



Inteligência Artificial Responsável na Educação

Desafios Éticos e Oportunidades



Ciências da Computação
Região Autónoma da Madeira
O habitat natural dos criadores

Título:

Inteligência Artificial Responsável na Educação | Desafios Éticos e Oportunidades
Ciências da Computação - O habitat natural dos criadores

Autor:

Rodolfo Duarte Pinto - Ciências da Computação | CSF Code.org Facilitador
Gabinete de Modernização das Tecnologias Educativas
Divisão de Tecnologias, Segurança e Infraestruturas
Direção de Serviços de Tecnologias e Ambientes Inovadores de Aprendizagem
Direção Regional de Educação

Contactos:

Rua D. João n.º 57, Quinta Olinda 9054 - 510 Funchal
Madeira - Portugal
Telefone: +351 291 705 860
Email: rodolfodu7@edu.madeira.gov.pt

Funchal, fevereiro de 2024

Índice

Introdução	4
1 - Reconhecimento Facial	5
2 - Sistemas de Recomendação em Redes Sociais	6
3 - Sistema de rastreamento de candidatos.....	7
4 - Algoritmos de aprovação de seguros, empréstimos e ajuda financeira.....	8
5 - Software Médico	9
6 - Reconhecimento de voz e assistentes de IA	10
7 - IA Generativa.....	11
8 - Pontuações de risco para populações vulneráveis	12
9 - Meio ambiente	13
Notas finais	14
Referências bibliográficas.....	15

Introdução

Numa realidade social e educativa em constante evolução, a presença e a influência dos sistemas de Inteligência Artificial (IA) tornam-se cada vez mais proeminentes. Este documento, "Inteligência Artificial Responsável na Educação - Desafios Éticos e Oportunidades", e que surge no seguimento do documento "Inteligência Artificial Responsável na Educação - Um Compromisso de Excelência" explora as considerações éticas sobre a integração da IA neste campo crucial. Desde o reconhecimento facial até aos algoritmos de aprovação, refletiremos sobre as diferentes facetas da IA que poderão moldar a nossa realidade.

Exploraremos diversas áreas de pesquisa em ética na IA, oferecendo uma visão abrangente e prática de como a tecnologia está a ser utilizada na Educação. Para cada área, destacaremos uma tecnologia ou setor específico que incorpora a IA com a partilha de exemplos concretos de casos de utilização.

Este documento, **elaborado com o suporte do curso code.org "AI Research Areas"**, servirá como um guia essencial não apenas para compreender o papel crucial da IA na Educação, mas também para ajudar os professores a perceberem a realidade da IA para além do contexto educacional. Assim é nosso compromisso aprofundar o conhecimento e fornecer um ponto de partida robusto para explorar a realidade da IA.

1 - Reconhecimento Facial

Como é que a tecnologia de reconhecimento facial, ao analisar rostos, pode perpetuar preconceitos e desafios éticos, e de que forma podemos lidar com estas questões na sua aplicação prática?

A tecnologia de reconhecimento facial, ao analisar rostos, pode inadvertidamente desenvolver preconceitos e desafios éticos, levantando questões fundamentais sobre a equidade e a justiça na sua aplicação prática. Esta tecnologia, que analisa características faciais para identificar ou verificar indivíduos, opera através da identificação de padrões a partir de um vasto conjunto de imagens de treino, considerando elementos como a distância entre os olhos e a espessura dos lábios.

Apesar de ser amplamente utilizada, desde o desbloqueio de telemóveis até à verificação de identidade, é crucial realçar que o reconhecimento facial tem revelado viés em relação a determinadas raças e géneros, levantando sérias preocupações éticas. Este viés pode resultar em resultados menos precisos e injustos para certos grupos demográficos.

Exemplos práticos de utilização:

Desbloqueio de telemóvel

Segurança aeroportuária

Aplicações de filtros em redes sociais

Deteção de crimes (ladrões)

Controlo de imigração

Lidar com estas questões implica uma abordagem ética no desenvolvimento e implementação deste tipo de tecnologia, incluindo a garantia de um conjunto de dados diversificados e representativos, bem como a implementação de algoritmos que minimizem viés e que promovam a equidade. É imperativo que a utilização do reconhecimento facial seja realizada de forma responsável, respeitando os direitos individuais e evitando discriminações injustas.

2 - Sistemas de Recomendação em Redes Sociais

Como é que os sistemas de recomendação nas redes sociais moldam a nossa percepção do mundo digital?

Os sistemas de recomendação nas redes sociais desempenham um papel decisivo na moldagem da nossa percepção do mundo digital. Estes algoritmos influenciam o que visualizamos, determinando a ordem dos conteúdos, através da seleção de anúncios, recomendação de conteúdos e até sugestões de amizades. Geralmente, utilizam métodos que consideram o histórico do comportamento do utilizador, os dados demográficos e as preferências de utilizadores semelhantes.

