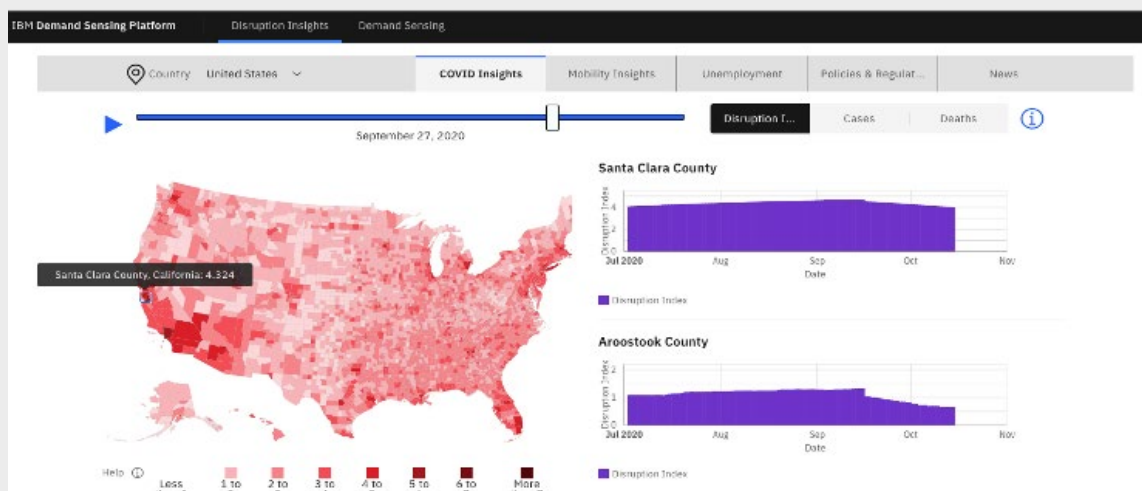
Um índice de risco localizado combina dados de taxas de infecção, mídia social, notícias, dados de Airbnb e muito mais. A análise inclui o impacto da doença na saúde, como os governos estão reagindo, como o comportamento público está mudando e o efeito de tudo isso na economia.

Os benefícios do projeto, porém, vão além de lidar com a pandemia. Dados de impressão sobre tópicos de notícias podem ajudar a antecipar novos padrões de comportamento. Ver um aumento nas notícias sobre esportes ao ar livre, por exemplo, ou sobre reservas de Airbnb nas montanhas, pode levar a uma campanha sobre caminhadas ou outros bens e serviços relacionados.¹⁰

Figura 2

Insight visual

Índice de lockdown da COVID-19



Se os destinatários dos insights não confiarem nos dados nos quais se baseiam, é improvável que os algoritmos de IA mais potentes e sofisticados tenham impacto significativo nos negócios.

Automatizar o máximo possível economiza tempo e recursos, enquanto encoraja a adoção mais ampla de métodos de IA, uma vez que a obtenção de dados não é mais uma tarefa difícil. Isso pode ajudar as equipes a evitarem ficar paralisadas por semanas ou meses em busca de permissão para usar os dados.

Nada na cultura de uma organização importa mais do que a confiança. Embora essencial dentro de uma organização, a confiança é ainda mais importante quando uma organização compartilha ou aceita dados de fora de suas fronteiras (consulte a Figura 3). Não importa o quão boa uma equipe de IA seja na coleta de dados e na construção de soluções que demonstrem valor. Se os destinatários dos insights gerados não confiarem nos dados, de onde eles vêm, como são usados, que suas distorções foram detectadas e sanadas ou que são transparentes, e que estão em conformidade com regulamentos e leis, então provavelmente os resultados estarão aquém do ideal. Na verdade, em tais circunstâncias, é improvável que os algoritmos de IA mais poderosos e sofisticados tenham muito impacto nos negócios.

As lições aprendidas com os recentes avanços conceituais e tecnológicos em blockchain fornecem uma fonte potencial de ajuda para o estabelecimento da confiança. Um dos princípios mais importantes do blockchain é que ele reúne as partes certas desde o início, aquelas que criariam ou quebrariam uma rede e que podem responder a perguntas como: “Quais dados serão compartilhados? Com quem?” Esse ecossistema minimamente viável, que também pode incluir reguladores, decide sobre a estrutura de incentivos, as estruturas de monetização e as regras de governança da rede.

