



Fostering Artificial Intelligence at Schools

Guia inclusivo para educadores (não- formais) no
contexto da educação em Inteligência Artificial

Índice

Introdução	4
Sobre o guia	5
Entre os objetivos do Guia FAlaS incluem-se	5
1. Desenvolver programas inclusivos de educação não-formal em Inteligência Artificial	6
1.1. As vantagens de aprender sobre IA num contexto não-formal	7
1.2. Promover reflexão sobre género, diversidade e inclusão no ambiente de aprendizagem	8
1.3. Práticas de Ensino inclusivas na educação não-formal	10
2. Diversidade e inclusão na inteligência artificial	14
2.1. Inteligência artificial – o que é?	15
2.2. Viés em Inteligência Artificial	17
2.3. Porque é que género, diversidade e inclusão são importantes para a IA?	23
2.4. Mulheres que estão a mudar o campo da inteligência artificial	24
3. Guia para atividades de promoção de uma abordagem inclusiva da inteligência artificial	28
3.1 Atividade 1: O jogo das <i>tags</i>	30
3.2 Atividade 2: Desenha a descrição	33
3.3 Activiteit 3: Raad de persoon	37
3.4 Atividade 4: Viés de género usando LearningML com dataset 1	41
3.5 Atividade 5: Viés de género usando o LearningML com dataset 2	47
3.6. Atividade 6: Identificar vieses na IA	52
Conclusões	57
Definições e glossário	59
Referências	60
Créditos das Imagens	62

Impressão

2023: CollectiveUP info@collectiveup.be

Editores: CollectiveUP (BE), Universidade do Rei Juan Carlos (ES), Teatro Circo Braga (PT), Vrije Universiteit Brussel (BE).

Autores: Liliana Carrillo, Chrysanthi Katrini, Paul Fenton, Loredana Bucseanu (CollectiveUP), Marjon Blondeel (Vrije Universiteit Brussel), Meritxell Diaz, Antonio Romero, Gregorio Robles (King Juan Carlos University), Sara Borges, Joana Miranda (Teatro Circo de Braga).

Layout por: Jurate Laugalyte (CollectiveUP), Carlos Nivia Otero (CollectiveUP).

Traduzido por: Frederick Ducatelle (Dutch), Ana Oliveira Monteiro (Portuguese), Gregorio Gobles and Meritxell Díaz Coque (Espanhol).

Sugerimos que este relatório seja citado como: Carrillo, L., Katrini, C., Fenton, P., ... (2023). *Inclusion guide for (non-formal) educators in the context of Artificial Intelligence education*. https://www.agileforcollaboration.eu/sdc_download/720/?key=shx1aegquebhqkkehrudj7xnOyc8yo

Esta publicação foi escrita no âmbito do projeto Erasmus+ FAIaS – Fostering Artificial Intelligence at Schools, que pretende apoiar os educadores na integração e uso dos conceitos da Inteligência Artificial nas suas sessões de formação, criando e partilhando recursos sobre a IA que podem ser usadas nesse contexto.

CollectiveUP agradece ao Digital Belgium Skills Fund (DBSF) por cofinanciar o projeto AI4InclusiveEducation (www.ai4inclusiveeducation.be) com um número de bolsa DBSF2023-14. O DBSF cofinanciou e apoiou parcialmente a realização do projeto Erasmus+ FAIaS, criando um impacto na Bélgica.

Esta publicação (em inglês, espanhol, holandês e português) e outras publicações e ferramentas produzidas no projeto podem ser descarregadas gratuitamente em: www.fosteringai.net

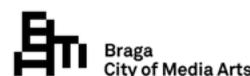
Este trabalho está licenciado sob a Creative Commons Attribution-ShareAlike 4.0 International License. Para ver uma cópia desta licença, visite <http://creativecommons.org/licenses/by-sa/4.0/>



Aviso legal

O FAIaS – Fostering Artificial Intelligence at Schools é um projeto de parceria, cofinanciado pela Comissão Europeia. Esta publicação reflete apenas as opiniões dos autores. A Comissão Europeia não pode ser responsabilizada por qualquer uso que possa ser dado à informação contida nesta publicação.

FAIaS – Fostering Artificial Intelligence at Schools – Projeto Erasmus+, Tipo de Ação “Cooperation partnerships in Scholl Education”, Número da bolsa: 2020-1-ES01-KA201-083047.



Introdução

Numa era em que a inteligência artificial (IA) está a tornar-se cada vez mais presente no nosso quotidiano, é de extrema importância o desenvolvimento ético e responsável desta tecnologia transformadora. *Fostering Artificial Intelligence at Schools (FAIaS)* é um projeto inovador e transformador, dedicado a promover uma inteligência artificial ética e responsável. O seu objetivo principal é promover o conhecimento sobre IA nas crianças e jovens. Ao fazê-lo, o FAIaS aspira a contribuir para a formação de uma geração que esteja não só informada sobre IA, mas também envolvida de forma ativa no pensamento crítico e na tomada de decisões responsáveis. Com a prevalência crescente da IA na nossa sociedade, este projeto reconhece a necessidade urgente de abordar os complexos desafios éticos e as implicações sociais e, em particular, a necessidade de os jovens estarem mais informados e capacitados em relação à utilização da IA.

Através de artigos, estudos de caso e orientações práticas, o projeto explora várias dimensões da ética e do impacto social da IA, para serem utilizados na educação não formal. Ao examinar exemplos do mundo real e ao basear-se em experiências de especialistas, o FAIaS fornece perspectivas e orientações valiosas para nortear o desenvolvimento e implementação de sistemas de IA responsáveis.

Este guia serve de orientação, fornecendo um modelo para guiar nos complexos dilemas éticos que surgem à medida que a IA se integra cada vez mais nos vários aspetos da nossa vida, incluindo nos cuidados de saúde, nas finanças, na educação e outros. Este “Guia Inclusivo para educadores (não-formais)” aborda tópicos como viés algorítmico, questões sobre privacidade, responsabilidade, transparência e o impacto da IA no mercado de trabalho. O livro defende uma aproximação holística ao desenvolvimento da IA – uma abordagem que dê prioridade ao bem-estar e à capacitação dos indivíduos e das comunidades, assegurando simultaneamente que os sistemas IA sigam os princípios de equidade, igualdade e justiça.

Sobre o guia

Este guia é um recurso útil para educadores não formais que desejem incorporar conceitos de inteligência artificial (IA) nos seus conteúdos e formação, com um foco nas práticas de diversidade e inclusão. Dessa forma, este é um complemento do documento “Fomentar a Inteligência Artificial nas Escolas – Orientações para docentes em educação formal” desenvolvido no âmbito do projeto FAlaS.

Capítulo 1 Contextualiza as oportunidades criadas pela educação não formal para aprender sobre IA; introduz uma série de conceitos relacionados com a consciencialização para as questões de género, diversidade e inclusão no ambiente de aprendizagem; e fornece exemplos de práticas de ensino inclusivas na educação não formal.

Capítulo 2 Mergulha nos desafios de desenvolver sistemas de IA inclusivos, introduzindo o conceito do enviesamento na IA, as suas causas, o impacto e soluções possíveis. Também discute por que razão é que o género, a diversidade e a inclusão são importantes na criação de tecnologias IA. A isto acrescenta casos de algumas mulheres que estão a mudar a área da inteligência artificial, servindo de exemplo e inspiração.

Capítulo 3 Consiste em guias práticos para atividades que promovem uma abordagem inclusiva da inteligência artificial, incluindo uma descrição e instruções passo-a-passo para cada atividade.

Finalmente, este guia apresenta um glossário de termos e outros recursos para ser usados na prática pedagógica.

Entre os objetivos do Guia FAlaS incluem-se

- Compreender os conceitos básicos da IA e o seu impacto na sociedade.
- Aprender sobre práticas que promovam a diversidade e a inclusão no contexto de educação não formal.
- Compreender o papel dos educadores não-formais na promoção de um futuro mais inclusivo e equitativo para a IA.
- Reconhecer a importância da inclusão na IA e nas suas aplicações.
- Desenvolver competências para conceber e disponibilizar sessões de formação de IA inclusivas.
- Adquirir a competência de facilitar discussões e debates sobre IA e as suas implicações.
- Motivar estudantes para um pensamento crítico e reflexivo na área da IA.



1

Desenvolver programas
inclusivos de educação
não-formal em
Inteligência Artificial

1.1. As vantagens de aprender sobre inteligência Artificial num contexto não-formal

A Inteligência Artificial (IA) tem o potencial de revolucionar a forma como vivemos, trabalhamos e aprendemos. À medida que a IA continua a crescer em popularidade e importância, torna-se cada vez mais relevante que as pessoas compreendam e sejam capazes de trabalhar com a IA. Programas de educação não-formal podem desempenhar um papel importante no ensino da IA a uma ampla gama de pessoas, desde estudantes e profissionais até aquelas que estão apenas a começar a explorar o mundo da tecnologia.

Uma das principais vantagens da educação não-formal é a sua flexibilidade e adaptabilidade. Os programas de educação não-formal podem ser personalizados para atender às necessidades e interesses específicos de todas as pessoas, tornando-os uma plataforma ideal para o ensino de IA. Por exemplo, cursos e oficinas online podem ser pensados no sentido de proporcionar experiências práticas com IA, enquanto programas comunitários podem ajudar as pessoas a desenvolver as suas capacidades e conhecimentos em IA num ambiente de apoio e colaboração.

Outra vantagem da educação não-formal é a sua capacidade de alcançar comunidades sub-representadas. A IA é um campo em rápido crescimento, e é importante garantir que pessoas de todas as origens e com todo o tipo de experiências tenham acesso à educação e formação em IA. Os programas de educação não-formal podem ajudar a preencher a lacuna entre a educação formal e informal, oferecendo oportunidades de aprendizagem sobre IA a pessoas que não tenham acesso a programas formais de educação ou formação.

Para ensinar IA de forma eficaz em programas de educação não-formal, é importante focar tanto nos aspectos técnicos como nos aspectos éticos da IA. As competências técnicas, como programação e código, são essenciais para quem deseja trabalhar com IA. No entanto, é igualmente importante ensinar sobre as implicações éticas e sociais da IA, incluindo questões relacionadas com a privacidade, proteção de dados e viés nos sistemas de IA. Ao incluir tanto os aspectos técnicos como os aspectos éticos da IA nos programas de educação não-formal, as pessoas podem desenvolver uma compreensão abrangente da IA e estar mais bem preparadas para tomar decisões informadas sobre a sua utilização.

Portanto, os programas de educação não-formal têm o potencial de desempenhar um papel importante no ensino de IA a um vasto conjunto de pessoas, complementando os programas existentes na educação formal. Ao aproveitar as vantagens da educação não-formal, como flexibilidade e adaptabilidade, e ao focar nos aspectos técnicos e éticos da IA, as pessoas podem desenvolver as competências e conhecimentos necessários para serem bem-sucedidas num cenário tecnológico em constante mudança. Ao fornecer acesso à educação e formação em IA, os programas de educação não-formal podem

ajudar pessoas de todas as origens e experiências a alcançar todo o seu potencial e contribuir para o desenvolvimento de um mundo mais inclusivo e equitativo (3).

Entre algumas das ideias-chave em que a educação não-formal pode apoiar a aprendizagem, sobre Tecnologia em IA para os jovens incluem-se:

- **Fornecer experiências práticas:** programas não-formais podem fornecer experiências práticas com tecnologia, como workshops sobre código, programas de robótica, e oficinas criativas.
- **Promover a criatividade e a inovação:** ao fornecer aos/às estudantes uma variedade de instrumentos e recursos para experienciar a tecnologia, a educação não-formal pode promover as suas competências criativas e inovadoras.
- **Apoiar as comunidades sub-representadas:** a educação não-formal pode ajudar a combater a clivagem digital, fornecendo acesso à tecnologia e à educação tecnológica às comunidades que são tipicamente sub-representadas nestas áreas.
- **Encorajar a aprendizagem ao longo da vida:** a educação não-formal pode fomentar nos jovens o gosto pela aprendizagem sobre tecnologia e prepará-los para um ambiente tecnológico em constante mudança.
- **Construir competências práticas:** os programas de educação não-formal podem fornecer competências práticas em áreas como a programação computacional, o *web design* e o desenvolvimento de aplicações, que podem ajudá-los a construir competências profissionais e a prepararem-se para carreiras no domínio da tecnologia.

Em suma, o papel da educação não-formal na aprendizagem sobre IA e Tecnologia é proporcionar um ambiente complementar, centrado na pessoa e motivador, que ajude os jovens a desenvolver as competências e o conhecimento que necessitam para prosperar na era digital (4).

1.2. Promover reflexão sobre género, diversidade e inclusão no ambiente de aprendizagem

Promover a reflexão sobre as questões de género, a diversidade e inclusão em programas de educação não-formal sobre IA é crucial para promover uma indústria tecnológica mais equitativa. A IA é um campo em rápido crescimento e é essencial que indivíduos de todos os contextos e experiências tenham acesso à educação e à prática em IA. Os programas de educação não-formal podem desempenhar um papel importante na criação de oportunidades para que pessoas de comunidades sub-representadas aprendam sobre IA e desenvolvam as competências necessárias para não só contribuir para uma tecnologia mais justa mas também poderem desenvolver carreiras de sucesso nestas áreas.

Para promover a diversidade e a inclusão nos programas de educação não-formal sobre IA, é fundamental assegurar, em primeiro lugar, que os programas sejam flexíveis, acessíveis e inclusivos. Além de se certificar que isso aconteça, devem ser assegurados serviços

de apoio para garantir que pessoas de comunidades sub-representadas possam participar efetivamente nesses programas. Outra questão relevante neste sentido é garantir que os programas sejam ministrados por formadores que reflitam a diversidade dessa comunidade e com experiência em trabalhar com pessoas da mesma.

Outro aspeto importante para promover a diversidade e inclusão em programas de educação não-formal sobre IA é focar tanto nos aspectos técnicos quanto éticos da IA. Competências técnicas, como código e programação, são essenciais para pessoas que desejam trabalhar com IA. No entanto, também é importante ensinar sobre as implicações éticas e sociais da IA, incluindo questões relacionadas com a privacidade, proteção de dados e viés nos sistemas de IA. Ao incluir tanto os aspectos técnicos quanto éticos da IA em programas de educação não-formal, as pessoas podem desenvolver uma compreensão abrangente da IA e estar mais bem preparadas para tomar decisões informadas sobre seu uso (5).

Podemos melhorar a qualidade da educação para todos os/as estudantes criando um ambiente de trabalho inclusivo. **Mas o que é que isso significa?**

Género, diversidade e inclusão na educação referem-se ao reconhecimento e valorização das identidades diversas, experiências e perspetivas de estudantes, educadores e outros envolvidos no sistema educativo. Envolve a criação de um ambiente de aprendizagem inclusivo e equitativo, onde todas as pessoas se sintam valorizadas e apoiadas, independentemente do seu género, raça, etnia, religião, orientação sexual, habilidade, idade ou contexto socioeconómico. Isso inclui abordar questões de discriminação, preconceito e marginalização e implementar estratégias para apoiar o sucesso académico e pessoal de todos os estudantes. Género, diversidade e inclusão na educação também envolve promover ativamente a equidade e justiça social, e desafiar sistemas e estruturas que perpetuam a desigualdade e a exclusão. (6)

1.3. Práticas de Ensino inclusivas na educação não-formal

Como educador/a, o seu papel é fundamental para ajudar os estudantes a construírem uma base sólida para o futuro. Promover um ambiente inclusivo significa reconhecer as diferenças entre estudantes, assim como as suas semelhanças. Significa garantir que todas as pessoas se sintam à vontade com quem são e dar-lhes oportunidades de explorar os seus interesses fora do contexto escolar. Ao fazer isso, ajudamos a criar um ambiente mais inclusivo para todos os envolvidos. Aqui estão algumas maneiras de o fazer:

1. **Certifique-se se está consciente dos seus próprios preconceitos e estereótipos**

Se se aperceber que está a fazer suposições sobre as pessoas com base na sua aparência ou comportamento, dê um passo atrás e tente olhar sob uma perspetiva diferente! Faça um esforço para questionar as suas suposições. Pode nem sequer se aperceber que fez certas considerações, mas esse é um bom começo para tomar consciência dos seus próprios preconceitos.

Todas as pessoas têm crenças e atitudes em relação a grupos de pessoas com base na sua raça ou etnia, género, peso corporal e outros traços, por exemplo. Essas crenças e atitudes em relação a grupos sociais são conhecidas como **preconceitos**. Preconceitos são crenças sobre alguém ou sobre um grupo específico de indivíduos, não fundamentadas em factos conhecidos. (7).