Desde as notícias que visualizamos até aos memes que partilhamos, os sistemas de recomendação têm um impacto significativo no nosso percurso online. No entanto, surgem preocupações relacionadas com a propagação da desinformação e o potencial de as redes sociais nos limitarem a bolhas informativas, reduzindo desta forma a exposição a diversas perspetivas.

Exemplos práticos de utilização:

Organização do *feed* de notícias

Partilha de anúncios

Recomendação de conteúdos

Sugestão de amigos

Apresentação de notícias

É fundamental promover a transparência e garantir a diversidade nas fontes de informação. Ao fazermos isso, podemos contribuir para uma experiência online mais equilibrada e enriquecedora.

3 - Sistema de rastreamento de candidatos

Como podemos garantir que os Sistemas de Rastreamento de Candidatos (ATS¹) automatizem eficientemente o recrutamento sem introduzirem discriminação, assegurando uma seleção imparcial de candidatos?

Garantir que os Sistemas de Rastreamento de Candidatos (ATS) automatizem eficientemente o recrutamento sem introduzirem a discriminação, assegurando uma seleção imparcial dos candidatos, é um desafio crucial neste ambiente dinâmico de recrutamento. Os ATS, ferramentas automatizadas que facilitam a gestão de volumes significativos de candidaturas, são essenciais para otimizar processos de recrutamento e contratação, incluindo a publicação de vagas e o tratamento de candidaturas.

Estes sistemas operam com o conhecimento de candidaturas anteriores bem-sucedidas, utilizando critérios específicos para identificarem atributos desejáveis para os futuros candidatos. No entanto, crescem preocupações significativas em relação à possível introdução de discriminação no processo de seleção por parte destas ferramentas.

Exemplos práticos de utilização:

Análise e verificação de antecedentes

Admissão na universidade (ou outros casos)

Seleção de currículos

Abordar estas preocupações implica a implementação de práticas éticas e a consideração de medidas que possam garantir a imparcialidade no processo de seleção. É fundamental que os ATS sejam programados para evitar vieses, com a adoção de critérios de avaliação justos e transparentes. Além disso, a monitorização contínua e a avaliação regular dos algoritmos são essenciais para assegurarem que estes sistemas promovem a diversidade e a igualdade de oportunidades.

¹ https://en.wikipedia.org/wiki/Applicant_tracking_system

4 - Algoritmos de aprovação de seguros, empréstimos e ajuda financeira

Como é que a IA aplicada em algoritmos de aprovação de seguros, empréstimos e auxílios financeiros impacta a equidade no acesso a cuidados, à educação, a oportunidades e até à independência financeira?

A IA pode auxiliar na tomada de decisões financeiras apropriadas, especialmente quando a carga de trabalho e as necessidades são elevadas. Estes sistemas funcionam de maneira semelhante aos Sistemas de Rastreamento de Candidatos, ou seja, dependem de diversos algoritmos para aprenderem com dados passados. Por exemplo, um sistema de IA pode analisar o histórico de quem foi aprovado para empréstimos e quem os pagou pontualmente. Pode utilizar estes dados para prever quem tem maior probabilidade de reembolsar o empréstimo a tempo e, assim, decidir a quem será concedido um empréstimo.

No entanto, tal como nos ATS, estes sistemas podem aprender estereótipos ou reforçar preconceitos sistémicos ao reproduzirem normas culturais. Isso pode resultar na exclusão contínua de certas populações no acesso a cuidados, à educação, a oportunidades e à independência financeira.

Exemplos práticos de utilização:

Aprovação de empréstimo

Atribuição de ajuda financeira

Cálculo do seguro

Torna-se fundamental refletir e avaliar sobre como a aplicação destes algoritmos podem inadvertidamente perpetuar desigualdades, marginalizando grupos específicos. Assegurar a equidade e a justiça na utilização destes algoritmos, de modo a evitar discriminação injusta e garantir o acesso igualitário a oportunidades financeiras.

5 - Software Médico

Num cenário onde a IA desempenha um papel vital na medicina, desde a detecção de doenças até à contribuição no desenvolvimento de novos fármacos, surge uma interrogação crucial: até que ponto podemos confiar na IA para moldar o futuro da saúde?

A IA tem inúmeras aplicações na medicina, desde a detecção de doenças até ao auxílio no desenvolvimento de novos medicamentos. Os hospitais estão a incorporar a IA de várias formas, como a visão computacional que analisa exames radiológicos ou algoritmos treinados com dados de pacientes para auxiliarem na determinação da cobertura de seguros de saúde.