Um conceito importante da governança de dados é preservar o conhecimento e os insights agregados, mas não os dados confidenciais que podem ter implicações regulamentares. Um princípio de blockchain pode ajudar aqui também: uma abordagem baseada em permissão e validação de rede preserva a transparência, a integridade de dados, a linhagem de dados e procedência. Isso pode ajudar a resolver uma das principais preocupações expressas por 66% dos profissionais de TI sobre como mitigar a falta de clareza sobre a procedência dos dados de treinamento de IA.¹²

Claro, a boa governança não termina com o rastreamento dos dados nos quais um modelo de IA é treinado. Ela também inclui o exame das decisões tomadas por um ser humano com base nesses dados, o que é essencial para a explicabilidade, especialmente quando essas decisões podem ser contestadas.

Figura 3

O dilema do compartilhamento

Os dados confiáveis são altamente valorizados, mas muito bem protegidos

Obter dados confiáveis é importante

79%



O cliente exigirá transparência e privacidade na troca de dados no futuro

78%



Disposição para compartilhar dados proprietários com parceiros de negócios em troca de valor

48%



Fonte: Pureswaran, Veena; Parm Sanha; Smitha Soman. “Promovendo o comércio global com blockchain: como liberar valor de mercados confiáveis e interconectados,” P: “Até que ponto você concorda com as seguintes declarações sobre dados confiáveis?” IBM Institute for Business Value. Maio de 2020.

Métodos e ferramentas de dados padrão não podem ser aplicados inalterados à IA.

Use as ferramentas certas

Muitos aplicativos de IA transformam dados brutos em sinais e encontram padrões e insights em séries temporais e outros conjuntos de dados muito grandes. Normalmente, para revelar esses sinais, os conjuntos de dados precisam ser processados centenas de vezes.

Imagine um grande varejista com mais de 600 milhões de SKUs. Então, imagine o que é necessário para processá-los e a capacidade para fazer isso diariamente *centenas de vezes*. Certamente não é um problema trivial, pois esse grau de poder de cálculo seguramente requer habilidades de dados avançadas e também envolve custos financeiros, incluindo os custos ambientais.

E, assim, surge um desafio de metadados que pode ser maior do que o desafio de dados original. Pense em uma única captura de imagem para um veículo inteligente e todos os dados contextuais que ele gera: data, hora, localização, objetos na imagem e sua velocidade em relação uns aos outros, bem como o mundo fora da imagem (contexto ambiental), e rapidamente a lista fica muito longa.

Os métodos e ferramentas de dados padrão não podem ser aplicados inalterados à IA. Uma organização precisa de ferramentas adequadas para preparar, refinar, limpar, combinar e reutilizar dados para IA (consulte “Perspectiva: Chief Data Office da IBM - Ferramentas para automatizar a governança de dados”).

Por exemplo:

- *Pipelines de dados reutilizáveis e tecnologia para suportar a catalogação de dados*, o que pode automatizar alguns aspectos da governança de dados e da validação de linhagem. Isso facilita a reutilização de dados e evita o restabelecimento de todas as aprovações e bases iniciais para obtê-los e usá-los, uma vez que muitas licenças estipulam que parte dos dados pode ser usada e como pode ser usada.
- *Ferramentas de virtualização de dados* que reduzem custos e simplificam alguns aspectos do uso de dados, permitindo que os dados sejam representados e manipulados em um novo ambiente sem que eles sejam, de fato, movidos.
- *Ferramentas de modelagem “sem código”* que permitem a aplicação visual quase arrastar e soltar de modelos de IA para conjuntos de dados para usuários de negócios e outros sem experiência em engenharia de IA.

Estreitamente relacionado a isso está o tipo de ambiente que as empresas devem cultivar. Pessoas em toda a empresa, cientistas de dados, analistas de dados, proprietários de negócios, devem ser capazes de obter facilmente os dados de que precisam. Um armazenamento de recursos (pense nisso como uma interface amigável no topo de um catálogo de dados tradicional), pode facilitar a localização de dados confiáveis e modelos de IA que já foram catalogados e examinados. Ele suporta a reutilização em toda a empresa, não apenas para usuários técnicos, ajudando a empresa a obter o máximo valor de seus dados e investimentos em IA.

Uma vez que o viés de dados pode prejudicar os resultados de IA, e até mesmo destituir grupos chave de partes interessadas, ferramentas que podem detectar e ajudar a mitigar o viés devem fazer parte da plataforma de habilitação da empresa, junto com ferramentas de visualização de dados que podem destacar possíveis problemas com os dados.