Porque é que os preconceitos existem? Na maioria dos casos, os preconceitos formam-se devido à tendência do cérebro humano categorizar novas pessoas e novas informações. Para aprender rapidamente, o cérebro liga novas pessoas ou ideias a experiências passadas. Uma vez que algo de novo tenha sido colocado numa categoria, o cérebro responde a isso da mesma maneira que responde a outras coisas nessa categoria. Mais concretamente:

- **Muita informação:** o nosso mundo possui tanta informação que precisamos de atalhos e truques para reconhecer o que é importante;
- **Sem sentido suficiente:** preenchemos lacunas de informação usando o que pensamos já saber;
- **Necessidade de agir rapidamente:** precisamos avaliar e aplicar informações, tentar prever o futuro e agir com base em novos conhecimentos – mas não temos muito tempo;
- **Do que é que nos devemos lembrar?** Não nos podemos lembrar de tudo, então tentamos reter o que é importante para o futuro e esquecer o resto!

Aqui estão alguns exemplos de preconceitos que os educadores podem ter e que podem prejudicar o processo educativo:

- **Preconceito racial:** Um/a educador/a pode inconscientemente ter estereótipos negativos sobre certos grupos étnicos, e isso pode afetar o seu estilo de ensino e comportamento em relação aos estudantes desses grupos (8).
- **Preconceito de gênero:** Um/a educador/a pode ter estereótipos de gênero que afetam as suas interações com estudantes do sexo masculino e feminino. Por exemplo, pode assumir que os meninos são melhores em matemática e ciências, enquanto as meninas são melhores em línguas e artes (9).
- **Preconceito etário:** Um/a educador/a pode ter preconceitos em relação a estudantes de diferentes idades. Por exemplo, pode assumir que as pessoas mais jovens são imaturas e têm dificuldade de concentração, enquanto as mais velhas são inflexíveis demais para aprender coisas novas (10).
- **Preconceito em relação a competências:** Um/a educador/a pode ter preconceitos em relação a estudantes com deficiências ou necessidades especiais, assumindo que são incapazes de aprender ou participar da mesma forma que os outros estudantes (11).
- **Preconceito socioeconómico:** Um/a educador/a pode inconscientemente ter preconceitos com base no estatuto socioeconómico de um aluno ou aluna, assumindo que pessoas de famílias com um rendimento mais baixo são menos inteligentes ou motivados (12).

2. Esteja consciente do tipo de linguagem que usa

Ser inclusivo significa usar palavras como “todos” em vez de frases como “meninos e meninas” ou “aqueles que se identificam como meninos ou meninas”. Tente evitar fazer suposições sobre a identidade de gênero, perguntando aos estudantes que pronomes preferem (por exemplo: ele/dele, ela/dela, eles/delas) em vez de presumir que todos usam pronomes masculinos.

3. Certifique-se que está ciente das necessidades de todos os estudantes

Tal significa dedicar tempo para fazer perguntas sobre como se sentem e o que querem aprender durante a aula. Certifique-se que inclui todos e todas nas discussões e atividades, além de oferecer oportunidades para aqueles que possam sentir-se excluídos à partida.

4. Faça um esforço para criar um ambiente seguro para os seus estudantes

Isso pode significar estabelecer expectativas claras de comportamento e aplicá-las de forma consistente. Também significa estar atento ao que diz e ao que faz à frente dos estudantes, além de garantir que não os/as excluiu involuntariamente ou os/as deixa desconfortáveis.

5. Por fim, se está a enfrentar dificuldades enquanto educador/a em qualquer um destes passos, é importante procurar apoio

Muitas organizações têm pelo menos um membro da equipa treinado para ajudar educadores/as que enfrentam questões de diversidade. Se a sua organização não tiver essa ajuda, tente procurar parceiros locais que ofereçam esse tipo de apoio (13).

Programas de educação não-formal têm o potencial de desempenhar um papel fundamental na promoção da igualdade de género, diversidade e inclusão na IA, tanto no que diz respeito à criação de um ambiente de aprendizagem inclusivo, como no destaque à importância da diversidade no desenvolvimento de sistemas de IA. Ao aproveitar as vantagens da educação não-formal, como a flexibilidade e a adaptabilidade, e ao focar-se nos aspetos técnicos e éticos da IA, as pessoas de comunidades sub-representadas podem desenvolver as competências e os conhecimentos necessários para ter sucesso nesse campo em rápido crescimento. Ao disponibilizar o acesso à educação e à formação em IA, os programas de educação não-formal podem contribuir para a construção de uma indústria de tecnologia mais inclusiva e equitativa, onde pessoas de todos os contextos e experiências podem alcançar o seu máximo potencial e contribuir para o desenvolvimento de um mundo mais justo.

As práticas de ensino inclusivas são essenciais para criar um ambiente de aprendizagem favorável a todos os estudantes. Inclusão significa que todos os estudantes são valorizados e respeitados, independentemente das suas origens, experiências e competências. Essa abordagem à educação ajuda a promover o envolvimento dos estudantes, o sucesso académico e o bem-estar social e emocional.

Um dos princípios-chave do ensino inclusivo é reconhecer e respeitar a diversidade das pessoas na sala de aula. Isso inclui reconhecer e valorizar as diferentes origens, culturas e experiências que os estudantes trazem para o ambiente de aprendizagem. Ao abraçar essa diversidade, os docentes podem ajudar os estudantes a sentirem-se vistos e valorizados, e podem criar um sentido de comunidade na sala de aula.

Outro aspecto importante do ensino inclusivo é assegurar as condições apropriadas a estudantes com necessidades diversas. Isso pode incluir o uso de tecnologia, como legendas ou descrições em áudio, para apoiar estudantes com deficiências auditivas ou visuais, ou fornecer métodos alternativos de avaliação para estudantes com dificuldades de aprendizagem. Os formadores também devem ser flexíveis nas suas abordagens de ensino, adaptando os seus métodos às necessidades individuais de cada pessoa.

O ensino inclusivo também envolve criar um ambiente de aprendizagem seguro e solidário, onde todos os estudantes se sintam à vontade para se expressar e fazer perguntas. Isso significa criar uma cultura, livre de discriminação e assédio, e promover ativamente o respeito e a aceitação de todos os estudantes. Os formadores podem alcançar isso estabelecendo expectativas claras de comportamento, promovendo relacionamentos positivos entre estudantes e abordando pronta e efetivamente quaisquer incidentes de *bullying* ou discriminação.

Além disso, as práticas de ensino inclusivas procuram envolver os estudantes no processo de aprendizagem e incentivá-los a desempenhar um papel mais ativo na sua própria educação e formação. Isso implica oferecer oportunidades para que estes participem ativamente em debates e discussões na sala de aula, façam perguntas e partilhem as suas ideias e perspetivas. Os formadores também devem incentivar a sua colaboração, ajudando-os a desenvolver importantes competências sociais e comunicacionais (14).

Estão aqui alguns exemplos de práticas de ensino inclusivas que podem ser usadas em programas educativos de IA:

- **Representação diversificada no material:** Incluir exemplos de diferentes pessoas, e culturas no material do curso e nos estudos de caso.
- **Acessibilidade:** Utilizar a tecnologia e métodos alternativos de aprendizagem para tornar o curso acessível a todos os estudantes, independentemente das suas competências.
- **Linguagem Inclusiva:** Evitar o uso de linguagem discriminatória, tal como linguagem de género ou estereotipada, e optar por uma linguagem inclusiva nas discussões e nas atividades na sala de aula.
- **Estimular diferentes perspetivas:** Criar oportunidades para que as pessoas partilhem as suas perspetivas e experiências únicas, e incentivar a diversidade de perspetivas nos debates na sala de aula.
- **Trabalho de grupo:** Incentivar ao trabalho em grupos diversos, ajudando-os a construir relacionamentos positivos e a desenvolver importantes competências sociais e comunicacionais.
- **Abordagens de ensino flexíveis:** Ser flexível nas abordagens de ensino e adaptar-se às necessidades individuais das pessoas, em vez de adotar uma abordagem única para todos.
- **Fornecer contexto:** Incluir o contexto histórico e cultural no ensino de IA para ajudar os estudantes a entender o seu desenvolvimento e impacto em diferentes comunidades.
- **Estimular debates éticos:** Incentivar discussões sobre as implicações éticas e sociais da IA e como esta pode impactar diferentes comunidades e indivíduos.
- **Fornecer oportunidades para reflexão:** Oferecer oportunidades para que os estudantes reflitam sobre os seus próprios preconceitos e experiências, e incentivá-los a tornarem-se defensores ativos da diversidade, equidade e inclusão no campo da IA.

Ao incorporar estas práticas de ensino inclusivas, os/as educadores/as podem ajudar os estudantes a desenvolver uma compreensão mais diversa e inclusiva da IA, e capacitá-los com as competências necessárias para contribuir para um futuro mais equitativo (15).



2

Diversidade e inclusão
na inteligência artificial

2.1. Inteligência artificial – o que é?

O objetivo do nosso guia é dar uma visão geral da percepção pública sobre a IA e, por essa razão, não nos debruçaremos a fundo nos seus aspectos técnicos, mas na forma como molda o mundo à nossa volta e porque é que é importante.

Pode encontrar mais informação, instruções sobre o ensino da IA, e planos de aula no “Guia para professores em educação formal”, desenvolvido no âmbito do projeto FAlaS – Fostering Artificial Intelligence at Schools, que pode ser descarregado gratuitamente aqui: <https://fosteringai.github.io/project/result1/>

A Inteligência Artificial (AI) é uma disciplina das ciências da computação que tenta replicar e desenvolver a inteligência humana e os seus processos implícitos, através de computadores. O termo foi cunhado em 1956 na Conferência de Dartmouth pelo reconhecido cientista computacional John McCarthy, embora algumas experiências remontem ao período imediatamente após a II Guerra Mundial.

Hoje, a inteligência artificial (IA) é considerada uma das tecnologias definidoras da última década e, provavelmente, também da próxima. Atualmente, é um “tópico quente”: a cobertura mediática e a discussão pública sobre IA são quase impossíveis de evitar. Empresas como a Google, o Facebook e a Apple estão a desenvolver bots inteligentes e assistentes de IA que prometem mudar a forma como trabalhamos, nos divertimos e aprendemos. Para onde quer que olhe, há evidências disso: o seu smartphone ouve os seus comandos e entende o que está a dizer; os veículos autónomos estão a tornar-se uma realidade; a robótica continua a evoluir em muitas fábricas por todo o mundo. A computação em nuvem tornou a IA mais acessível do que nunca. Como resultado, *developers* estão a criar uma ampla gama de produtos com IA: interagindo com clientes por meio de chatbots; ajudando empresas a prever o que os clientes desejam e quando irão comprar; realizando diagnósticos médicos automaticamente numa questão de minutos, em vez de dias ou semanas (16).

Ainda que não haja acordo sobre uma definição única da IA, existe um consenso em quatro abordagens principais: duas delas são baseadas nos seres humanos (sistemas que pensam como eles e sistemas que agem como eles), e outras duas, baseadas em aspectos racionais (sistemas que pensam racionalmente e sistemas que agem racionalmente).

Alguns exemplos de sistemas de IA incluem:

- Software de reconhecimento de fala, que é capaz de compreender reconhecer e transcrever palavras ditas;
- Sistemas de reconhecimento de imagem, que pode identificar objetos ou pessoas em imagens ou vídeos;
- Processamento de linguagem natural (NLP), que permite a máquinas compreenderem e gerarem linguagem humana;
- Algoritmos de tomadas de decisão, que podem analisar dados e tomar decisões baseadas nessa análise.

Na educação e formação há muitas formas de usar a IA para melhorar a experiência e os resultados da aprendizagem. Alguns exemplos são (17):

- **Sistemas de aprendizagem adaptativos:** estes usam a IA para personalizar as experiências de aprendizagem para cada estudante, baseando-se nos seus pontos fortes, nas suas fraquezas e estilo de aprendizagem. Este sistema pode ajustar o conteúdo e ritmo do material de aprendizagem para melhor se encaixar nas necessidades individuais do aluno.
- **Sistemas tutoriais inteligentes:** estes usam a IA para fornecer tutorias personalizadas e *feedback* aos estudantes, à medida que estes desempenham tarefas e resolvem problemas. Estes sistemas podem ajudar os estudantes a identificar áreas onde precisam de apoio adicional e disponibilizar diretrizes e recursos para os ajudar a aprender de forma mais eficiente.
- **Tradução de idiomas e ferramentas de aprendizagem de línguas:** a IA pode ser usada para traduzir texto e discurso de uma língua para a outra, facilitando a aprendizagem e comunicação de estudantes com pessoas que falem outras línguas.
- **Ferramentas de avaliação e pontuação:** a IA pode ser usada para pontuar exercícios e avaliações, dando *feedback* imediato e libertando os/as educadores/as, para que estes se possam focar em tarefas mais complexas.
- **Assistentes virtuais:** Assistentes virtuais com recurso a IA que podem ajudar estudantes a encontrar informação, concluir tarefas e responder a questões, libertando os/as docentes para que estes se possam focar no ensino e na aprendizagem.
- **Recomendações personalizáveis:** a IA pode analisar o histórico da performance de um/a estudante para recomendar materiais de aprendizagem e recursos que serão mais úteis para o/a envolver e estimular.

Porque a inteligência artificial está a moldar cada vez mais o mundo em que vivemos, é importante, não apenas aprender sobre seu impacto e possíveis implicações, mas também sobre como ser ativo no desenvolvimento da tecnologia de forma a beneficiar toda a sociedade. Aqui estão algumas razões pelas quais aprender sobre inteligência artificial é importante:

- **Preparação para o futuro:** A inteligência artificial está a ser rapidamente integrada em muitas indústrias, e aprender sobre a mesma pode ajudar a prepararmo-nos para futuras oportunidades de emprego e garantir que estamos capacitados a prosperar num ambiente de trabalho em constante mudança.
- **Compreensão dos seus impactos:** A inteligência artificial tem o potencial de ter impactos significativos na sociedade, tanto positivos quanto negativos. Ao aprender sobre IA, podemos ter uma melhor compreensão das suas possíveis implicações, incluindo a possibilidade de substituição de empregos, preocupações com privacidade e considerações éticas.
- **Tomar decisões informadas:** A inteligência artificial está a ser cada vez mais usada para tomar decisões que afetam a vida das pessoas, desde cuidados de saúde até à justiça penal. Aprender sobre inteligência artificial pode ajudar-nos a tomar decisões informadas sobre essas aplicações e advogar por um uso responsável e ético da mesma.
- **Fomentar a inovação:** A inteligência artificial tem o potencial de impulsionar a inovação tecnológica e criar produtos e serviços. Ao aprender sobre inteligência artificial,

podemos compreender melhor as suas possíveis aplicações e contribuir para moldar o seu desenvolvimento de maneira alinhada com os nossos valores e prioridades.

→ **Desenvolver competências de pensamento crítico:** Aprender sobre inteligência artificial requer pensamento crítico e competências para avaliar informações complexas. A IA pode ajudar as pessoas a desenvolverem as suas competências de pensamento crítico e análise de informações.

Aprender sobre IA é essencial para que as pessoas compreendam o seu impacto e estejam preparadas para navegar nas suas possíveis implicações. Ao investir na educação sobre a IA, os indivíduos podem assegurar que estão preparados para participar na construção de um futuro mais equitativo e responsável (18).

2.2. Viés em Inteligência Artificial

Os sistemas automáticos não são intrinsecamente neutros. Os computadores têm de ser programados por pessoas e, como consequência, refletem preconceitos e preferências de quem programa os algoritmos IA por detrás destes sistemas. Aqui temos alguns exemplos de possíveis causas de vieses em IA.

Viés na inteligência artificial refere-se à tendência dos sistemas de IA em exibir certos preconceitos ou estereótipos nos seus processos de tomada de decisão e resultados. Tal pode ocorrer devido aos dados que foram utilizados para treinar o sistema de IA, bem como aos algoritmos e modelos utilizados para construir e operar o sistema. O viés na IA acontece porque os seres humanos escolhem os dados que os algoritmos utilizam e decidem como serão aplicados os resultados desses algoritmos. Sem testes extensivos e equipas diversificadas, é fácil que vieses inconscientes entrem nos modelos de *machine learning*. Assim, os sistemas de IA tendem a automatizar e perpetuar esses modelos tendenciosos.

Uma forma muito comum de introduzir vieses nos sistemas IA é através de **dados enviesados**. Por exemplo, suponha que quer automatizar o processo de decisão acerca da aceitação ou não de estudantes numa escola ou universidade. Suponha que, no passado, se aceitaram sobretudo homens. Se se utilizarem estes dados históricos para treinar o novo sistema, provavelmente este viés de género é replicado.

Ao usar dados do mundo real, como artigos de notícias ou publicações das redes sociais, os sistemas de IA tendencialmente reforçam preconceitos existentes. As aplicações de tradução automática, por exemplo, tendem a apresentar um viés de género em prejuízo das mulheres, quando traduzem a partir de línguas que usam pronomes neutros no género. Exemplos típicos são “ela toma conta das crianças” e “ele investe” (19).