A IBM ²até lançou um projeto para utilizar a IA no diagnóstico do cancro. No entanto, têm ocorrido falhas em projetos significativos da IA médica, como esta ferramenta de diagnóstico de cancro da IBM que recomendou combinações fatais de medicamentos e algoritmos de visão computacional que falharam em detetarem cancros de pele em tons de pele mais escuros.

Exemplos práticos de utilização:

Radiologia

Cobertura de seguro

Investigação sobre medicamentos

Triagem Manchester

É imperativo assegurar a precisão e equidade dos algoritmos, garantindo que são treinados com conjuntos de dados representativos de diversas populações. A transparência e a investigação ética são essenciais para construir a confiança na utilização da IA na moldagem do futuro da saúde.

² <https://teducativas.madeira.gov.pt/s/TpV6CwB>

6 - Reconhecimento de voz e assistentes de IA

Como é que a tecnologia de Reconhecimento de Voz e Assistentes de IA estão a moldar as nossas interações diárias?

Ao longo da última década, os assistentes de voz baseados em IA tornaram-se cada vez mais presentes no nosso quotidiano. Desde o "Hey Siri?", ao "Ok Google" até ao "Alexa!", a tecnologia de reconhecimento de voz avançou para a capacidade de processar a fala em tempo real. Estas tecnologias não só proporcionam vantagens, mas também melhoram a acessibilidade para aqueles que enfrentam barreiras ao acederem a ecrãs, inserirem texto ou realizarem gestos no ecrã.

No entanto, é importante destacar algumas considerações éticas relacionadas ao reconhecimento de voz e aos assistentes de IA. A comercialização predominante destes assistentes como femininos tem sido alvo de críticas, levantando preocupações sobre as representações de género. Além disso, a constante escuta destes dispositivos suscita questões significativas de privacidade. Outro ponto a salientar é a tendência destes assistentes de voz para cometerem erros, ou mal-entendidos, especialmente quando confrontados com sotaques regionais ou expressões locais.

Exemplos práticos de utilização:

Acessibilidade

Casa inteligente

Tradução

Jogos

Embora as vantagens e a acessibilidade sejam evidentes, é fundamental abordar as preocupações éticas, garantindo a transparência na recolha e utilização de dados, bem como promovendo o desenvolvimento de assistentes de voz que reconheçam e compreendam a diversidade linguística e cultural. Este equilíbrio é essencial para garantir uma adoção responsável e ética desta tecnologia inovadora.

7 - IA Generativa

Como é que podemos equilibrar a promessa fascinante da Inteligência Artificial Generativa, capaz de criar diversos tipos de conteúdo, com a necessidade crítica de garantir a credibilidade e evitar possíveis distorções na informação que consumimos?

Muito do que denominamos como IA refere-se especificamente à Inteligência Artificial Generativa: uma tecnologia que pode produzir diversos tipos de conteúdo, incluindo textos, imagens, áudios e dados sintéticos. Exemplos incluem o ChatGPT para linguagem e software de geração de imagens como o Midjourney ou Dall-E. Estas tecnologias são treinadas com enormes conjuntos de dados, seja de texto ou de imagens, utilizando padrões nesses dados para gerarem novo conteúdo.

Contudo, esta capacidade traz consigo implicações significativas para a credibilidade do conteúdo que consumimos. A IA Generativa tem o poder de criar histórias noticiosas que nunca aconteceram, até mesmo gerar vídeos de políticos a dizerem coisas que nunca disseram. Além disso, é propensa a viés, podendo reforçar estereótipos relacionados com a raça, o género, a incapacidade entre outros.

Exemplos práticos de utilização:

Chatbots

Arte, Música, Vídeo

Notícias

Educação

Conteúdo personalizado

Jogos

O desenvolvimento ético e a implementação responsável de medidas de controlo são determinantes para assegurar que a IA Generativa contribui para o avanço positivo e confiável da sociedade, evitando potenciais impactos negativos na qualidade da informação e na perceção pública.

8 - Pontuações de risco para populações vulneráveis

Como podemos equilibrar a utilização da IA na avaliação de riscos para populações vulneráveis, considerando a delicada linha entre benefícios e os potenciais impactos negativos?

A IA proporciona a capacidade de analisar dados, como diagnósticos anteriores, idade e outras informações, para prever a probabilidade de alguém necessitar de tratamento hospitalar no próximo ano. Esta avaliação de risco estende-se a diversas áreas, desde atividades criminosas, negligência infantil, ameaças à segurança e até à deteção de possíveis dificuldades de aprendizagem dos alunos.

Contudo, a complexidade reside na implementação destes algoritmos. Embora prometam *insights* valiosos, alguns sistemas revelam-se invasivos e propensos a incorreções, desencadeando acusações injustas e prejudicando oportunidades e acessos.

