

Relatório de Viagem - AI CON USA 2024

Henrique Faulhaber

Em junho de 2024, participei do AI CON USA em Las Vegas, um evento de Inteligência Artificial com palestras, workshops e stands de empresas do setor.

A palestra de abertura foi feita por Donna Sarkar da Microsoft, que destacou a onipresença da inteligência artificial (IA) no ambiente profissional atual, especialmente desde o final de novembro de 2022. Ela explicou o funcionamento da IA generativa baseada nos modelos conhecidos como transformadores generativos pré-treinados (GPT), que criam novos conteúdos com base em grandes conjuntos de dados pré-existentes e passam por um processo de pré-treinamento que exige grande esforço computacional.

Donna também abordou o conceito de pesquisa semântica, que vai além da simples pesquisa por palavras-chave para encontrar dados contextualmente semelhantes. Ela enfatizou a importância de integrar a IA com dados específicos de negócios, garantindo que os resultados sejam mais relevantes e úteis para cada caso.

Sarkar identificou vários papéis profissionais emergentes no ecossistema da IA, como especialistas em aprendizado de máquina, solucionadores de problemas, especialistas em dados, construtores de soluções e validadores de IA. Ela discutiu a importância de cada um desses papéis e a necessidade contínua de aprimoramento profissional.

Ela também abordou os desafios e oportunidades da IA, destacando as dificuldades associadas aos prompts de IA e a necessidade de dados de treinamento de alta qualidade. Além disso, enfatizou a importância das habilidades humanas, que continuam insubstituíveis mesmo em um mundo cada vez mais dominado pela IA.

Por fim, Donna falou sobre tendências futuras na IA. Ela prevê a criação de produtos de IA específicos para indústrias, o desenvolvimento de pequenos modelos de linguagem (SLMs) que podem operar localmente sem a necessidade de conexão com a nuvem, e a crescente importância da gestão de dados pessoais. Também mencionou a necessidade de melhorar a criação de prompts e o desenvolvimento de bibliotecas de prompts, facilitando o uso eficaz da IA.

Em sua palestra Nevra Ledwon da DecisionBrain falou sobre otimização é um ramo crucial da IA. Ela descreveu como a otimização pode ser usada para criar planos e cronogramas eficientes, utilizando regras de negócios e técnicas de machine learning para prever demandas e necessidades.

Ela compartilhou um exemplo prático de uma empresa de manutenção de elevadores que usou otimização matemática para planejar a manutenção dos elevadores. Com a pandemia de COVID-19, o uso de elevadores diminuiu, mas as empresas ainda precisavam cumprir contratos de manutenção. A otimização ajudou a empresa a oferecer uma garantia de tempo de atividade, ajustando os planos de manutenção com base no uso real dos elevadores.

Ledwon destacou os desafios da otimização, especialmente a complexidade matemática envolvida e o grande número de soluções possíveis. Mesmo com os avanços na velocidade de processamento e computação em nuvem, a otimização continua sendo uma tarefa complexa.

Nevra forneceu vários exemplos de aplicações reais de otimização, como precificação de passagens aéreas e quartos de hotel, roteirização de entregas, otimização de portfólios de investimentos e agendamento de tripulações aéreas. Ela explicou como a otimização pode melhorar significativamente a eficiência operacional e ajudar as empresas a alcançar seus objetivos de negócios.

Ela também mencionou que, atualmente, até mesmo empresas de médio porte estão implementando sistemas baseados em otimização personalizada pois avanços em computação em nuvem e técnicas de modelagem tornaram essas soluções mais acessíveis. Ela ressaltou a importância de plataformas de desenvolvimento de aplicativos que permitem transformar modelos de otimização em aplicativos completos baseados na web.

Hien Luu da DoorDash falou sobre o impacto da IA generativa e da adoção dessa tecnologia nas empresas. Ele começou destacando como as imagens geradas por IA se tornaram incrivelmente realistas e impressionantes em termos de detalhes, iluminação e ângulos de luz.