Quando analisamos a classificação de dados, **a forma como os dados são recolhidos é muito importante. Tanto a subamostragem como a sobreamostragem podem conduzir a vieses**. Um exemplo de subamostragem é quando sistemas de deteção de faces são sobretudo treinados para reconhecer caras de homens brancos, e, como consequência, não reconhecem caras de mulheres mais escuras tão bem como caras

de homens mais brancos, por exemplo Sobreamostragem, por outro lado, pode conduzir à sobre-representação de certos grupos. Por exemplo, crimes cometidos em áreas frequentadas pela polícia estão mais representados no conjunto de dados e o modelo IA treinado nesses dados irá provavelmente refletir esta desproporção e conseqüentemente essas áreas serão marcadas como mais perigosas.

As pessoas podem também **reforçar vieses em sistemas IA já implementados**. Por exemplo, um sistema IA pode já mostrar anúncios de cartões de crédito com juros elevados a pessoas com um baixo grau de escolaridade. Elas podem clicar neste anúncio sem se aperceber que isto vai reforçar o viés já existente e que irão continuar a receber tais sugestões.

O viés na IA pode ter impactos negativos significativos na sociedade e nos indivíduos. Os impactos potenciais do viés na IA incluem:

Resultados discriminatórios

Os sistemas IA que estão enviesados podem produzir resultados que criem desvantagens injustas para certos grupos de pessoas, como mulheres ou grupos minoritários. Isso pode conduzir à discriminação e desigualdade em áreas como o emprego e cuidados de saúde.

O algoritmo dos anúncios do Facebook

Em 2019, o Facebook (20) foi processado pelo Departamento de Habitação e Desenvolvimento Urbano dos Estados Unidos porque se constatou que a empresa permitia que os seus anunciantes direcionassem deliberadamente anúncios com base no gênero, raça e religião, todas categorias protegidas pelo sistema legal do país. Anúncios de empregos para funções em enfermagem ou trabalho de secretariado eram sugeridos principalmente a mulheres, enquanto anúncios de empregos para motoristas de táxi eram exibidos em maior número a homens, especialmente homens pertencentes a minorias étnicas. O algoritmo aprendeu que anúncios de imóveis tinham mais probabilidades de obter estatísticas de interação positivas quando exibidos a pessoas brancas, fazendo com que não fossem exibidos a grupos minoritários.

Esse problema decorre da forma como a máquina de IA aprende. Como é característico dos algoritmos de *machine learning*, a plataforma de anúncios criou um padrão a partir dos dados que lhe foram fornecidos, mas esse padrão já refletia desigualdades sociais existentes e, se não fosse corrigido, poderia contribuir para as propagar ainda mais. Essa prática é conhecida como “publicidade baseada na discriminação” e foi descoberta como algo que perpetua estereótipos prejudiciais e discriminação na sociedade. Esse problema destaca a importância das considerações éticas no desenvolvimento e uso da IA. É fundamental que os algoritmos sejam projetados e monitorizados de forma a não perpetuar ou ampliar preconceitos e discriminação já existentes.

Oportunidades limitadas

Vieses em sistemas IA podem limitar oportunidades para certos grupos de pessoas. Por exemplo, se um sistema IA é enviesado contra as mulheres, pode não recomendá-las para certos cargos ou programas educativos, conduzindo a uma ausência de progresso e crescimento para essa franja da população.

A ferramenta de recrutamento enviesada da Amazon

Em 2018 (21), a Reuters informou que a Amazon estava a trabalhar num sistema de recrutamento IA projetado para agilizar o processo de contratação, ler currículo e selecionar os candidatos mais qualificados. Infelizmente, a IA parecia ter uma questão com as mulheres, e ficou evidente que o algoritmo tinha sido programado para replicar as práticas de contratação existentes, o que significa que também reproduziu os seus preconceitos. Tal ocorreu porque os modelos de computador da Amazon foram treinados para avaliar candidatos observando padrões em currículo enviados para a empresa ao longo de um período de 10 anos. A maioria desses currículos era de homens, refletindo a predominância masculina na indústria de tecnologia.

De facto, o sistema de IA da Amazon aprendeu que candidatos do sexo masculino eram preferíveis. O sistema penalizava currículos que incluíam a palavra “mulheres”, como por exemplo “capitã do clube de xadrez feminino”. E desvalorizava os currículos de estudantes de faculdades exclusivamente femininas. Em vez de ajudar a eliminar os preconceitos presentes no processo de recrutamento, o algoritmo simplesmente automatizou-os. A Amazon reconheceu este enviesamento e eliminou o sistema.

O viés na contratação e no recrutamento pode acontecer quando certos grupos de pessoas são injustamente prejudicados ou excluídos de oportunidades de emprego, com base em fatores como raça, género, idade, religião, orientação sexual ou estatuto de deficiência. Isso pode conduzir à falta de diversidade no ambiente de trabalho e ter consequências negativas, tanto para os indivíduos afetados, como para a empresa como um todo.

É importante que as empresas estejam cientes do risco de viés em seus processos de contratação e recrutamento e tomem medidas para mitigá-lo. Isso pode incluir o uso de fontes de recrutamento diversas, a implementação de critérios de avaliação justos e objetivos, e a formação dos funcionários em questões de diversidade, equidade e inclusão.

Desinformação

Vieses nos sistemas IA podem conduzir à disseminação de informação falsa ou enganadora. Por exemplo, se um sistema IA for tendencioso contra um certo grupo de pessoas, pode propagar desinformação sobre esse grupo, levando a estereótipos negativos e a discriminação.

Sistemas de recomendação de notícias que usam algoritmos IA para personalizar conteúdos baseados nos hábitos de leitura e preferências podem também perpetuar desinformação. Se esses algoritmos são enviesados em relação a determinadas perspectivas, podem reforçar crenças existentes e difundir informação falsa ou enganadora.

Tomada de decisão ineficaz

Vieses nos sistemas IA podem conduzir a uma tomada de decisão de baixa qualidade, pois o sistema IA pode não considerar de forma correta todos os fatores relevantes. Isto pode conduzir a consequências negativas em áreas como os cuidados de saúde, justiça e finanças.

Racismo na disponibilização de cuidados de saúde nos Estados Unidos

Em 2018, uma equipa da Universidade da Califórnia, Berkeley (22), identificou um problema com um sistema de IA que estava a ser usado para fornecer cuidados de saúde a 200 milhões de pacientes nos Estados Unidos e que resultou em que pacientes negros recebessem um nível de atendimento abaixo do padrão. O problema decorreu do facto de o sistema atribuir valores de risco usando o custo previsto dos cuidados de saúde como variável determinante, e como os pacientes negros muitas vezes tinham frequentemente menos capacidade para pagar, ou eram percebidos como tendo recursos financeiros limitados para adquirir cuidados de alta qualidade, o sistema de IA basicamente aprendeu que não tinham direito a tal padrão de qualidade no atendimento. De modo geral, as pessoas negras receberam pontuações de risco mais baixas do que as pessoas brancas, apesar de os pacientes negros serem estatisticamente mais propensos a ter condições comórbidas e, portanto, de facto, enfrentarem níveis mais altos de risco. Isso, por sua vez, significava que pacientes negros tinham menos probabilidade de aceder ao padrão qualidade de atendimento e de cuidados necessário e mais probabilidade de experimentar efeitos adversos, como resultado de isso lhes ter sido negado. Após esta constatação, a equipa da UC Berkeley trabalhou com a empresa responsável pelo desenvolvimento da ferramenta, no sentido de encontrar outras variáveis além do custo, para a atribuição dos valores de risco esperados, reduzindo o viés em 84%.

Existem inúmeros casos em que o viés na IA levou ao racismo na prestação de cuidados de saúde. Um exemplo é o uso de algoritmos para prever que pacientes

estão em risco de desenvolver certas condições ou doenças. Esses algoritmos são frequentemente treinados com dados que refletem as experiências de saúde de certas populações, o que significa que podem não prever com precisão os riscos de indivíduos de outros grupos raciais ou étnicos. Tal pode ter como resultado que os prestadores de cuidados de saúde aloquem recursos e tratamentos, de forma desproporcional, a certos grupos raciais ou étnicos, em vez de fornecer iguais cuidados a todos os pacientes.

Em suma, é crucial que os provedores de cuidados de saúde e empresas de tecnologia estejam cientes do potencial do viés na IA e tomem medidas para o mitigar. Isso pode envolver o uso de conjuntos de dados diversos para treinar algoritmos, implementar medidas para prevenir a discriminação e garantir que os indivíduos tenham a oportunidade de fornecer *feedback* sobre o uso da IA na área da saúde.

Viés na IA pode ter consequências sérias, pode conduzir ao tratamento injusto de certos grupos e pode perpetuar desigualdades já existentes. É importante reconhecer e abordar os vieses nos sistemas de IA.

O que é que podemos fazer acerca dos vieses na inteligência artificial? [23]

Mudar a forma como as pessoas são educadas e formadas sobre ciência e tecnologia

Uma forma de reduzir o viés na IA é mudar a forma como as pessoas são educadas e formadas sobre a ciência e tecnologia. Essa educação pode incluir proporcionar uma representação mais diversificada nos currículos de ciência e tecnologia, destacar o impacto dos preconceitos históricos e atuais nessa área e enfatizar a importância das considerações éticas no desenvolvimento e uso da IA.

Outra abordagem pode ser incorporar mais oportunidades de aprendizagem prática e experimental que permitam aos estudantes envolverem-se diretamente e compreender os possíveis vieses e as implicações da IA. Tal pode incluir projetos ou estudos de caso que desafiam os estudantes a identificar e abordar vieses em sistemas de IA ou a considerar as implicações éticas do seu uso.

Além disso, fornecer aos formadores mais recursos e apoio, de modo que estes entendam e abordem essas questões pode ajudar a criar um ambiente de aprendizagem mais inclusivo e equitativo para todos os estudantes. Tal pode incluir oportunidades de desenvolvimento profissional, acesso a investigação e recursos, e apoio contínuo, formação e orientação de especialistas na área.

Estar ciente dos contextos em que a IA pode ajudar a corrigir vieses, bem como do alto risco da IA agravar o viés

Ao implementar a IA, é importante antecipar os domínios potencialmente propensos a vieses injustos, como aqueles dos exemplos anteriores de sistemas tendenciosos ou com dados distorcidos. Tanto indivíduos quanto empresas têm algum grau de responsabilidade social e, portanto, temos a obrigação de regular os nossos processos de modelação para garantir que as nossas práticas são éticas. As organizações precisam de se manter atualizadas para saber como e onde a IA pode melhorar a equidade – e aprender onde os sistemas de IA têm tido dificuldades.

As empresas devem estabelecer um modelo de correção de vieses dentro da sua estratégia geral de IA, que contenha um portfólio de ações técnicas, operacionais e organizacionais:

- **Estratégias técnicas** envolvem ferramentas que podem ajudar a identificar as fontes potenciais de viés e revelar os traços nos dados que afetam a precisão do modelo;
- **Estratégias operacionais** incluem melhorar os processos de coleta de dados usando red teams (equipas vermelhas) internas e auditores terceirizados;
- **Estratégias organizacionais** incluem estabelecer um ambiente de trabalho onde as métricas e processos são apresentados de uma forma transparente.

Considerar os sistemas *Human-in-the-Loop*

O objetivo da tecnologia *Human-in-the-Loop* é fazer o que nem um ser humano nem um computador podem realizar sozinhos. Quando uma máquina não consegue resolver um problema, os humanos devem intervir e resolver o problema por ela. Ao fazer isso, é criado um ciclo contínuo de *feedback*. Com *feedback* contínuo, o sistema aprende e melhora o seu desempenho a cada execução subsequente. Como resultado, o *Human-in-the-Loop* leva a conjuntos de dados raros mais exatos e ao aprimoramento da segurança e precisão.

A maioria dos algoritmos de *machine learning* mais populares requer grandes quantidades de dados rotulados para produzir resultados precisos. No entanto, existem muitos casos em que não há dados não rotulados suficientes como base. Por exemplo, se estiver à procura de exemplos de notícias falsas num idioma com apenas alguns milhares de falantes, pode ainda não existir nenhum exemplo de notícias falsas nesse idioma. Portanto, o algoritmo não tem uma base para aprender. Nesse caso, manter os humanos no processo pode garantir o mesmo nível de precisão, mesmo para tipos de dados mais raros. Esse é o caso do Facebook, que ainda mantém humanos envolvidos na monitorização e na abordagem de atividades nas plataformas de media sociais.

2.3. Porque é que género, diversidade e inclusão são importantes para a inteligência artificial?

Os sistemas de IA são desenhados e construídos por seres humanos e, muitas vezes, refletem os preconceitos e estereótipos das pessoas que os desenvolvem. Se a equipa de desenvolvimento da IA não for diversificada, o mais provável é que o sistema de IA perpetue preconceitos e estereótipos prejudiciais. Equipas que apresentam diversidade em termos de raça, género, orientação sexual, idade ou condições económicas desenvolvem melhores sistemas de IA, porque pessoas diversificadas pensam de forma diferente e conseguem ter mais e novas ideias, são melhores a identificar preconceitos e promovem um pensamento mais criativo. A diversidade cognitiva é essencial no desenho da IA.

Quando o YouTube (24) lançou a aplicação para upload de vídeos para iOS, entre 5 e 10% dos vídeos enviados pelos utilizadores estavam virados ao contrário. Estes vídeos foram filmados de forma incorreta? Não. O problema estava no design inicial da aplicação. Este foi projetado para utilizadores destros, mas os telemóveis geralmente giram 180 graus quando segurados com a mão esquerda. Sem se aperceber, o YouTube criou uma aplicação que funcionava melhor para a equipa de *developers*, que era quase exclusivamente constituída por destros.

No entanto, a realidade é que os dados podem ser interpretados em função de quem os lê. A interpretação dos dados feita pela inteligência artificial pode ser tão tendenciosa quanto a inteligência humana que está por trás. O que pode acontecer quando uma excelente equipa de IA não é cognitivamente diversa? Como foi mostrado no exemplo anterior, o algoritmo de recrutamento da Amazon apresentou viés de género, descartando candidatas femininas e habilitadas para determinados cargos após aprender que a percentagem de mulheres nessas posições era menor.

Promover a igualdade de género, a diversidade e a inclusão na IA pode ajudar a garantir que as tecnologias de IA sejam desenhadas e desenvolvidas com uma abordagem centrada no ser humano. As mulheres e outros grupos sub-representados trazem perspectivas, experiências e habilidades únicas, o que pode levar ao desenvolvimento de tecnologias de IA mais criativas, inovadoras e eficazes que atendam melhor às necessidades da sociedade. Quando os sistemas de IA não são diversos e inclusivos, podem não corresponder adequadamente às necessidades e interesses de todos os membros da sociedade. Isto pode conduzir à falta de confiança nos sistemas de IA e a que não sejam adotados por certos grupos.

2.4. Mulheres que estão a mudar o campo da inteligência artificial

A contribuição das mulheres é fundamental para o futuro da Tecnologia e da IA. Ao aumentar a visibilidade do papel das mulheres no campo da Tecnologia e da IA e ao promovê-las, procuramos inspirar profissionais de educação não-formal a criar programas de formação mais inclusivos, onde o género, a diversidade e a inclusão desempenham um papel central no desenvolvimento de atividades e ambientes de formação.

É relevante ressaltar que a lista não é exaustiva, mas pretende oferecer um panorama geral das mulheres extraordinárias que são uma força motriz no campo da IA em todo o mundo. Ao ultrapassar os limites do que é possível e ao expandir o conhecimento sobre o vasto e diversificado potencial da IA na saúde, nas infraestruturas, clima, energia, linguagem (entre outros), essas mulheres desempenham um papel importante como inspiração para a próxima geração de cientistas, engenheiras e empreendedoras.

Carolina Barcenas [25]

Vice-Presidente Sénior para a Investigação na Visa

Lidera uma equipa, impulsionando a inovação em quatro áreas de grande relevância: Inteligência Artificial e *machine learning*, onde é responsável por explorar e desenvolver formas avançadas de aproveitar os dados para criar valor comercial para a Visa através de técnicas de IA; Blockchain; Segurança e Comércio do Futuro. A equipa concentra-se na investigação fundamental e orientada para produtos. A Carolina Bárcenas assumiu a co-liderança do grupo Visa Women in Technology em Austin e participa frequentemente em eventos que apoiam a emancipação das mulheres, como a Conferência de Mulheres do Texas, Cimeiras de mulheres latino-americanas em Tecnologia e a Girlstart. É a força por trás do programa de estágios da Universidade comunitária, com o foco em candidatos não tradicionais. Trabalhou tanto na indústria quanto no meio académico e possui mais de 20 anos de experiência em soluções de *machine learning* na *fintech*. Antes de começar a trabalhar na Visa, passou sete anos no PayPal, onde era responsável por gerir risco de pequenos e médios vendedores de comércio eletrónico. A sua especialização é em técnicas avançadas de data mining. Possui um doutoramento em Estatística Aplicada do Instituto de Tecnologia da Geórgia, como Bolseira Fulbright.