Ferramentas bem projetadas também podem gerar confiança. Por exemplo, imagine uma empresa de petróleo e gás com um tesouro de dados geográficos, topográficos e sísmicos. Sua equipe de trabalhadores de campo, incluindo cientistas e engenheiros de várias disciplinas, está pronta para usá-los, mas não possui as habilidades de programação necessárias para trabalhar com modelos de IA que os cientistas de dados construíram. Integrar as ferramentas certas com os modelos e dados pode dar às equipes de campo responsáveis pelas operações diárias a capacidade de se envolverem totalmente com a IA sem que precisem ser codificadores qualificados em ciência de dados.

A escolha das abordagens e ferramentas analíticas corretas pode ajudar a facilitar o foco no valor: por exemplo, vinculando o desempenho algorítmico às métricas de negócios. Em uma ilustração do mundo real, um dono da mercearia europeu estava usando IA para prever onde melhor investir em locais futuros. Ao analisar o contexto atual de onde as lojas existentes estavam, ele precisava ser capaz de distinguir entre lojas pequenas e muito grandes.

Para selecionar as métricas corretas, a equipe precisava combinar um conhecimento profundo da ciência de dados com uma visão ampla dos objetivos de negócios. Para os interessados, o erro de raiz quadrada média logarítmica provou ser uma abordagem muito melhor do que o erro de raiz quadrada média tradicional para capturar as enormes diferenças de escala, um caso claro de compreensão da intenção do negócio e da melhor ferramenta para combinar com ele.¹³

As ferramentas certas são importantes, mas o uso bem-sucedido da IA requer a integração dessas ferramentas como parte de uma plataforma de dados estratégicos mais ampla. Com essa plataforma estabelecida, as equipes também se beneficiam das habilidades multidisciplinares no início de um projeto, incluindo estrategistas, analistas de negócios, especialistas funcionais (SMEs), engenheiros de dados, cientistas de dados, organizadores de dados e gerentes de produto e de projeto.

A implementação de IA bem-sucedida também exige que essas equipes tenham estruturas organizacionais alinhadas com sua abordagem aos dados. A maneira como uma organização lida com os dados tende a imitar a estrutura da organização. Uma empresa firmemente enraizada em feudos e silos terá dificuldades para compartilhar dados. As empresas que criam uma capacidade de dados de IA bem-sucedida acertam em seguir um modelo operacional no qual a experiência e a prática de IA, independentemente de onde estejam, se alinham com muitas das características do negócio.

Em última análise, porém, garantir que as práticas de dados estejam à altura da tarefa de suportar as tecnologias de IA atuais não é suficiente. As abordagens de gerenciamento de dados devem evoluir além do que está no horizonte da IA, como técnicas neuro-simbólicas, e antecipar suas necessidades exclusivas.

Perspectiva: Chief Data Office da IBM — Ferramentas para automatizar a governança de dados

Integrar, catalogar e administrar ativos de dados para uma empresa com mais de 100 anos, com aproximadamente 600 entidades e negócios conhecidos, é um grande desafio. Uma ferramenta importante para ajudar é o glossário da IBM, um conjunto de termos de negócios que descreve os dados que são a base para a governança. Inicialmente, seu uso era quase totalmente manual, exigindo ampla comunicação com as equipes de governança e os proprietários da fonte de dados.

Em longo prazo, essa não seria uma abordagem extensível e exigiria centenas de líderes de governança combinados com um número igual de SMEs para padronizar e integrar os dados. Em uma tentativa de ajudar a automatizar o estágio de curadoria e fornecer metadados para ajudar os especialistas, as equipes recorreram à IA para rotular os dados e automatizar algumas das funções de dados padrão. A iniciativa, chamada Geração automatizada de metadados, ajudou a reduzir o enorme acúmulo de ativos de dados que precisam de integração e curadoria de seis a oito semanas para menos de 10 dias, com reduções significativas na demanda de tempo dos membros da equipe.

A abordagem também melhorou a qualidade dos dados, simplificando ainda mais a governança de dados e os requisitos regulamentares. A expansão do uso da tecnologia acelerou muito a capacidade de toda a empresa de controlar os dados. Seu sucesso levou a uma solicitação para fortalecê-la e transformá-la em uma oferta comercial.

A governança foi uma parte essencial desse fortalecimento. A governança com falha gera problemas sérios, então a equipe permitiu que a ferramenta capturasse os dados imutáveis que poderiam validar a proveniência e a linhagem dos dados para fornecer garantias de que os modelos baseados neles poderiam ser facilmente mantidos de maneira sustentável e proativa.

Plano de ação

Lidando com o dilema dos dados

Embora tirar a IA do laboratório e colocá-la em produção total esteja longe de ser uma tarefa trivial, identificamos as principais ações que as empresas podem realizar para acelerar o caminho para o ajuste de escala da IA.