Hien destacou os avanços significativos na tecnologia de IA generativa, incluindo suas capacidades de criar, raciocinar e interagir. Ele explicou que essas tecnologias têm o potencial de transformar várias indústrias, oferecendo uma maneira mais humana de interação e criação de conteúdo.

Ele também abordou a importância da multimodalidade, onde as tecnologias de IA podem entender, interpretar e gerar saídas em diferentes modos, como texto e imagem. Essa capacidade permite uma compreensão mais abrangente do mundo e pode ser aplicada em diversas áreas, desde ferramentas educacionais imersivas até assistentes digitais personalizados.

Luu identificou vários desafios na adoção da IA, como a falta de clareza, resistência organizacional e escassez de talentos. Ele enfatizou a necessidade de lideranças fornecerem clareza e direção na adoção da IA generativa. Ele também discutiu a importância de conectar iniciativas de IA ao valor de negócios, garantindo que as empresas entendam os problemas que estão tentando resolver e o retorno esperado sobre o investimento.

O palestrante sugeriu várias estratégias para a adoção eficaz da IA, incluindo começar com pequenos projetos gerenciáveis, fomentar a colaboração entre equipes e desenvolver continuamente as habilidades e capacidades técnicas necessárias. Ele enfatizou a importância da experimentação e da abordagem gradual, começando com casos de uso de produtividade interna e expandindo conforme a confiança e a compreensão da tecnologia aumentam.

Sudeep Kesh da S&P Global Ratings falou sobre como a inteligência artificial (IA) e outras tecnologias podem mudar fundamentalmente o cenário dos negócios. Ele destacou a importância de integrar essas tecnologias nos negócios para aproveitar novas oportunidades e enfrentar os riscos associados.

Kesh introduziu um framework para entender a IA nos negócios, chamado "cabeça, coração e mãos". Ele explicou que "cabeça" se refere ao pensamento científico e à medição, "coração" aborda a motivação intrínseca e "mãos" se concentra na ação intencional e curiosa. Ele destacou a importância de agir com intenção e ser curioso como uma criança, sem deixar que o conhecimento atual interfira no aprendizado.

Kesh também abordou os desafios na adoção da IA, como a clareza organizacional, resistência à mudança e escassez de talentos. Ele enfatizou a necessidade de governança sólida e de uma abordagem cuidadosa para gerenciar os riscos associados à IA. Ele mencionou que empresas e governos estão colaborando para enfrentar a desinformação gerada pela IA e desenvolver legislação adequada para regulamentar seu uso.

Kesh sugeriu que as empresas tratem seus modelos de negócios como hipóteses e usem a IA para explorar novas capacidades e oportunidades. Ele destacou a importância de focar nos dados como combustível para os algoritmos de IA e de considerar a operacionalização e a prontidão das empresas para adotar essas tecnologias.

Na mesa redonda sobre o Futuro da IA, os participantes discutiram sobre a Inteligência Artificial Geral (AGI) e suas implicações. Os painelistas começaram discutindo o que significa AGI. Jason Arbon destacou que a definição de AGI ainda é vaga e que é vista como uma inteligência não especializada, capaz de realizar tarefas que qualquer humano médio pode fazer, em várias disciplinas. Tariq King acrescentou que a inteligência geral vai além das tarefas específicas, incluindo funções cognitivas e emocionais.

Kevin Piles e Donna Sarkar apontaram que, embora haja muito medo em torno da AGI, atualmente não há evidências de que estamos próximos de alcançá-la. Donna enfatizou que muitas das preocupações com AGI são exageradas, pois a IA atual ainda falha em tarefas simples como agendar reuniões sem erros. Ela destacou a importância de entender as reais capacidades e limitações da IA antes de prever cenários catastróficos.