No meu caso, comecei a pensar que teria uma carreira enquanto física; no entanto, descobri que a minha paixão era resolver problemas e criar abstrações matemáticas para serem utilizadas em tomadas de decisão. Depois de perceber a minha paixão profissional, reajustei os meus objetivos e segui uma carreira em machine learning. Os objetivos mudam e, à medida que avança, não tenha medo de se adaptar. Em última análise, é fundamental trabalhar em algo que nos apaixona.

Stephanie Lampkin [26]

Fundadora técnica e CEO da Blendoor

Como fundadora e CEO da Blendoor, Stephanie Lampkin é apaixonada por utilizar a inteligência aumentada e a análise de pessoas para cruzar uma força de trabalho diversificada de candidatos com empresas, garantindo o anonimato dos candidatos no processo. Ela acredita que isso pode minimizar o viés inconsciente, dando mais oportunidades a pessoas qualificadas e contribuir para criar melhores empresas e economias mais sustentáveis. Uma das funções de Stephanie Lampkin é recolher dados para o BlendScore, um relatório publicado pela Blendoor que avalia empresas de tecnologia nas suas práticas de recrutamento, remuneração, retenção e promoção de mulheres e minorias sub-representadas.

Fundei a Blendoor três meses depois de ter uma experiência de entrevista bastante surpreendente com a Google. Na altura, eu morava em Nova Iorque e acabara de me licenciar no MIT Sloan. Quando estava a ser entrevistada para o cargo de Chefe Analista na Google, a equipa de recursos humanos disse-me que o cargo de Vendas ou o de Marketing seria o mais indicado para mim, porque não me viam como suficientemente preparada para a vaga. Antes dessa entrevista, já tinha experiência em programação e conhecimento em Ciências da Computação, tinha uma licenciatura em Engenharia na Universidade de Stanford, tinha trabalhado na Microsoft durante 5 anos e meio, e acabara de concluir um MBA. Havia alguma coisa sobre este feedback não me pareceu totalmente correto, mas recusei, educadamente, a posição de Vendas/Marketing e segui em frente. Mais tarde, descobri que, na época, de todos os trabalhadores da Google apenas 2% eram afro-americanos, 3% latinos e cerca de 25% mulheres. A narrativa que eles estavam a contar aos media era que era um problema de falta de talentos, que eles simplesmente não conseguiam encontrar mulheres e pessoas de cor qualificadas o suficiente. Isto levou-me a um momento de reflexão e a criar uma plataforma que facilitasse a conexão entre candidatos e empresas, para que estas possam ter acesso a um grupo mais amplo de talentos diversos. Foi assim que a Blendoor surgiu.

Freyja Jørgensen [27]

Gestora de inovação na Simula/Gründergarasjen, e Chefe do Grupo NORA.startup

Freyja Jørgensen tem muita energia e é o tipo de pessoa que simplesmente faz as coisas acontecerem. Um dos grupos mais ativos dentro do Consórcio de Investigação em Inteligência Artificial da Noruega (NORA) é a NORA.startup. A iniciativa NORA.startup é uma porta de entrada para uma maior cooperação entre a academia, incubadoras e empresas no campo da inteligência artificial, *machine learning* e robótica. Freya Jørgensen possui uma formação académica que cruza biologia molecular e psicologia. Possui uma ampla

experiência em trabalhar na interseção entre academia e startups, e esteve envolvida em diversos projetos educacionais e de inovação. Anteriormente, Jørgensen trabalhou na promoção da ciência e da tecnologia na educação, mas agora dedica o seu tempo ao apoio de empresas de tecnologia em fase inicial na incubadora Gründergarasjen. Também está a trabalhar no sentido de expandir ainda mais a rede NORA.startup e apoiar as startups de IA em particular. Jørgensen desempenha um papel fundamental na estruturação interna da iniciativa, além de gerir os eventos que contribuirão significativamente para a expansão do ecossistema de IA na Noruega.

Desde jovem, eu sabia que, de alguma forma, queria trabalhar com pessoas e ciência. Tinha o sonho de me tornar pediatra ou investigadora. Ter um avô com um doutoramento em química e um excelente professor de biologia no ensino secundário ainda me inspirou mais. Mais tarde, ao envolver-me em diversos projetos educacionais e de inovação na Simula, fiquei muito entusiasmada em trabalhar num ambiente dinâmico e interdisciplinar. Isto levou-me ao campo extremamente inspirador e em rápido crescimento das startups de tecnologia e IA. A IA e outras tecnologias vão continuar a desempenhar um papel central nas nossas vidas, e acho importante apoiar e compreender esse campo. Todos os dias aprendo algo novo e enfrento novos desafios e possibilidades no trabalho que faço – isso é uma das coisas principais que me motiva no meu trabalho.

Rebekah Agwunobi [28]

Estudante na Choate Rosemary Hall e vencedora da bolsa Caroline D. Bradley

O interesse de Rebekah Agwunobi por computadores foi apoiado pela sua família desde muito cedo e tudo começou com uma aula de JavaScript no 3º ano do 1º ciclo. Desde então, descobriu a IA e as suas aplicações para o bem comum no programa AI4ALL da Universidade de Stanford, passou dois verões como estagiária de investigação na USC, criou um programa de *machine learning* na sua escola secundária e muito mais. Juntamente com o interesse em tecnologia, é uma ativista apaixonada que se preocupa com questões como o encarceramento em massa, a reforma prisional e a importância da diversidade no STEM. Rebekah Agwunobi tem esperança que se possa usar a IA e as ciências da computação como ferramentas eficazes para resolver problemas do mundo real.

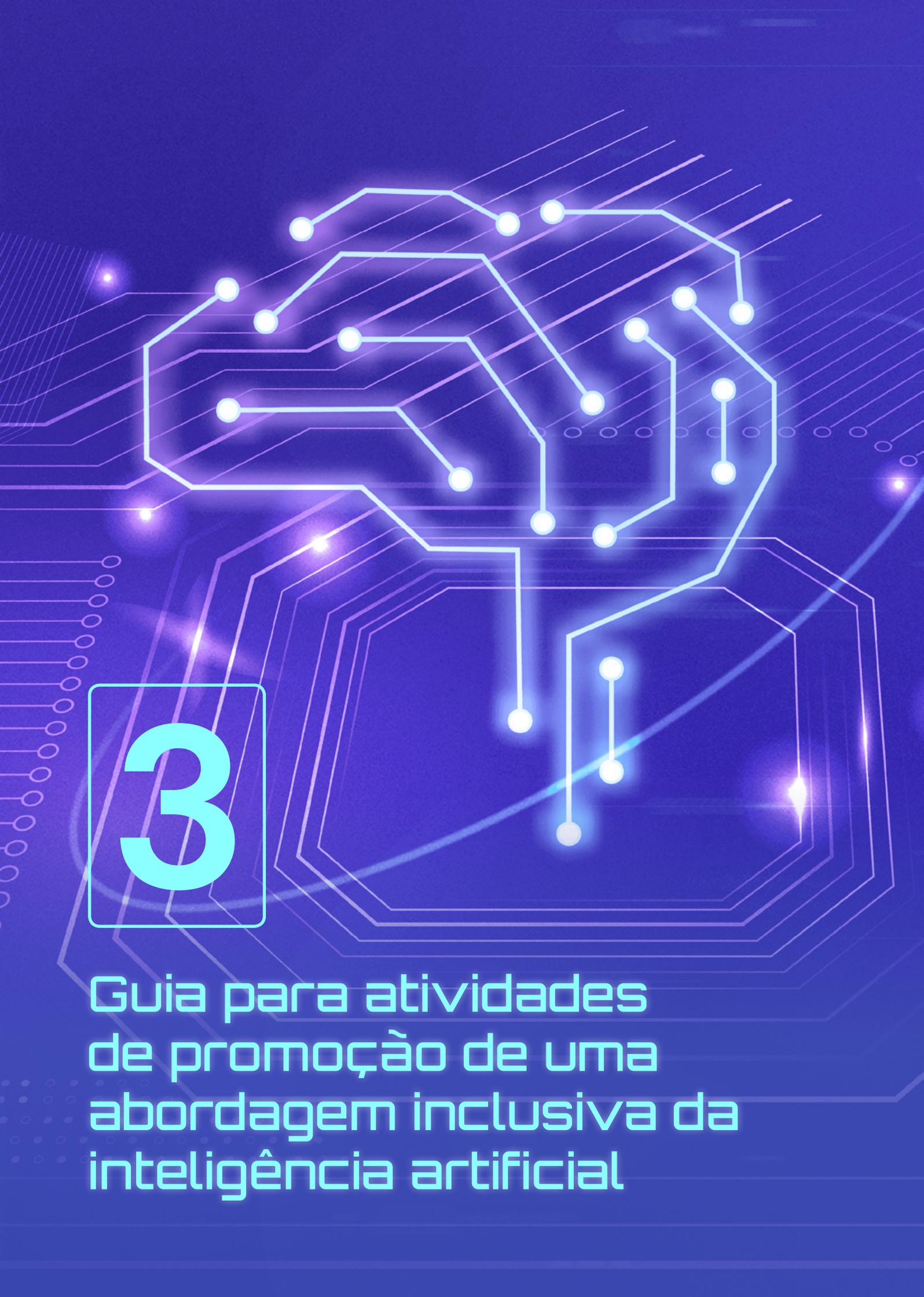
Eu queria alguma coisa incorporada no meu horário do curso que eu achasse empoderador e que me permitisse explorar coisas que me interessavam. Foi por isso que no último período estudei machine learning. Na aula, discutimos algumas das técnicas gerais que aprendi no AI4ALL de Stanford, e também abordamos conceitos mais avançados. Participamos nas competições Kaggle, e aprendi muito sobre diferentes aplicações de machine learning em áreas como criação artística e música. Aprendi muita teoria por trás da matemática que estava a fazer.

Abbey Lin [27]

Chief compliance officer da Cryon, ex-Chefe de controle e de Conformidade na Microsoft Noruega, Cofundadora e membro do Conselho de Administração da Oslo.ai, Fundadora da Ethical AI Resources

Abbey Lin trabalhou na área de Ética e Conformidade nos últimos 15 anos. Foi por meio dessas experiências que foi vendo os desafios éticos e as oportunidades de moldar a IA e quis que mais pessoas estivessem conscientes disso. Durante a sua licença maternidade, há alguns anos, criou o site e feed de notícias Ethical AI Resources, que reúne informações e as últimas notícias sobre uma IA responsável, e tornou-se uma fonte popular no Twitter www.twitter.com/Alethical (@aithical). Além disso, Abbey é cofundadora do Oslo.AI, que está a ajudar a construir e apoiar uma comunidade de entusiastas de IA em Oslo, através de eventos e oportunidades de networking. Abbey também é membro do conselho da Visualyst.co, uma startup de IA focada em vídeo compliance.

Se alguém se preocupa com diversidade e inclusão, também tem de se preocupar com a IA responsável. A IA tem a capacidade de impactar positivamente o mundo, mas também pode, e já tem, impactado negativamente comunidades já em si marginalizadas. A IA responsável precisa de ser considerada em todas as etapas do desenvolvimento do produto, e os consumidores de IA precisam de a questionar e exigir transparência sobre como o produto foi criado.



3

Guia para atividades
de promoção de uma
abordagem inclusiva da
inteligência artificial

Esta secção inclui um conjunto de atividades práticas, que organizações de educação e formação não formal podem incluir nos seus programas para ensinar as pessoas sobre IA.

As atividades propostas introduzem diferentes conceitos de IA e incluem descrições passo-a-passo de cada atividade, materiais a serem utilizados e recursos educacionais complementares para apoiar educadores criar e executar atividades envolventes e com significado.

Antes de começar, aqui estão algumas dicas para criar um ambiente de aprendizagem acolhedor e inclusivo para todas as pessoas, independentemente de suas origens ou conhecimento prévio sobre IA:

- Comece por apresentar o conceito de IA e explicar as suas várias aplicações. Pode usar vídeos, textos e outros materiais para apresentar uma visão geral da IA e do seu potencial impacto na sociedade.
- Certifique-se que enfatiza a importância da diversidade, equidade e inclusão no campo da IA. Isto pode incluir discutir as formas como a IA tem sido usada para perpetuar preconceitos e debater estratégias para criar sistemas de IA mais inclusivos e equitativos.
- Use estudos de caso e exemplos do mundo real para ilustrar as várias formas em que a IA está a ser utilizada em diferentes indústrias e setores. Isso pode ajudar os estudantes a ver as aplicações práticas da IA e entender como esta está a impactar o mundo à sua volta.
- Incentive o pensamento crítico sobre as implicações éticas da IA e a discussão em torno das potenciais consequências do uso da IA em diferentes contextos. Pode usar sugestões e perguntas para facilitar essa conversa.
- Considere incorporar atividades práticas, como exercícios ou projetos de código, para dar às pessoas a oportunidade de aplicar o que aprenderam e explorar conceitos de IA de forma mais interativa.
- Incentive os estudantes a manterem-se atualizados sobre os últimos desenvolvimentos em IA e a continuarem a aprender sobre a área mesmo após o fim das sessões. Isto pode incluir informar-se sobre notícias relevantes, participar em conferências e workshops e juntar-se a comunidades e organizações online relevantes nesta área.

3.1 Atividade 1: O jogo das *tags* [29]

Objetivos de aprendizagem

- Melhorar a comunicação não-verbal:** Os participantes irão utilizar gestos e sinais para formar grupos sem recorrer à comunicação verbal, estimulando assim competências de comunicação eficazes.
- Pensamento estratégico:** Os participantes irão desenvolver competências de pensamento estratégico ao planear e adaptar os seus movimentos durante o jogo.
- Comunicação e trabalho em equipa:** Os participantes vão comunicar de forma eficaz e trabalhar em equipa, coordenando os seus movimentos e estratégias com os seus colegas de equipa.
- Resolução de problemas:** Os participantes vão-se envolver na resolução de problemas ao antecipar e contrariar os movimentos dos taggers e encontrar formas criativas para evitar serem *tagados*.
- Espírito desportivo e jogo justo:** Os participantes aprenderão a importância do jogo justo, respeito e espírito desportivo ao seguir as regras do jogo e tratar os outros com justiça e respeito.
- Consciencialização dos vieses:** Os participantes tomarão consciência de seus próprios preconceitos, principalmente sobre o que é o viés de afinidade e o que todos temos em comum, mesmo que sejamos todos diferentes.

Materiais:

Papel de rascunho

Fita cola

Caneta ou lápis

Papel colorido

Descrição da atividade

Prepare-se com etiquetas ou adesivos (*tags*) para cada participante. Tem a opção de criar as suas próprias etiquetas recortando formas de papel colorido ou utilizar adesivos. Certifique-se de que as etiquetas estão disponíveis numa variedade de cores, formas, padrões e tamanhos. Por exemplo, pode ter várias etiquetas roxas, mas elas podem ser em forma de círculos, quadrados e triângulos, ou com padrões como quadrados, pontos e riscas.

O facilitador/líder começa a distribuir as *tags* individuais a todos os participantes. Cada pessoa deve colar o seu tag nas suas roupas, posicionando-o entre a cintura e os ombros. Assim que todos estiverem etiquetados, os participantes começam a movimentar-se pelo espaço, prestando atenção nos outros ao seu redor.

3. Guia para atividades de promoção de uma abordagem inclusiva da inteligência artificial

O facilitador/líder introduz então algumas variações, pedindo a todos para se movimentarem de formas diferentes, como saltar ou dar pulinhos. Após alguns minutos de movimentação, o líder instrui os participantes a formarem pequenos grupos sem utilizar comunicação verbal. Torna-se crucial observar como se dividem em grupos. Os participantes podem considerar os seus *tags* e organizar-se com base na cor ou forma. Uma vez que os grupos tenham sido formados, todos se distribuem novamente, continuando a movimentar-se pelo espaço.

O processo é repetido várias vezes, com os participantes a reagrupar-se sem falar. Conforme o jogo avança e mais grupos são formados, os participantes reúnem-se eventualmente num grande círculo. Neste momento, o facilitador/líder inicia uma discussão, incentivando todos a partilhar como decidiram formar os seus grupos. A conversa explora se alguém viu além dos *tags* ou criou intencionalmente grupos diversos abrangendo diferentes formas, cores, padrões e tamanhos.