Práticas líderes para adotantes de IA menos estabelecidos:
empresas nas fases de consideração, avaliação e testes

Valor de negócios

Concentre-se no problema e se a IA pode resolvê-lo. Desenvolva e adapte casos de uso de IA para demonstrar o benefício para outros que podem ser coproprietários ou ter dados úteis, incluindo parceiros de negócios e terceiros. Os workshops de Design Thinking, em particular, podem incluir diversos interessados desde o início. Incluir aqueles que controlam conjuntos de dados importantes pode dar às partes interessadas um senso de voz, propriedade compartilhada e interesse no resultado do projeto de IA, além de ajudar a abrir o acesso a dados valiosos.

Integração

Comece com o que está à mão. Concentre-se nos dados realmente disponíveis. Não se distraia com a manipulação de “dados de brinquedo” ou mesmo dados sintéticos para experimentos. No outro extremo, não fique tentado a integrar dados que você não pode obter facilmente ou provavelmente não tem permissão para usar.

Governança

Reduza criteriosamente. Coloque em modo de curadoria de dados ou, se garantido, feche projetos de IA que não possam mostrar um aumento constante nos dados de treinamento ou adesão real das partes interessadas para compartilhar dados.

Ferramentas

Não esteja sozinho. Nenhuma organização possui todos os dados, ferramentas ou plataformas de que precisa. A complexidade de integrar e administrar dados de fontes distintas com diferentes proprietários e restrições de uso pode ser imensa. Escolha cuidadosamente os parceiros com os quais você pode colaborar e que podem fornecer dados utilizáveis, ferramentas de nível superior e uma plataforma integrada. Dessa forma, você não gasta a maior parte de seus recursos buscando, preparando e gerenciando tudo por conta própria.

Práticas líderes para adotantes de IA mais estabelecidos:
empresas nas fases de implementação, operação e otimização

Valor de negócios

Invista em “jornalismo de dados” para comunicar valor. Construa uma cultura eficaz para dados e IA confiáveis. Faça relatórios regulares sobre os resultados e o potencial do projeto por meio de eventos e jornalismo de dados ativos. Entrevistar partes interessadas para ajudar outros grupos a compreender a sua perspectiva e necessidades. Crie histórias que ilustrem boas práticas de dados e os benefícios dos dados aproveitados para IA.

Dê vida ao elemento humano e aos impactos positivos que os projetos de IA têm nas pessoas. E não se esqueça de imaginar o futuro: o que um grande negócio ou publicação técnica estará dizendo sobre sua empresa (e equipe) em cinco anos?

Integração

Iterar, integrar e automatizar. Aplique métodos ágeis (sprints iterativos, em particular) para identificar desde o início problemas com dados que podem dificultar o sucesso. Identifique outros conjuntos de dados, externos e internos, que podem enriquecer os insights e atingir os objetivos de negócios. Construa mecanismos para reutilizar dados e ferramentas para automatizar e evitar retrabalho.

Governança

Escale práticas de dados para escalar a IA. Durante os testes, encontre respostas para perguntas operacionais. Por exemplo, como você invocará seus modelos, com que frequência os modelos precisarão ser ajustados, a taxa de ingestão de dados (streaming versus um trabalho agendado?) o volume de dados em produção e o tamanho do hardware requerido. Para implementar na produção no mundo real, você precisará considerar outros requisitos de dados relacionados à governança, volume e função de organizadores de dados.

Ferramentas

Capacite “cientistas de dados cidadãos” (com cuidado). Um ambiente de armazenamento de recursos e ferramentas de modelagem “sem código/low-code” permite que usuários de negócios e outros sem experiência em engenharia de IA aproveitem a IA por conta própria. Mas, uma governança robusta é essencial para evitar que cidadãos bem-intencionados fiquem “descontrolados”.