Os painelistas concordaram que a transição para AGI não será abrupta, mas gradual, comparando com o recente impacto do ChatGPT, que trouxe a IA para o mainstream de forma rápida. Jason Arbon destacou que as duas principais barreiras para a AGI são a falta de memória de longo prazo e a incapacidade de auto-aperfeiçoamento. Ele sugeriu que superar essas barreiras pode não ser tão distante quanto se imagina.

Donna Sarkar e outros painelistas enfatizaram a necessidade de colocar humanos no circuito para garantir que a IA seja usada de forma segura e ética. Eles discutiram a importância de as empresas implementarem "guard rails" e governança robusta para evitar o uso indevido da tecnologia. A mesa redonda também mencionou que o governo e as organizações precisam colaborar para estabelecer regulamentações adequadas.

Os painelistas discutiram como a IA está sendo integrada em várias indústrias e a necessidade de especialistas dentro dessas indústrias para entender e aplicar a tecnologia

de forma adequada. Donna enfatizou que é crucial que os profissionais em todas as indústrias compreendam o que a IA pode e não pode fazer, e trabalhem para identificar tanto os cenários positivos quanto os negativos.

A mesa redonda concluiu que, embora a AGI seja um tópico de intenso debate e especulação, é crucial focar nas capacidades atuais da IA e nas maneiras de integrá-la de forma segura e ética nos negócios e na sociedade. Os painelistas enfatizaram a importância da colaboração contínua entre empresas, governos e especialistas para navegar os desafios e oportunidades apresentados pela IA.

Matthew Gunter da GitHub falou sobre a utilização do Copilot, ferramenta de IA generativa da Microsoft no processo de desenvolvimento de software. Ele citou um estudo que conduziu envolvendo cerca de 100 desenvolvedores, em que deram o Copilot para aproximadamente metade deles, com um vídeo de um minuto sobre como usá-lo. Todos os participantes tinham que criar um servidor HTTP em JavaScript. Aqueles que usaram o Copilot concluíram a tarefa em 55% menos tempo. Embora os resultados tenham sido surpreendentes, perceberam que seria mais difícil para as empresas verem todo esse impacto em uma escala organizacional.

Matthew enfatizou que há vários desafios na adoção do Copilot, pois a adoção inicial era rápida, mas depois se estabiliza. Havia resistência ou apatia que impedia todos os desenvolvedores de adotarem o Copilot automaticamente. Organizações focadas em métricas tiveram muitos desafios para ver qualquer impacto significativo em suas tendências de telemetria. Quando o Copilot causava mudanças positivas na telemetria, havia ceticismo sobre se realmente era devido ao Copilot ou apenas uma flutuação aleatória.

Matthew disse que o Copilot deve ser entendido como um copiloto, alguém que se senta no cockpit com o piloto e ajuda a realizar diferentes tarefas necessárias para completar a missão de voo. Mas também deve ser visto como um sistema especialista, fornecendo informações e insights em tempo real. Como um tradutor, ajudando com novas linguagens e frameworks. Como uma rede de comunicação, transportando contexto de uma sessão para outra. Como um modelo preditivo, ajudando a prever o impacto de mudanças no código. E finalmente, como um nível de automação, similar ao que vemos na indústria automobilística.

Na palestra foi citado que um estudo da McKinsey de 2023 analisou outras ferramentas e encontrou benefícios semelhantes, mas também uma variação no nível de benefício. Mesmo para tarefas complexas de codificação, o Copilot conseguiu economizar 10% do tempo do desenvolvedor. Outro estudo do MIT, não sobre o Copilot, mas sobre o ChatGPT, mostrou como a estrutura da tarefa muda quando os desenvolvedores usam essas ferramentas. O brainstorming manteve o mesmo tempo, mas a redação das primeiras versões foi reduzida de 50% para 25% do tempo, enquanto a edição aumentou. Desenvolvedores juniores se beneficiaram completando tarefas com maior sofisticação e economizando tempo, enquanto desenvolvedores mais experientes economizaram apenas tempo.