Um guia passo-a-passo

Passo 1

Quem está a facilitar o jogo deve distribuir uma tag para cada participante. Cada pessoa deve colar o seu tag na roupa, entre a cintura e os ombros.

Passo 2

Todos devem movimentar-se pelo espaço e manter uma distância segura entre si.

Passo 3

Após um ou dois minutos, quem está a liderar o jogo deve pedir a todos que formem pequenos grupos sem falar.

Passo 4

Depois de formar os grupos, todos devem separar-se novamente e movimentar-se no espaço.

Passo 5

Assim que os grupos tiverem sido formados, todos devem se separar novamente e continuar a movimentar-se pelo espaço.

Passo 6

Todos devem reunir-se num grande círculo, mantendo uma distância segura entre si.

Discussão

A pessoa a liderar a atividade deve perguntar:

- Como é que decidiram a constituição dos grupos?
- Alguém viu para além das *tags*?
- Será que alguém formou intencionalmente um grupo diverso com uma variedade de formas, cores, padrões e tamanhos?

Reflexão

Através desta atividade, o conceito de diversidade foi apresentado aos participantes e, especificamente, o viés de afinidade, que é uma forma de viés inconsciente, foi explorado.

- Os indivíduos ficaram surpreendidos ao aprender sobre o viés de afinidade?
- Como se sentiram ao refletir sobre as formações de grupo durante o jogo dos *tags* e sobre a lista de pessoas em quem confiam?
- Quais são algumas das vantagens e desafios associados à colaboração em grupos, abrangendo indivíduos semelhantes, mas também diversos?

Os participantes são incentivados a refletir sobre os vários grupos aos quais pertencem, como uma turma escolar, uma equipa desportiva ou um círculo de amigos. Estes devem considerar se esses grupos consistem principalmente em indivíduos que são semelhantes ou diferentes de si mesmos. Em seguida, os participantes devem formar pares e envolver-se em conversas sobre estratégias para melhorar a compreensão da diversidade nas suas vidas quotidianas e nas suas interações. Por turnos, os participantes podem partilhar as suas ideias, se assim desejarem.

3.2 Atividade 2: Desenha a descrição [30]

Objetivos de aprendizagem

1. **Desenvolvimento das competências de observação:** Os participantes vão aprimorar a sua capacidade de observar e prestar atenção aos detalhes.
2. **Estímulo da criatividade:** Os participantes vão exercitar a sua criatividade e imaginação ao interpretar descrições em forma visual.
3. **Melhoria das competências de comunicação:** Os participantes vão praticar a comunicação eficaz ao transmitir com precisão as descrições por meio dos seus desenhos.
4. **Estímulo ao pensamento crítico:** Os participantes envolvem-se no pensamento crítico ao analisar as descrições e traduzi-las em representações visuais.
5. **Promoção do trabalho em equipa:** Os participantes vão colaborar uns com os outros, partilhando as suas interpretações e trabalhando juntos para resolver desafios.

Materiais

Canetas ou lápis de cor

Papel de rascunho

Impressões das fotografias e descrições das pessoas, objectos, cenas ou conceitos escolhidos

Descrição da atividade

Para iniciar a atividade, os participantes reúnem-se em grupo e recebem materiais de desenho, como papel e lápis. O facilitador ou participante designado assume o papel de descritor.

Este vai ler uma descrição de um objeto, cena ou conceito, sem revelar o que é. As descrições podem ser específicas ou abstratas, permitindo uma variedade de interpretações. Enquanto o facilitador/descritor lê a descrição, os participantes ouvem atentamente e visualizam os detalhes. Em seguida, têm um tempo determinado para traduzir as suas imagens mentais em desenhos no papel. Os desenhos devem capturar a essência do objeto ou da cena descrita.

Assim que acabar o tempo para desenhar, os participantes partilham os seus desenhos com o grupo. O descritor revela o objeto ou cena original, e os participantes comparam os seus desenhos para ver o grau de precisão com que capturaram a descrição. Tal incentiva a reflexão e a discussão, permitindo que os participantes explorem as diferenças e semelhanças nas suas interpretações.

3. Guia para atividades de promoção de uma abordagem inclusiva da inteligência artificial

Após cada ronda, os papéis de descritor e participantes podem ser alternados, garantindo que todos tenham a oportunidade de praticar a descrição e o desenho. A atividade pode ser personalizada variando a complexidade das descrições, introduzindo temas diferentes ou incorporando limites de tempo para adicionar um elemento de desafio.

Um guia passo-a-passo

Passo 1

Quem está a liderar a atividade deve ler a descrição do objeto, cena, pessoa ou conceito. Pode utilizar o exemplo abaixo, de Katherine Johnson, que foi uma das primeiras mulheres afro-americanas a trabalhar para a NASA sendo premiada com a Medalha Presidencial da Liberdade em 2015 [31].

Exemplo

Descrição da imagem: Nesta cena, vemos uma pessoa a receber um prémio presidencial em 2015. Esta pessoa foi matemática, nasceu em 1918 e deu importantes contributos nos primeiros tempos da NASA, tendo sido pioneira na utilização de computadores para muitas tarefas.



Passo 2

Todos devem desenhar a pessoa com base na descrição.

Passo 3

Dar a toda a gente uns minutos para finalizarem os seus desenhos.

Passo 4

Incentive todos a partilharem o que desenharam. Compare as semelhanças e diferenças entre os desenhos:

- Como estão vestidos?
- Qual é a sua estatura (alto/a, baixo/a, gordo/a, magro/a)?
- Qual é a cor da sua pele?
- Têm um género?
- Quantos anos têm?

Passo 5

Mostre a todos a imagem da primeira pessoa e revele o nome. Discuta qual o desenho que se assemelha mais à imagem real da pessoa. Quais são as diferenças mais significativas entre os desenhos e a imagem real?

Discussão

Pergunte às pessoas por que desenharam a descrição da forma como o fizeram. O que é que na descrição as levou a pensar que a pessoa tinha uma aparência específica? O que as fez escolher certa idade, raça, género, tamanho ou aparência para essa pessoa?

Reflexão

Nesta atividade, todos desenharam pessoas com base em descrições do trabalho que elas realizaram e do impacto que tiveram. Peça a todos que reflitam sobre os desenhos que diferiram muito da imagem real. Muitas vezes, associamos certas características a certos empregos, personalidades ou conquistas. Isso é um viés inconsciente e é causado por vários fatores, como aprender História apenas de uma perspectiva ou ver programas de TV e filmes em que os argumentos e personagens se inspiram em estereótipos ignorantes ou prejudiciais.

1. Reflexão individual: Após acabarem os desenhos, os participantes podem reservar um momento para refletir sobre suas próprias interpretações, considerando questões como:

- Como interpretei a descrição?
- Em que elementos me foquei e porquê?
- Captei a essência do objeto ou da cena descrita?
- Que desafios enfrentei durante o processo de desenho?
- O que podia ter feito de diferente para melhorar o meu desenho?

Incentivar os participantes a refletirem individualmente permite que eles ganhem informações sobre o seu próprio processo de pensamento, escolhas artísticas e áreas a melhorar.

2. Reflexão em grupo: Após partilharem os seus desenhos com o grupo, os participantes podem participar numa sessão de reflexão em grupo facilitada pelo descritor ou por um líder designado. Esta sessão pode incluir a discussão e a exploração das diferentes interpretações e variações nos desenhos. Algumas perguntas para reflexão em grupo podem incluir:

- Que semelhanças e diferenças notamos nos nossos desenhos?
- Que aspetos da descrição foram difíceis de captar visualmente?

3. Guia para atividades de promoção de uma abordagem inclusiva da inteligência artificial

- Como é que as variações nas nossas interpretações contribuíram para uma compreensão mais rica da descrição?
- Algum desenho ofereceu perspectivas ou insights únicos?

A reflexão em grupo permite que os participantes tenham uma perspetiva mais ampla das diversas interpretações e incentiva o diálogo aberto e a apreciação de diferentes abordagens. Também fomenta um sentido de aprendizagem coletiva e experiências partilhadas.

Explore várias abordagens para reduzir o viés inconsciente. Pergunte se alguém pode propor sugestões. Por exemplo, pode:

- Análise retratos positivos de indivíduos de diversos géneros, raças ou capacidades, referir livros como o livro de Martin Luther King Jr “Meninos Pequenos, Grandes Sonhos” [edição. Nuvem de Letras, 2021]” e “Exceptional Women in Black History” de Vashti Harrison ou filmes como “O Falcão Manteiga de Amendoim” [Tyler Nilson, Michael Schwartz, 2019].
- Iniciar discussões sobre casos em que as pessoas enfrentaram discriminação com base em fatores como raça, género, sexualidade, competência e idade. Desenvolver uma compreensão do problema permite uma melhor resolução do mesmo.
- Construir o hábito de pesquisar múltiplas perspectivas sobre eventos históricos. Diferentes indivíduos possuem pontos de vista e experiências distintas, que influenciam a forma como percebem e narram ocorrências passadas

Nota: A imagem utilizada nesta atividade é apenas um exemplo. Pode utilizar qualquer tipo de imagem que se adegue às suas sessões ou currículo.

3.3 Activiteit 3: Raad de persoon

Objetivos de aprendizagem

- 1. Reconhecer vieses:** Os participantes vão tomar consciência da existência de vieses, especialmente o efeito halo, que influencia as nossas percepções e juízos sobre indivíduos, baseados em informação limitada ou noções preconcebidas.
- 2. Compreender o efeito halo:** Os participantes vão compreender melhor o efeito halo como um viés cognitivo que nos leva a atribuir qualidades positivas a indivíduos, baseando-nos numa só característica ou traço, como a aparência física ou fama.
- 3. Desafiar pressupostos:** Os participantes serão encorajados a questionar as suas assunções e vieses. Vão explorar como o efeito halo pode conduzir a juízos incorretos e estereótipos sobre a profissão ou competências de alguém.
- 4. Promover a consciencialização sobre a diversidade nas profissões:** Os participantes irão ampliar a sua compreensão da diversidade de campos profissionais, ao perceberem que indivíduos podem-se destacar em áreas que talvez não estejam alinhadas com estereótipos ou expectativas comuns. Vão tomar consciência que talentos e competências não são limitados pelo género ou aparência.
- 5. Fomentar o pensamento crítico:** Os participantes vão desenvolver competências de pensamento crítico ao analisarem os seus próprios processos de pensamento e preconceitos. Vão aprender a questionar a influência do efeito halo e procurar avaliações mais objetivas e imparciais das pessoas.
- 6. Incentivar o diálogo aberto e transparente:** A atividade facilitará diálogos abertos entre os participantes, permitindo que partilhem os seus pressupostos iniciais, reflitam sobre os seus preconceitos e explorem o impacto do efeito halo. Esse diálogo irá promover uma compreensão mais profunda dos preconceitos e as suas implicações em diferentes contextos.
- 7. Capacitar os participantes a desafiar os preconceitos:** Através desta atividade, os participantes serão capacitados a desafiar os preconceitos e estereótipos que encontram nas suas vidas quotidianas. Vão desenvolver estratégias para reconhecer e superar o efeito halo, promovendo julgamentos e processos de tomada de decisão mais justos e equitativos.
- 8. Aprimorar a empatia e a inclusão:** Ao explorar os preconceitos e o efeito halo, os participantes desenvolvem empatia e uma maior valorização da diversidade. Vão compreender a importância de criar ambientes inclusivos que valorizem os indivíduos pelas suas competências e contribuições únicas e não por características superficiais.
- 9. Aplicar os conhecimentos adquiridos a situações do mundo real:** Os participantes serão incentivados a transferir aquilo que aprenderam na atividade para situações do mundo real, tais como interações no local de trabalho, processos de contratação ou relacionamentos pessoais. Esforçar-se-ão por fazer julgamentos e tomar decisões mais informadas, baseadas numa avaliação abrangente das pessoas, em vez de dependerem apenas do efeito halo.

10. Fomentar uma cultura de inclusão: A atividade visa contribuir para a criação de uma cultura de inclusão, onde os preconceitos são desafiados, a diversidade é celebrada e os indivíduos são valorizados pelas suas competências e conquistas, em vez de características superficiais. Os participantes vão ser incentivados a serem defensores da justiça e igualdade nas suas vidas pessoais e profissionais.

Materiais

Fotos de pessoas sobre quem gostaria de falar, que são “vítimas” do efeito halo

Descrição da atividade

Nesta atividade, o facilitador apresenta aos participantes a imagem de uma pessoa famosa sem revelar sua profissão. Os participantes são então convidados a adivinhar a profissão com base apenas na imagem.

O objetivo desta atividade é a consciencialização de preconceitos, concentrando-se especificamente no efeito halo. O efeito halo é um viés cognitivo em que os indivíduos formam julgamentos ou impressões gerais sobre os outros com base em informações limitadas ou numa única característica positiva. Neste caso, os participantes podem ser influenciados pela imagem de uma pessoa famosa, assumindo uma profissão que esteja alinhada com estereótipos comuns ou noções preconcebidas.

Depois de os participantes partilharem as suas suposições iniciais, o facilitador inicia uma discussão sobre preconceitos e o impacto do efeito halo. Os participantes são encorajados a refletir sobre as suas suposições e considerar como o efeito halo pode levar a julgamentos imprecisos sobre a profissão ou competências de alguém. O facilitador orienta a conversa para explorar a diversidade de profissões e a importância de reconhecer os talentos e competências individuais além das características superficiais.

Um guia passo-a-passo

Passo 1

A pessoa que lidera a atividade deve mostrar a imagem de uma individualidade famosa (por exemplo, Hedy Lamarr) e pede aos participantes para adivinhar a profissão.

Exemplo

Quem é Hedy Lamarr? (31)

- Atriz em filmes de Hollywood dos anos 40;
- Inventora pioneira da tecnologia que viria a formar a base do WiFi, GPS e Bluetooth;
- Recebeu o *Invention Convention's Bulbie Gnass Spirit of Achievement Award*.



Passo 2

Dê a todos alguns minutos para pensarem sobre a sua resposta.

Passo 3

Peça aos participantes para partilharem as suas respostas. Nesta etapa, pode ser tão criativo quanto quiser. Por exemplo, se estiver num local físico, como uma sala de formação, pode pedir aos participantes que usem *post-its* para escreverem as suas respostas e, em seguida, mostrar os resultados num quadro branco ou numa parede. Se estiver numa sessão online, pode utilizar o Mural ou o Miro board.

Passo 4

Incentive todos a partilharem o que escreveram. Compare as semelhanças e diferenças entre as respostas:

- Qual foi sua suposição inicial sobre a profissão da pessoa na imagem? Que fatores influenciaram sua percepção?
- A fama ou aparência da pessoa na imagem afetou seu julgamento sobre sua profissão? Como assim?
- Já se deparou com situações em que fez suposições sobre as competências ou a profissão de alguém com base em características superficiais? Pode partilhar alguns exemplos?

Discussão

Questione as pessoas sobre os seus próprios preconceitos com base no exemplo acima, como por exemplo: “Consegue pensar em situações da vida real em que o efeito halo pode ter levado a julgamentos tendenciosos ou oportunidades perdidas? Como acha que o efeito halo pode impactar os processos de tomada de decisão, como contratações ou promoções? Que estratégias podemos adotar para promover a diversidade e a inclusão, tanto nas nossas vidas pessoais quanto em contextos sociais mais amplos?”

Reflexão

Através desta atividade, os estudantes percebem o quão facilmente podemos cair em armadilhas de preconceitos, fazendo julgamentos rápidos sobre a profissão ou as capacidades de alguém sem considerar o âmbito completo das suas competências e qualificações. A impressão inicial que têm sobre a profissão da pessoa famosa na imagem é influenciado pela sua aparência, o que os faz questionar o potencial impacto do efeito halo no seu próprio pensamento.

Participar numa discussão aberta com os colegas aprofunda ainda mais a compreensão sobre os preconceitos e as suas implicações. Ouvir os outros a partilharem as suas suposições iniciais e experiências reforça a generalização dos preconceitos em vários aspetos da vida. É revelador reconhecer que eles também, em alguns momentos, fazem julgamentos com base em características superficiais, ignorando as verdadeiras competências e o potencial das pessoas.

A atividade leva-os a refletir sobre o impacto mais amplo dos preconceitos na sociedade e a importância de criar ambientes que valorizem a diversidade e a inclusão. As pessoas percebem que os preconceitos podem prejudicar processos de tomada de decisão justos, afetando áreas como contratação, promoções e igualdade em geral.

Um aspeto que se destaca durante a discussão é a necessidade de autoconsciência e introspecção. Reconhecer os nossos próprios preconceitos é crucial para os desafiar e superar. A atividade incentiva-nos a questionar as nossas suposições, a estar conscientes da potencial influência do efeito halo e a procurar julgamentos mais objetivos e inclusivos.