Exemplos práticos de utilização:

Bem-estar das crianças

Justiça

Avaliação segurança

Deteção de dificuldades de aprendizagem

Pontuação social

Pontuação de risco para a saúde

Será necessário desenvolver e implementar estratégias que assegurem a equidade, a transparência e a responsabilidade na utilização da IA para a avaliação de riscos, de modo a garantir que os benefícios sejam maximizados sem comprometer a justiça e os direitos individuais.

9 - Meio ambiente

Como é que podemos equilibrar os benefícios da IA na preservação ambiental com a consciência do impacto ambiental intrínseco à criação de modelos de IA?

A aplicação da IA na preservação ambiental tem revelado benefícios significativos, desde a previsão de catástrofes naturais até à proteção dos oceanos e biodiversidade. No entanto, é essencial abordar a consciência do impacto ambiental intrínseco à criação de modelos de IA, reconhecendo que todas as atividades relacionadas com a IA dependem de computadores físicos que consomem energia para trabalhar.

Embora a IA possa ser uma aliada nos desafios ambientais e de sustentabilidade, é importante destacar que a criação e o treino de grandes algoritmos de IA têm uma pegada de carbono considerável. Este processo pode resultar no aumento das emissões de gases de efeito estufa, contribuindo para o aquecimento global.

Exemplos práticos de utilização:

Previsão de catástrofes naturais

Monitorização do declínio da biodiversidade

Previsão sobre as alterações climáticas

Ajuda às pessoas a manterem-se seguras em caso de fenómenos meteorológicos

Deteção de poluição

Apoio à vida selvagem

Para mitigar o impacto ambiental da IA, será necessário promover práticas de desenvolvimento sustentável, investir em fontes de energia renovável para alimentar servidores de IA e incentivar a eficiência energética na computação. Ao fazê-lo, podemos garantir que a IA contribua efetivamente para a preservação ambiental, sem comprometer a saúde do planeta.

Notas finais

Num mundo em constante transformação e impulsionado pela IA, é imperativo reconhecermos não apenas os seus avanços inovadores, mas também as considerações éticas que acompanham a sua integração em várias esferas da sociedade. Este documento explora diversas áreas de pesquisa em ética na IA, destacando tecnologias específicas e disponibilizando exemplos práticos que ilustram a aplicação transversal da IA.

À medida que abraçamos a IA como uma ferramenta poderosa, será preciso adotarmos uma abordagem ética no seu desenvolvimento e implementação. A consciência dos vieses, a transparência na utilização de dados, a promoção da equidade e a responsabilidade na criação de algoritmos tornam-se fundamentais para garantir que a IA contribua positivamente para a sociedade.

Enquanto nos maravilhamos com as conquistas da IA na medicina, na educação, no ambiente e em outras áreas, devemos também abordar os desafios éticos que surgem. Ao promovermos uma discussão contínua sobre a ética na IA e incentivarmos a investigação e o desenvolvimento responsável, podemos moldar um futuro onde a IA é uma aliada confiável para o progresso humano.

Este documento serve como um reforço à exploração contínua destes temas éticos, para que seja possível garantir que a IA é utilizada de maneira responsável, benéfica e equitativa, respeitando os valores fundamentais da nossa sociedade em constante evolução.

Uma palavra de agradecimento sincero ao projeto code.org pela generosa partilha de recursos educativos que serviram como base fundamental para a elaboração deste documento. A disponibilidade destes materiais não apenas enriquece o portefólio educativo, como possibilita a respetiva contextualização destas questões éticas da IA na nossa realidade educativa.

Referências bibliográficas

Code.org. (2024). AI Research Areas. Disponível em: <https://code.org/> (Acedido a 6 de fevereiro de 2024).

Code.org, CoSN, Digital Promise, European EdTech Alliance, Larimore, J., and PACE (2023). AI Guidance for Schools Toolkit. Disponível em: <https://teachai.org> (Acedido a 6 de fevereiro de 2024).

Buolamwini, J. (2023). Unmasking AI. Random House.

Giannini, S. (2023). Generative AI and the future of education. UNESCO. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000385877/PDF/385877eng.pdf.multi> (Acedido a 6 de fevereiro de 2024).

Kaufman, D. (2022). Desmistificando a inteligência artificial. Autêntica.

OpenAI. (2021). ChatGPT (Versão 3.5). Disponível em: <https://openai.com/chatgpt> (Acedido a 6 de fevereiro de 2024).

UNESCO. (2023). Artificial Intelligence. Disponível em: <https://www.unesco.org/en/artificial-intelligence> (Acedido a 6 de fevereiro de 2024).



Este documento é de utilização gratuita ao abrigo de uma licença Internacional
Atribuição – Não Comercial – Compartilha Igual 4.0
(CC BY-NC-SA 4.0)