Notas e fontes

- 1 Dado não publicado da pesquisa do IBM Institute for Business Value sobre o Valor da IA (n = 2.765 executivos de negócios de nível C nas principais indústrias, áreas funcionais e geografias).
- 2 “2021 CEO Study: Find your essential.” IBM Institute for Business Value. Fevereiro de 2021. <https://www.ibm.com/thought-leadership/institute-business-value/c-suite-study/ceo>
- 3 “Global AI Adoption Index 2021.” Consulta matinal em nome da IBM. Abril de 2021. <https://newsroom.ibm.com/IBMs-Global-AI-Adoption-Index-2021>
- 4 Linthwaite, Rachel. “Overcome Obstacles To Get To AI At Scale.” Forrester. Janeiro de 2020. <https://www.ibm.com/downloads/cas/VBMPEQLN>; “IDC Survey Finds Artificial Intelligence to be a Priority for Organizations But Few Have Implemented an Enterprise-Wide Strategy.” IDC. 8 de julho de 2019. <https://www.idc.com/getdoc.jsp?containerId=prUS46534820>
- 5 Ibid.
- 6 Ibid.
- 7 “Global AI Adoption Index 2021.” Consulta matinal em nome da IBM. Abril de 2021. <https://newsroom.ibm.com/IBMs-Global-AI-Adoption-Index-2021>
- 8 “AI Global Survey.” IDC. Maio de 2019; Linthwaite, Rachel. “Overcome Obstacles To Get To AI At Scale.” Forrester. Janeiro de 2020. <https://www.ibm.com/downloads/cas/VBMPEQLN>
- 9 Poole, Erin. “The new frontier: Industries navigate supply chain challenges quicker with demand sensing.” IBM Smarter Business Review. 21 de maio de 2020. <https://www.ibm.com/blogs/services/2020/05/21/the-new-frontier-industries-navigate-supply-chain-challenges-quicker-with-demand-sensing/>
- 10 “How AI and data models help governments fight Covid-19.” Financial Times. 2020. <https://www.ft.com/partnercontent/ibm/how-ai-and-data-models-help-governments-fight-covid-19.html> ; <https://emergentalliance.org/about>
- 11 “AI Global Survey.” IDC. Maio de 2019; Linthwaite, Rachel. “Overcome Obstacles To Get To AI At Scale.” Forrester. Janeiro de 2020. <https://www.ibm.com/downloads/cas/VBMPEQLN>
- 12 “Global AI Adoption Index 2021.” Consulta matinal em nome da IBM. Abril de 2021. <https://newsroom.ibm.com/IBMs-Global-AI-Adoption-Index-2021>
- 13 Experiência de especialistas. Para uma discussão dos dois métodos, consulte Saxena, Sharoon. “Qual a diferença entre RMSE e RMSLE?” Analytics Vidhya. 26 de junho de 2019. <https://medium.com/analytics-vidhya/root-mean-square-log-error-rmse-vs-rmlse-935c6cc1802a>.

Sobre o Expert Insights

Expert Insights representam as opiniões de líderes de pensamento sobre assuntos relevantes de negócios e relacionados a tecnologia. Eles são baseados em conversas com os principais especialistas no assunto de todo o mundo. Para obter mais informações, entre em contato com o IBM Institute for Business Value no e-mail iibv@us.ibm.com.

© Copyright IBM Corporation 2021

IBM Brasil Ltda
Rua Tutóia, 1157
CEP 04007-900
São Paulo – SP
Brasil

Produzido nos Estados Unidos da América
Junho de 2021

IBM, o logotipo IBM e ibm.com são marcas comerciais da International Business Machines Corp., registradas em muitos países no mundo todo. Outros nomes de produtos e serviços podem ser marcas comerciais da IBM ou de outras empresas. Uma lista atualizada de marcas comerciais IBM está disponível na Web em “Copyright and trademark information” no endereço ibm.com/legal/copytrade.shtml.

Este documento estava atualizado na data de publicação inicial e pode ser alterado pela IBM a qualquer momento. Nem todas as ofertas estão disponíveis nos países onde a IBM opera.

AS INFORMAÇÕES NESTE DOCUMENTO SÃO OFERECIDAS NO ESTADO EM QUE SE ENCONTRAM (“AS IS”) SEM QUALQUER GARANTIA, EXPLÍCITA OU IMPLÍCITA, INCLUINDO MAS SEM SE LIMITAR A GARANTIAS DE COMERCIALIZABILIDADE, ADEQUAÇÃO A UM PROPÓSITO ESPECIAL E QUALQUER GARANTIA OU CONDIÇÃO DE NÃO VIOLAÇÃO. Os produtos IBM são garantidos de acordo com os termos e condições dos acordos sob os quais são fornecidos.

Esse relatório tem a intenção de oferecer apenas orientação geral. Não se destina a substituir pesquisa detalhada ou o bom-senso profissional. A IBM não será responsável por qualquer perda sofrida por qualquer organização ou pessoa que utilize esta publicação.

Os dados usados neste relatório podem ser derivados de fontes que não a IBM, e a IBM não verifica, valida ou audita tais dados. Os resultados do uso de tais dados são fornecidos “no estado em que se encontram”, e a IBM não faz qualquer garantia, expressa ou implícita.