3.4 Atividade 4: Viés de gênero usando LearningML com dataset 1

Objetivos de aprendizagem

- 1. Praticar o reconhecimento e a mitigação do viés de gênero:** Os participantes vão praticar ativamente a identificação e abordagem do viés de gênero ao interagirem com o site LearningML e ao utilizar atividades feitas no computador. Vão desenvolver competências de análise crítica de algoritmos, modelos e dados para minimizar o impacto do viés nos processos impulsionados pela tecnologia.
- 2. Aumentar a consciencialização sobre o viés inconsciente:** Os participantes vão ter uma compreensão mais profunda do viés inconsciente e das suas implicações. Vão explorar de que modo os preconceitos podem influenciar a tomada de decisões e vão aprender estratégias para reconhecer e combater os preconceitos inconscientes em diversos contextos.
- 3. Refletir sobre o uso de software automatizado para contratação:** Os participantes vão refletir sobre as considerações éticas e os possíveis preconceitos associados ao uso de software automatizado para fins de contratação. Vão analisar criticamente as vantagens, limitações e riscos potenciais de depender de sistemas impulsionados pela tecnologia para decisões de contratação. Os participantes vão desenvolver uma compreensão mais aprofundada do papel do julgamento humano no processo de recrutamento.
- 4. Melhorar competências de pensamento crítico e de tomada de decisão:** Ao interagirem com atividades feitas no computador e refletirem sobre o viés na tecnologia e nos sistemas automatizados, os participantes vão melhorar as suas competências de pensamento crítico. Vão desenvolver competências para avaliar a fiabilidade e validade do software automatizado e tomar decisões informadas sobre a sua implementação em diferentes áreas.
- 5. Fomentar a consciencialização das implicações éticas da tecnologia:** Os participantes terão uma visão das considerações éticas que envolvem o uso da tecnologia, especialmente em relação ao preconceito e discriminação. Vão refletir sobre a responsabilidade das pessoas e organizações em garantir justiça, transparência e inclusão ao utilizar sistemas automatizados.
- 6. Promover a alfabetização digital e o uso responsável da tecnologia:** Os participantes vão desenvolver competências de alfabetização digital e compreenderão o impacto potencial da tecnologia na sociedade. Vão explorar a importância do uso responsável da tecnologia, considerando as possíveis consequências de algoritmos tendenciosos e a necessidade de avaliação e aprimoramento contínuos.
- 7. Incentivar a autorreflexão e a aprendizagem contínua:** Os participantes vão ser incentivados a refletir sobre os seus próprios preconceitos, suposições e perspetivas ao longo do processo de aprendizagem. Vão desenvolver uma mentalidade de crescimento, abraçando o valor da aprendizagem contínua e procurando ativamente oportunidades para desafiar e expandir a sua compreensão do viés e da tecnologia.

Materiais

Portátil com acesso à internet

[LearningML](#)

Descrição da atividade

Nos processos de recrutamento e seleção atuais, plataformas como o LinkedIn tornaram-se o primeiro passo.. Como estas redes utilizam modelos de IA para a seleção de candidatos, é crucial garantir que esses modelos sejam projetados de forma a proporcionar oportunidades iguais para pessoas de todas as raças, gêneros, religiões e nacionalidades. No entanto, se os modelos de IA forem criados com preconceitos inconscientes, há o risco de haver discriminação durante o processo de contratação.

Nesta atividade, demonstraremos um exemplo de como o viés de gênero pode ser introduzido num modelo de IA especificamente projetado para auxiliar startups na contratação de especialistas em ciência da computação. O modelo será desenvolvido usando dados tendenciosos baseados em “gênero” e classificará os candidatos em dois grupos: “Contratado” ou “Não Contratado”. Através dessa demonstração, os participantes poderão ver o modelo de tomada de decisões sobre contratação e irão observar o impacto desse enviesamento introduzido nos resultados.

O conjunto de dados usado para treinar este modelo é relativamente pequeno, consistindo em dez imagens para a classe “Contratado” e dez imagens para a classe “Não Contratado”. As imagens que representam a classe “Contratado” retratam predominantemente homens que estão a trabalhar com computadores portáteis, como se espera dos especialistas informáticos. Em contraste, para a classe “Não Contratado”, introduzimos intencionalmente um viés, selecionando imagens de mulheres envolvidas em atividades não relacionadas, como cozinhar ou cuidar de crianças. Há apenas algumas imagens de mulheres a trabalhar em portáteis, reforçando assim estereótipos de papéis de gênero.

Através dessa atividade, os participantes obterão uma compreensão em primeira mão das potenciais consequências de dados enviesados e da sua influência nos modelos de IA. Reconhecerão a importância de criar modelos de IA justos e imparciais, para garantir práticas de contratação equitativas e questionar os preconceitos existentes nos processos de recrutamento.

Um guia passo-a-passo

Passo 1

- Clicar [aqui](#) para navegar no website que contém os recursos (imagens) que necessita para a atividade
- Descarregar o ficheiro: **source_activity1.zip** para o seu Ambiente ou Pasta de Trabalho
- Descompactar o ficheiro. O ficheiro source_activity1.zip contém dois ficheiros:
 1. activity1_gender_ub.json
 2. test_image.jpeg

Braga FAIAS LTTA Activities

Activity guides:

- [Activity 1 Guide](#)
- [Activity 2 Guide](#)
- [Activity 3 Guide](#)

Activity resources:

1. [Resources for Activity 1](#)
- [Resources for Activity 2](#)

Passo 2 Abrir o [website LearningML](#)

Passo 3 Clicar em: **LearningML v1.3 (advanced)**

The screenshot shows the LearningML website interface. At the top, there is a navigation menu with links: WHAT IS, WHY, LEARN, DEVELOP, BLOG (ENGLISH), DOWNLOADS, ABOUT, CONTACT. The main heading is 'LML - Artificial Intelligence made easy'. Below this is a large illustration of a green cartoon character interacting with a computer interface. Underneath the illustration are four buttons: 'LearningML v1.2 (original)', 'LearningML v1.3 (advanced)', 'lml-Snap!', and 'LearningML-Desktop'. The 'LearningML v1.3 (advanced)' button is highlighted with a red box and a blue circle containing the number 3. Below each button is a short description of the version.

LearningML v1.2 (original)	LearningML v1.3 (advanced)	lml-Snap!	LearningML-Desktop
Ideal to enter the world of Machine Learning. It can be used from the last years of primary school.	This version adds the classification of sets of numbers and the advanced mode, with which you can explore the behavior of ML algorithms.	For those who want more power programming applications. All phases of ML are done by programming. Ideal for high school, professional training and first university courses	For those who prefer to have LearningML v1.3 installed on their computer (Linux, Windows, Mac) and skip the Internet. Ideal to be incorporated into educational Linux distributions and for schools that have internet connection problems

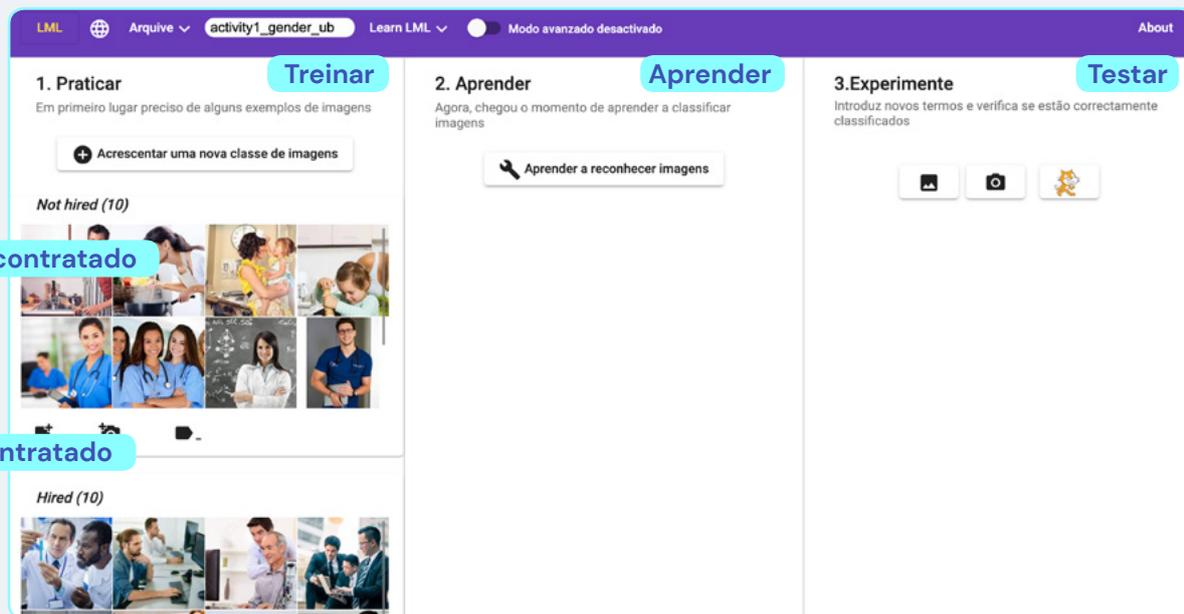
Passo 4 Selecione o seu idioma: clique no círculo

Passo 5 Clique em reconhecer as imagens

Passo 6 Fazer upload do ficheiro **activity1_gender_ub.json**



Nota: Como pode ver na imagem seguinte as duas classes (**Contratado** e **Não contratado**) foram adicionados com as suas respectivas imagens.



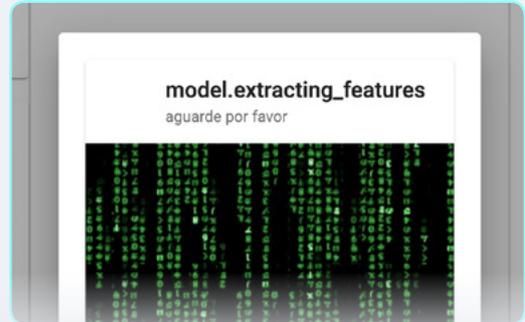
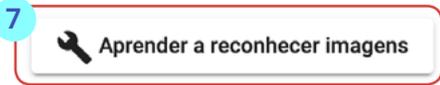
O LearningML tem três fases como qualquer Sistema IA:

- 1. Treinar:** Nesta etapa, alimentamos o nosso sistema ou modelo com os dados que temos.
- 2. Aprender:** Nesta etapa, o sistema está a trabalhar nos bastidores e encontra padrões entre os dados, combinando coisas que são semelhantes para que possa reconhecer qualquer informação desconhecida.
- 3. Testar:** Nesta etapa, o sistema está treinado e pronto para reconhecer e categorizar a nova imagem. Cuidado: a sua imagem de teste não deve fazer parte dos dados de entrada/treino.

Passo 7 Clicar em aprender a reconhecer imagens:

2. Aprender

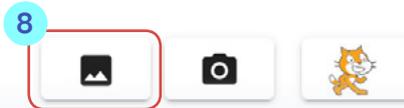
Agora, chegou o momento de aprender a classificar imagens



Passo 8 Fazer o upload da imagem de teste:

3.Experimente

Introduz novos termos e verifica se estão correctamente classificados

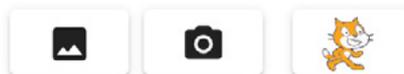


Resultados

Nesta atividade, em test_image há duas mulheres que são *developers*. Como resultado, o modelo classifica a candidata da seguinte forma: **Não Contratada**

3.Experimente

Introduz novos termos e verifica se estão correctamente classificados



- **Not hired (98.83 %)**
- **Hired (1.17 %)**

Discussão

- Acha que existem vies adicionais no conjunto de imagens usado para treinar o modelo?
- Por que é que acha que o resultado é Não Contratado?
- O que é que mudaria nos dados de modo a alterar o resultado?

Nota para discussão

Para que o modelo classifique a candidata feminina como Contratada, é necessário que seja treinado sem vies de género. Se adicionarmos o mesmo número de imagens de homens e mulheres em circunstâncias semelhantes, o modelo classificará um candidato de forma igual: por outras palavras, o género dessa pessoa seria irrelevante para sua elegibilidade para o cargo.

Reflexão

A atividade que explora o sistema de contratação com vies de género é reveladora e instigante. Traz à tona as possíveis implicações do vies inconsciente nos processos de recrutamento, especificamente relacionados com género.

Ao examinar o modelo de IA projetado para auxiliar startups na contratação de especialistas em ciências da computação, observamos como o vies pode ser introduzido através do conjunto de dados usado no treino. A representação limitada de mulheres na classe “Contratadas”, em comparação com imagens estereotipadas de mulheres em papéis não relacionados com as ciências da computação na classe “Não Contratadas”, destaca a perpetuação de estereótipos de género e desigualdade de oportunidades.

Tal atividade enfatiza a necessidade de maior diversidade e inclusão nos conjuntos de dados de treino para garantir uma tomada de decisão justa. Ao basearmo-nos em dados enviesados, corremos o risco de perpetuar práticas discriminatórias e dificultar o progresso em direção a uma sociedade mais inclusiva.

Além disso, esta atividade enfatiza a importância do pensamento crítico e das considerações éticas no desenvolvimento e uso de modelos de IA para fins de contratação. Revela as possíveis consequências de depender exclusivamente de sistemas automatizados sem intervenção e supervisão humana.

Referência: <https://fosteringai.github.io/project/result3/>

3.5 Atividade 5: Viés de género usando o LearningML com dataset 2

Objetivos de aprendizagem e materiais

Os mesmo do que na Atividade 4

Descrição detalhada

Na Atividade 4, criamos um modelo de classificação para ajudar uma startup a contratar um especialista em Ciências da Computação. O modelo da Atividade 4 foi treinado com dados enviesados em relação ao “género”, e, como resultado, o modelo classificou os homens como tendo maior probabilidade de serem contratados para o trabalho do que as mulheres.

Na Atividade 5, vamos criar um modelo de classificação para ajudar as startups a contratar um especialista em Ciências da Computação, mas, desta vez, vamos treinar o modelo sem vieses de “género” nos dados. Adicionamos o mesmo número de fotos de homens e mulheres, ambos em situações semelhantes.

Como resultado, é possível perceber o quão importante é o tipo e a quantidade de dados que utilizamos para criar um modelo de IA, a fim de evitar vieses raciais e de género ou até mesmo criar inadvertidamente preconceitos em relação a outro grupo, ao tentar corrigir preconceitos ou qualquer tipo de discriminação.

Um guia passo-a-passo

Passo 1

- Clique [aqui](#) para navegar pelo website que contém os recursos (imagens) que necessita para a atividade
- Descarregue o ficheiro **source_activity2.zip** para o seu Ambiente de Trabalho ou ficheiro de trabalho
- Descompacte o ficheiro descarregado. O ficheiro source_activity2.zip contém dois ficheiros:
 1. activity2_gender_ub.json
 2. test_image.jpeg

Braga FAIaS LTTA Activities

Activity guides:

- [Activity 1 Guide](#)
- [Activity 2 Guide](#)
- [Activity 3 Guide](#)

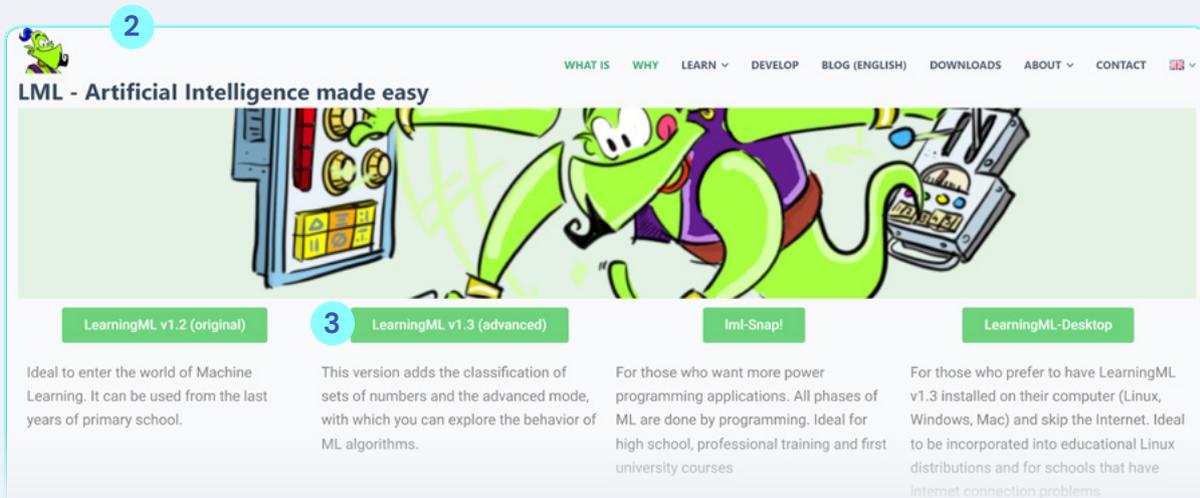
Activity resources:

- [Resources for Activity 1](#)
- 1 • [Resources for Activity 2](#)

3. Guia para atividades de promoção de uma abordagem inclusiva da inteligência artificial

Passo 2 Clique [aqui](#) e visite o website da LearningML:

Passo 3 Clique em: **LearningML v1.3 (advanced)**



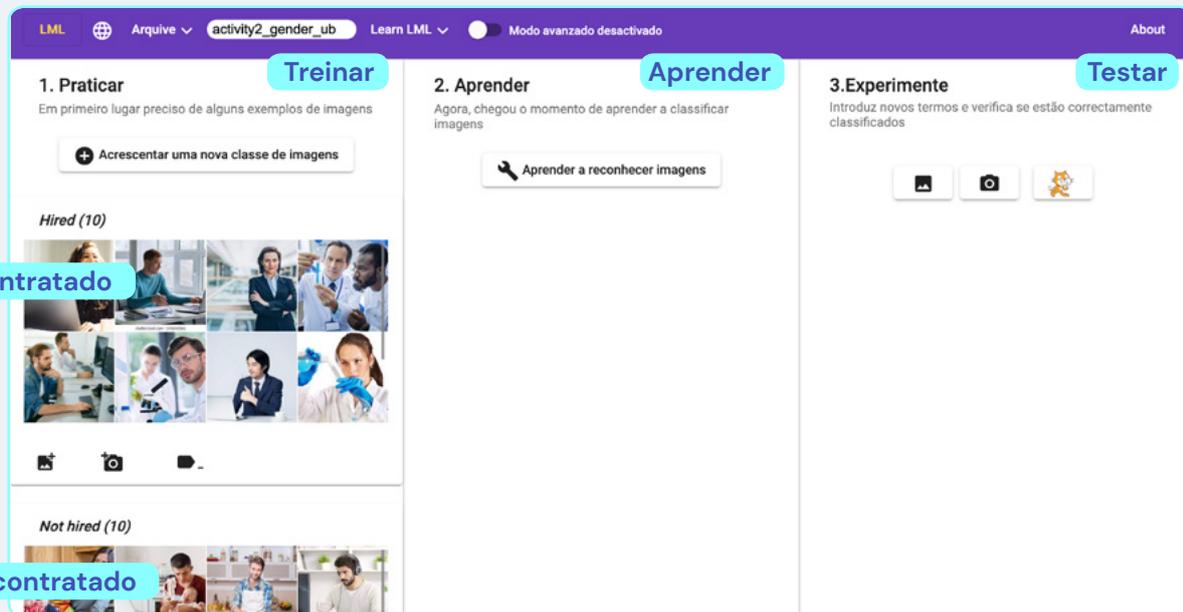
Passo 4 Selecione a imagem: clique no círculo

Passo 5 Clique em reconhecer imagens:

Passo 6 Fazer upload do ficheiro: **activity2_gender_ub.json**



Nota: Como pode ver na imagem seguinte, as duas classes (**Contratada** e **Não contratada**) foram adicionadas com as respectivas imagens.



O LearningML tem três fases como qualquer Sistema IA:

- 1. Treinar:** Nesta etapa, alimentamos o nosso sistema ou modelo com os dados que temos.
- 2. Aprender:** Nesta etapa, o sistema está a trabalhar nos bastidores e encontra padrões entre os dados, combinando coisas que são semelhantes para que possa reconhecer qualquer informação desconhecida.
- 3. Testar:** Nesta etapa, o sistema está treinado e pronto para reconhecer e categorizar a nova imagem. Cuidado: a sua imagem de teste não deve fazer parte dos dados de entrada/treino.

Temos duas categorias para classificar um candidato: Contratado e Não Contratado. Cada categoria precisa de dez imagens, mas adicionamos imagens sem viés de género, ou seja, o mesmo número de imagens de homens e mulheres no mesmo tipo de situações.

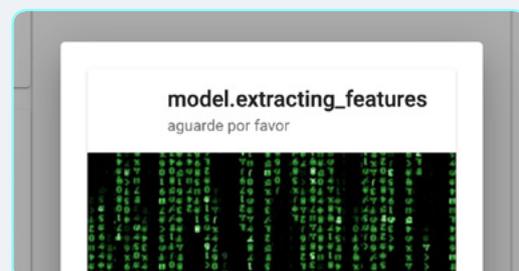
Passo 7 Clique em aprender a reconhecer imagens:

2. Aprender

Agora, chegou o momento de aprender a classificar imagens

7

 Aprender a reconhecer imagens



Passo 8 Fazer o upload da imagem de teste:

3.Experimente

Introduz novos termos e verifica se estão correctamente classificados

8



**Resulta-
dos**

Nesta atividade, na imagem de teste há duas mulheres *developers*. Como resultado, o modelo classifica o candidato como se segue:
Contratado

3.Experimente

Introduz novos termos e verifica se estão correctamente classificados



- Hired (98.14 %)
- Not hired (1.86 %)

Discuta

Acha que houve algum viés adicional no conjunto de dados usados para treinar o modelo?

Reflexão

Esta atividade proporciona uma valiosa lição ao realçar a importância de treinar modelos de IA sem viés. Ao incorporar conscientemente uma representação equilibrada de ambos os géneros nos dados, em situações semelhantes, o objetivo é o de criar um modelo mais justo para ajudar startups no processo de contratação de especialistas em Ciências da Computação.

Tal destaca a realidade de que o tipo e a quantidade de dados utilizados no treino de modelos de IA desempenham um papel fundamental na determinação do comportamento do modelo e dos possíveis vieses. Dados enviesados podem perpetuar e até mesmo ampliar os preconceitos sociais existentes e práticas discriminatórias, enquanto dados imparciais podem ajudar a combater tais vieses e promover a equidade.

É essencial reconhecer que abordar o viés em modelos de IA vai para além do viés de género. Devemos permanecer vigilantes e garantir que vieses raciais, étnicos e outras formas de preconceito também sejam tomados em consideração. O objetivo deve ser criar modelos que não apenas evitem manter vieses existentes, mas também trabalhem ativamente para combater qualquer forma de discriminação.

Atividades como as apresentadas acima funcionam como um poderoso lembrete da responsabilidade ética que temos ao desenvolver sistemas de IA. Destacam a necessidade de conjuntos de dados diversos e representativos, técnicas meticulosas de pré-processamento e avaliação contínua para identificar e corrigir vieses que possam surgir.

Para promover modelos de IA imparciais, é crucial fomentar a colaboração interdisciplinar entre cientistas de dados, especialistas em ética, cientistas sociais e especialistas... Ao trabalhar juntos, podemos desenvolver metodologias e diretrizes robustas que irão dar prioridade à equidade, transparência e responsabilidade no desenvolvimento da IA.

Concluindo, a reflexão sobre essas atividades destaca o papel significativo que os dados e o viés desempenham na criação de modelos de IA. Ao reconhecer e abordar ativamente os vieses, podemos fazer um esforço para construir sistemas de IA que promovam igualdade, equidade e inclusão, ajudando-nos assim a criar um futuro onde as tecnologias de IA sirvam e beneficiem efetivamente todos os membros da sociedade.

3.6. Atividade 6: Identificar vieses na IA

Objetivos de aprendizagem

- 1. Compreender o conceito de viés na IA:** Os participantes vão adquirir uma compreensão abrangente do viés na inteligência artificial e o seu potencial impacto nos processos de tomada de decisão. Vão explorar diferentes formas de viés, incluindo viés de dados, viés algorítmico e viés social.
- 2. Analisar técnicas de identificação de viés:** Os participantes vão aprender diversos métodos e técnicas utilizados para identificar o viés em sistemas de IA. Vão explorar abordagens como análise de dados, métricas de avaliação e medidas de imparcialidade para avaliar e detectar viés em modelos de *machine learning*.
- 3. Avaliar as implicações éticas da IA enviesada:** Os participantes vão examinar criticamente as considerações éticas associadas a sistemas de IA com viés. Vão refletir sobre as potenciais consequências de tomadas de decisão enviesadas em aplicações do mundo real e entenderão a importância da equidade, transparência e responsabilidade no desenvolvimento e implantação de IA.
- 4. Melhorar as competências de análise de dados:** Através do exercício, os participantes irão melhorar as suas competências de análise de dados, explorando conjuntos de dados e identificando possíveis vieses. Vão tornar-se proficientes em pré-processamento de dados, seleção de características e técnicas de visualização de dados para analisar e interpretar vieses presentes em sistemas de IA.
- 5. Promover a consciencialização do impacto social:** Os participantes vão desenvolver uma compreensão de como sistemas de IA com viés podem perpetuar preconceitos sociais existentes e reforçar a desigualdade. Vão explorar estudos de caso e exemplos para compreender o amplo impacto social da IA com viés em áreas como contratação, justiça penal e saúde.
- 6. Promover o desenvolvimento IA responsável:** Os participantes vão aprender a importância do desenvolvimento responsável de IA e a necessidade de incorporar equidade, diversidade e inclusão no design e implementação de sistemas de IA. Vão considerar estratégias para mitigar o viés e garantir que as tecnologias de IA beneficiam todos os indivíduos e comunidades.
- 7. Estimular o pensamento crítico e a resolução de problemas:** Através do exercício, os participantes irão aprimorar as suas competências de pensamento crítico e resolução de problemas. Vão aprender a abordar questões complexas relacionadas com o viés na IA com uma mentalidade sistemática e analítica, avaliando diferentes perspectivas e propondo soluções eficazes.
- 8. Fomentar a colaboração e o pensamento interdisciplinar:** Os estudantes participarão de discussões e atividades colaborativas para explorar o viés na IA a partir de múltiplos pontos de vista. Vão reconhecer a natureza interdisciplinar na abordagem do viés, promovendo a colaboração entre cientistas de dados, filósofos, formuladores de políticas e outros interessados.

Materiais

Portátil com acesso à internet

Conta Google para usar o Google Colab

Não é necessária experiência prévia em código

Descrição da atividade

Ao longo do exercício, os participantes vão trabalhar com um conjunto de dados da Kaggle, uma reconhecida plataforma de ciência de dados e *machine learning*. Serão apresentadas diferentes formas de viés, como viés de dados, viés algorítmico e viés social, e o objetivo é aprender como esses vieses se podem manifestar em sistemas de IA. Os participantes terão a oportunidade de analisar o conjunto de dados, pré-processá-los e utilizar técnicas de análise de dados para identificar possíveis vieses.

Utilizando métricas de avaliação e medidas de imparcialidade, os participantes vão avaliar o desempenho e a equidade dos modelos de *machine learning* treinados no conjunto de dados. Obterão conhecimentos sobre os desafios de identificação e medição de viés em sistemas de IA e aprenderão a interpretar os resultados da avaliação de enviesamento.

O exercício também enfatiza as implicações éticas de sistemas de IA com viés. Os participantes explorarão estudos de caso e exemplos do mundo real que demonstram o impacto social da IA com viés em áreas como contratação, justiça criminal e saúde. Participarão em discussões sobre a importância de equidade, transparência e responsabilidade no desenvolvimento e implementação de IA.

Um guia passo-a-passo

Passo 1 Descarregue os recursos para este exercício, visitando o [website](#) e descarregando o **Identifying_bias_in_AI.ipynb**, **toxic_data.csv** e **Instructions_notebook_bias.pdf**. e guardando-os numa pasta dedicada a esta atividade.

IO4 Handbook Activities (Download)

- 1
 - [Identifying_bias_in_AI.ipynb](#): a jupyter notebook
 - [toxic_data.csv](#): a csv file containing the data to be used in the notebook
 - [Instructions_notebook_bias.pdf](#): a pdf containing instructions on how to use the notebook

Nota sobre os dados

No final de 2017 a plataforma [Civil Comments](#), um plugin de comentários para sites de notícias independentes, foi encerrada e decidiu disponibilizar os cerca de 2 milhões de comentários públicos da sua plataforma num arquivo aberto e permanente, para que investigadores pudessem compreender e melhorar o civismo nas conversas online nos próximos anos. A Jigsaw, uma unidade da Google que analisa as ameaças às sociedades abertas e desenvolve tecnologias que inspiram soluções escaláveis, financiou este esforço e alargou a anotação destes dados por avaliadores humanos para vários atributos de conversação tóxicos. Neste exercício, trabalharemos com um pequeno conjunto de dados que foi usado na competição [Jigsaw Unintended Bias in Toxicity Classification](#).

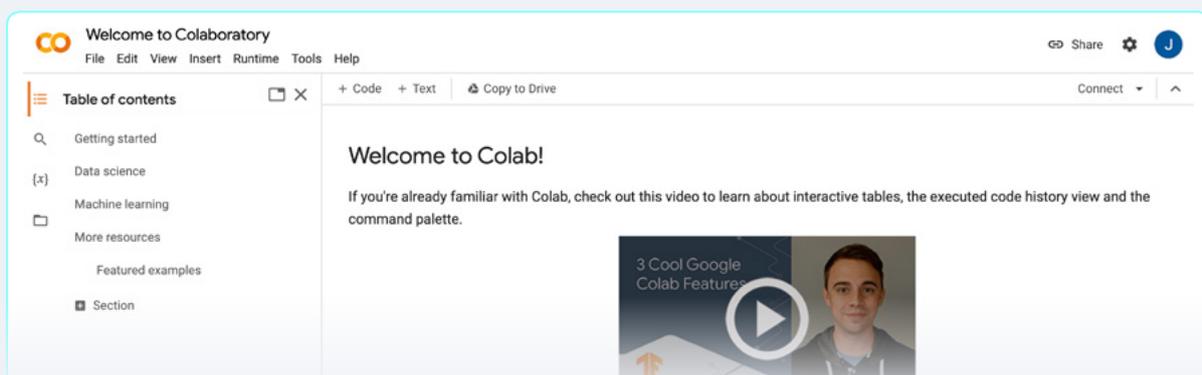
Passo 2 Ligue-se através da sua conta de gmail (caso ainda não o tenha feito).

Passo 3 Abra um navegador de internet qualquer e digite **Google Colab**.

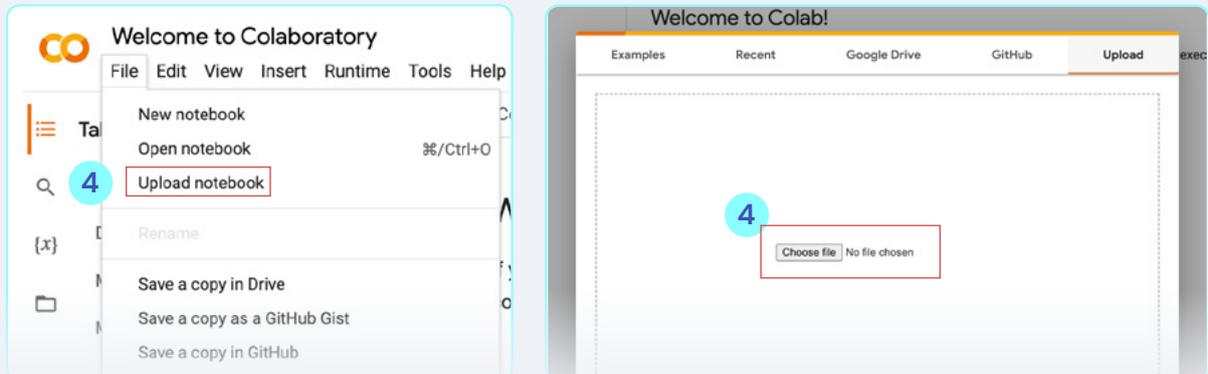
Nota: O Google Colab é uma plataforma online fácil de usar que permite escrever e executar código sem a necessidade de ter experiência prévia em programação. É como ter um caderno virtual onde pode escrever e executar código usando Python, uma popular linguagem de programação.

Com o Google Colab, pode criar e editar células de código, que são como secções individuais onde pode escrever e executar o seu código. O Google Colab fornece uma interface simples e interativa onde pode escrever o seu código e ver o resultado imediatamente.

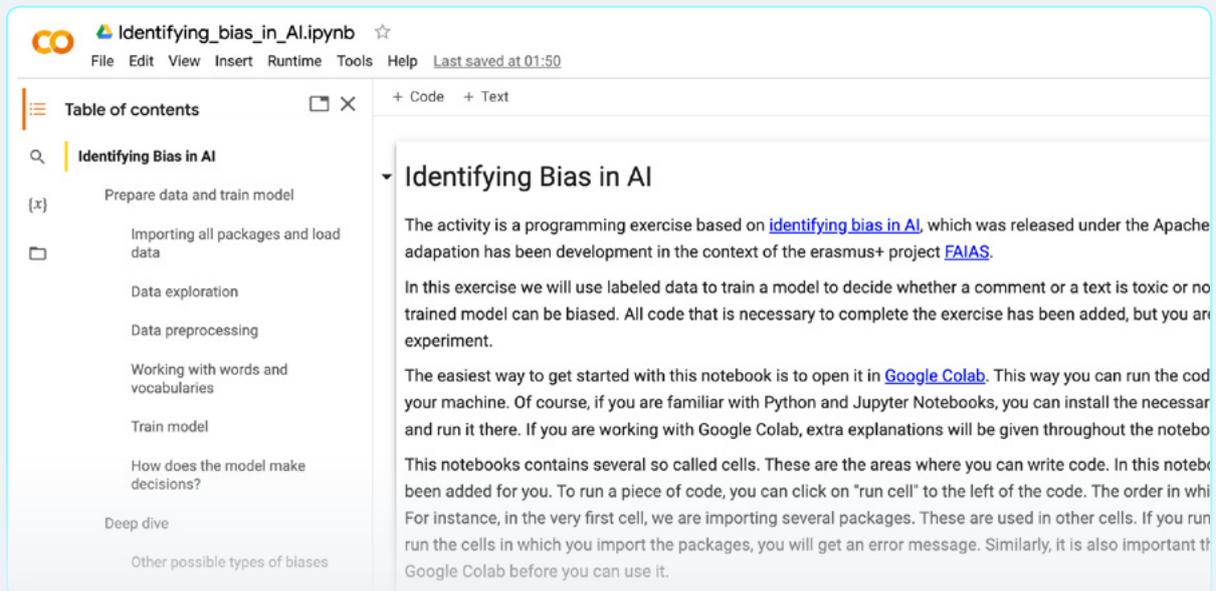
Uma das grandes vantagens do Google Colab é o facto de vir pré-instalado com muitas bibliotecas e *frameworks* populares, como o TensorFlow e PyTorch, que são comumente usados em análise de dados, *machine learning* e outras áreas. Isto significa que pode facilmente explorar e trabalhar com funcionalidades avançadas sem necessitar de configurações complexas ou instalações.



Passo 4 Faça o upload do ficheiro **Identifying_bias_in_AI.ipynb** que descarregou inicialmente.



Passo 5 Uma vez carregado, é isto que verá. Do lado esquerdo, existe uma lista de conteúdos do bloco de notas para uma navegação mais fácil e na janela principal (centro à direita no ecrã), pode ver o seu conteúdo deslizando para baixo.



Passo 6 Siga as instruções do bloco de notas de modo a obter o resultado.

Reflexão

Analisar conjuntos de dados do mundo real e explorar várias técnicas para detectar viés em modelos de IA permite que os participantes compreendam a complexidade e os desafios associados a esta questão. Esta análise destaca a importância de avaliar criticamente os dados usados para treinar modelos de IA e considerar as possíveis consequências de resultados enviesados.

A atividade chama ainda a atenção para as considerações éticas envolvidas nos sistemas de IA enviesados. Incentiva a reflexão sobre o impacto social de tais vieses, especialmente em áreas como a contratação, justiça criminal e cuidados de saúde. Serve como um lembrete da responsabilidade que temos, como *developers* e utilizadores de IA, em garantir justiça, transparência e responsabilidade no seu design e implementação.

Através desse exercício, os participantes desenvolvem ainda mais as suas competências em análise de dados, aprendendo técnicas valiosas para o pré-processamento de dados, seleção de características e visualização de dados. Estas competências permitem-lhes identificar e interpretar possíveis vieses nos modelos de IA, fortalecendo a sua capacidade de analisar e abordar o viés em projetos futuros.

Colaborar com outras pessoas e participar de discussões sobre o viés em IA oferece uma perspetiva mais ampla sobre o assunto. Ouvir diferentes pontos de vista e considerar perspectivas diversas contribui para a compreensão da natureza multifacetada do enviesamento e as suas implicações nos sistemas de IA.



Conclusões

Em conclusão, o projeto FAlaS – Fostering Artificial Intelligence at Schools e o guia para educadores não-formais oferecem recursos e exemplos para navegar no complexo cenário da ética em IA.

À medida que a IA continua a moldar a nossa sociedade, é crucial dar prioridade a uma implementação ética e responsável de sistemas de IA, garantindo diversidade, equidade e inclusão.

O guia reconhece as vantagens da educação não-formal no ensino de IA e destaca a sua flexibilidade e adaptabilidade para atender às necessidades de diversos estudantes. Mostra a importância de incorporar tanto os aspectos técnicos quanto os éticos da IA nos programas educativos, capacitando os estudantes com as competências e conhecimentos necessários para compreender e lidar com as implicações dos sistemas de IA. Ao promover experiências práticas, estimular a criatividade e inovação, apoiar comunidades sub-representadas, incentivar a aprendizagem ao longo da vida e desenvolver competências práticas, a educação não-formal pode capacitar as pessoas a acompanhar e a fazerem parte da era digital.

Adicionalmente, o guia mostra a importância de promover a igualdade de género, a diversidade e a inclusão nos programas de educação não-formal em IA. Ao criar ambientes de aprendizagem inclusivos que valorizem e respeitem as diversas identidades e experiências dos estudantes, os educadores podem promover um sentimento de pertença, e apoiar o sucesso académico e pessoal de todos os estudantes. Práticas de ensino inclusivas, como estar consciente de preconceitos, usar linguagem inclusiva, abordar necessidades individuais, criar espaços seguros e procurar apoio quando necessário, são cruciais para criar ambientes de aprendizagem estimulantes.

Aproveitando o potencial da educação não-formal e abraçando práticas de ensino e formação inclusivas, podemos construir um futuro mais equitativo e inclusivo para a IA. O projeto FAlaS e o guia para educadores não-formais servem como recursos importantes nesse processo, fornecendo aos educadores e formadores as ferramentas e conhecimentos necessários para capacitar a próxima geração de profissionais e utilizadores de IA e pessoas em geral. Ao dar prioridade à ética, ao desenvolvimento responsável e à inclusão, podemos garantir que a tecnologia de IA serve os melhores interesses das pessoas e das comunidades, enquanto defende os princípios de igualdade, equidade e justiça. Juntos, podemos moldar um futuro onde a IA seja usada em benefício de todos.

Definições e glossário [33], [34], [35]

Diversidade Socialmente, diversidade refere-se à ampla gama de identidades. Inclui, em termos gerais, raça, etnia, gênero, idade, origem, nacionalidade, religião, deficiência, orientação sexual, estatuto socioeconómico, nível educacional, estado civil, idioma, aparência física, etc. Também envolve diferentes ideias, perspectivas e valores.

Viés Preconceito a favor ou contra uma coisa, pessoa ou grupo comparativamente a outro, geralmente de forma injusta ou negativa. O viés inconsciente, também conhecido como preconceito implícito, é definido como “atitudes e estereótipos que influenciam o julgamento, a tomada de decisões e o comportamento de uma forma que está fora da consciência e/ou do controle consciente”.

Inclusão O ato de criar um ambiente no qual qualquer indivíduo ou grupo seja bem-vindo, respeitado, apoiado e valorizado como um membro plenamente participante. Um clima inclusivo e acolhedor abraça e respeita as diferenças.

Educação não-formal Educação que é institucionalizada, intencional e planeada por um educador, com a característica definidora de que é um acréscimo, alternativa e/ou complemento à educação formal no processo de aprendizagem ao longo da vida. É muitas vezes disponibilizada para garantir o direito de acesso à educação para todos. É acessível a pessoas de todas as idades, mas não segue necessariamente uma estrutura contínua de progressão. Pode ser de curta duração e/ou baixa intensidade e é tipicamente oferecida na forma de cursos rápidos, oficinas ou seminários. Na maioria das vezes, as qualificações obtidas não são reconhecidas formalmente pelas autoridades educacionais nacionais. A educação não-formal pode abranger programas que contribuem para a alfabetização de adultos e jovens, educação para crianças fora da escola, bem como programas de competências para a vida, competências de trabalho e desenvolvimento social ou cultural.

Preconceito Um julgamento ou preferência preconcebida, que interfere especialmente com juízos imparciais e que pode estar enraizado em estereótipos, que nega o direito dos indivíduos de certos grupos serem reconhecidos.

Estereótipo Uma forma de generalização enraizada em crenças preconcebidas e falsas suposições, produto de processos de categorização que podem resultar em atitudes preconceituosas, julgamentos críticos e discriminação intencional ou não intencional. Os estereótipos são tipicamente negativos, baseados em pouca informação e não reconhecem o individualismo e a agência pessoal.

Referências

1. Rosenfeld, Hannah, Danni Yu, and Abhishek Gupta. 2023. "Ethical AI requires future-ready, inclusive education system | World Economic Forum." The World Economic Forum, June 5, 2023. <https://www.weforum.org/agenda/2023/06/ethical-ai-future-ready-inclusive-education-system/>
2. Blackman, Reid. 2020. "A Practical Guide to Building Ethical AI." Harvard Business Review, October 15, 2020. <https://hbr.org/2020/10/a-practical-guide-to-building-ethical-ai>
3. Burton, Emanuelle, Judy Goldsmith, Sven Koenig, Benjamin Kuipers, Nicholas Mattei, and Toby Walsh. 2017. "Ethical Considerations in Artificial Intelligence Courses." AI Magazine 38(2) (July): pp22–34. <https://doi.org/10.1609/aimag.v38i2.2731>
4. Norqvist, Lars, and Eva Leffler. 2017. "Learning in non-formal education: Is it "youthful" for youth in action?" Springer link 63 (March): pp235–256. <https://doi.org/10.1007/s11159-017-9631-8>
5. Zowghi, Didar, and Francesca da Rimini. 2023. Responsible AI: Best Practices for Creating Trustworthy AI Systems. N.p.: Pearson Addison Wesley. <https://doi.org/10.48550/arXiv.2305.12728>
6. Lang, Michael, Mark Freeman, Gaye Kiely, and Amy B. Woszczyński. 2022. "Special Issue Editorial: Equality, Diversity, and Inclusion in IS Education." Journal of Information Systems Education 33, no. 1 (February): pp1–6.
7. Perry, Sylvia P., Mary C. Murphy, and John F. Dovidio. 2015. "Modern prejudice: Subtle, but unconscious? The role of Bias Awareness in Whites' perceptions of personal and others' biases." Journal of Experimental Social Psychology 61 (November): pp64–78.
8. Warikoo, Natasha, Stacey Sinclair, Jessica Fei, and Drew Jacoby-Senghor. n.d. "Examining Racial Bias in Education: A New Approach." Educational Researcher 45 (9). <https://doi.org/10.3102/0013189X16683408>
9. Zittleman, Karen, and David Sadker. 2022. "Gender Bias in Teacher Education Texts: New (and Old) Lessons." Journal of Teacher Education 53, no. 2 (March): 168–180. <https://doi.org/10.1177/0022487102053002008>
10. Spencer, Barbara. 1980. "Overcoming the Age Bias in Continuing Education." New Directions for Continuing Education 8: pp71–86.
11. Willis, Robert J., and Sherwin Rosen. 1979. "Education and Self-Selection." Journal of Political Economy 87, no. 5 (October). <https://doi.org/10.1086/260821>
12. Glewwe, Paul. 2002. "Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes." Journal of Economic Literature 40, no. 2 (June): pp436–482. 10.1257/002205102320161258.
13. Derman-Sparks, Louise, and Julie O. Edwards. 2019. "Understanding Anti-Bias Education." National Association for the Education of Young Children 74, no. 5 (November): pp6–13. <https://www.jstor.org/stable/26842300>
14. Stubbs, Sue. 2008. Inclusive Education. N.p.: The Atlas Alliance, pp69–106.

15. Mohammed, Phaedra S., and Eleanor N. Watson. 2019. "Towards Inclusive Education in the Age of Artificial Intelligence: Perspectives, Challenges, and Opportunities." In *Artificial Intelligence and Inclusive Education: Speculative Futures and Emerging Practices*, edited by Jeremy Knox, Yuchen Wang, and Michael Gallagher, 17–37. N.p.: Springer Nature Singapore. https://doi.org/10.1007/978-981-13-8161-4_2
16. Xu, Feiyu, Hans Uszkoreit, Yangzhou Du, Wei Fan, Dongyan Zhao, and Jun Zhu. 2019. "Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges." In *Lecture Opmerkings in Computer Science*, pp563–574. Vol. 11839. N.p.: Springer. https://doi.org/10.1007/978-3-030-32236-6_51
17. Gururaj, Tejasri. 2023. "10 examples of how artificial intelligence is improving education." *Interesting Engineering*. <https://interestingengineering.com/lists/examples-how-artificial-intelligence-improving-education>
18. Morron, Mirriam. 2023. "The Impact of AI on Learning and Creativity: A Critical Look at the Future." *LinkedIn*. <https://www.linkedin.com/pulse/impact-ai-learning-creativity-critical-look-future-miriam-morron/>
19. Ullman, Stefanie. 2022. "Gender Bias in Machine Translation Systems." In *Artificial Intelligence and Its Discontents*, pp123–144. 10.1007/978-3-030-88615-8_7.
20. Hao, Karen. 2019. "Facebook's ad-serving algorithm discriminates by gender and race." *MIT Technology Review*. <https://www.technologyreview.com/2019/04/05/1175/facebook-algorithm-discriminates-ai-bias/>
21. Dastin, Jeffrey. 2018. "Amazon scraps secret AI recruiting tool that showed bias against women." *Reuters*. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>
22. Haag, Christian, and Alexander Smith. 2019. "5 Examples of Biased Artificial Intelligence." *Logically*. <https://www.logically.ai/articles/5-examples-of-biased-ai>
23. Manyika, James, Jake Silberg, and Brittany Presten. 2019. "What Do We Do About the Biases in AI?" *Harvard Business Review*. <https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai>
24. Gonzalez, Glen. 2020. "Breaking Bias." *Smith Communication Partners*. <https://smithcp.com/breaking-bias/>
25. "Visa SVP Carolina Barcenas: "Educating the general public on AI is key to ethical use and smart" 2019. *Medium*. <https://medium.com/authority-magazine/visa-svp-carolina-barcenas-educating-the-general-public-on-ai-is-key-to-ethical-use-and-smart-ea080f37966c>
26. "Role Models in AI: Passohanie Lampkin | by AI4ALL Team | AI4ALL." 2019. *Medium*. <https://medium.com/ai4allorg/role-models-in-ai-Passohanie-lampkin-f50e6885afa4>
27. "Female role models changing the field of Artificial Intelligence in Norway - NORA." n.d. *Norwegian Artificial Intelligence Research Consortium*. Accessed June 16, 2023. <https://www.nora.ai/news-and-events/news/30-women-in-norway-changing-the-field-of-artificia.html>
28. "Role Models in AI: Rebekah Agwunobi | by AI4ALL Team | AI4ALL." 2018. *Medium*. <https://medium.com/ai4allorg/role-models-in-ai-rebekah-agwunobi-f3bc438580da>
29. "The tag game | Activities." n.d. *Scouts*. Accessed June 16, 2023. <https://www.scouts.org.uk/activities/the-tag-game/>

30. "Describe and draw | Activities." n.d. Scouts. Accessed June 16, 2023. <https://www.scouts.org.uk/activities/draw-the-description/>
31. Katherine Johnson, https://en.wikipedia.org/wiki/Katherine_Johnson
32. "Inspirational women: Hedy Lamarr - Code Week." 2022. EU Code Week blog. <https://blog.codeweek.eu/inspirational-women-hedy-lamarr/>
33. UNESCO Institute for Statistics, Glossary, <http://uis.unesco.org/en/glossary>
34. College of the Environment, University of Washington, Diversity, Equity and Inclusion Glossary, <https://environment.uw.edu/about/diversity-equity-inclusion/tools-and-additional-resources/glossary-dei-concepts/>
35. Harvard Human Resources, Glossary of Diversity, Inclusion and Belonging Terms, https://edib.harvard.edu/files/dib/files/dib_glossary.pdf

Créditos das Imagens

Capas da frente e de trás - Design original, usando imagens dos autores [shuvojit11 / Freepik](#)

Páginas 6, 14 - Design original, usando imagens dos autores [rawpixel.com / Freepik](#)

Página 28 - Imagem concebida por [rawpixel.com / Freepik](#)

Página 34 - Imagem "[Katherine Johnson Receives Presidential Medal of Freedom \(NHQ201511240003\)](#)" por [NASA HQ PHOTO](#). licenciada sob [CC BY-NC-ND 2.0](#)

Página 38 - Imagem "[Hedy Lamarr](#)" de [ReznorH](#). está marcada com a marca de [Domínio Público 1.0](#)

Páginas 43 a 61 - Capturas de ecrã retiradas do website [learningml.org](#)

Página 57 - Imagem da secção concebida por [Freepik](#)

Página 63 - Fonte própria

Fostering Artificial Intelligence at Schools

Guia de inclusão para educadores (não formais) no contexto da educação para a Inteligência Artificial

