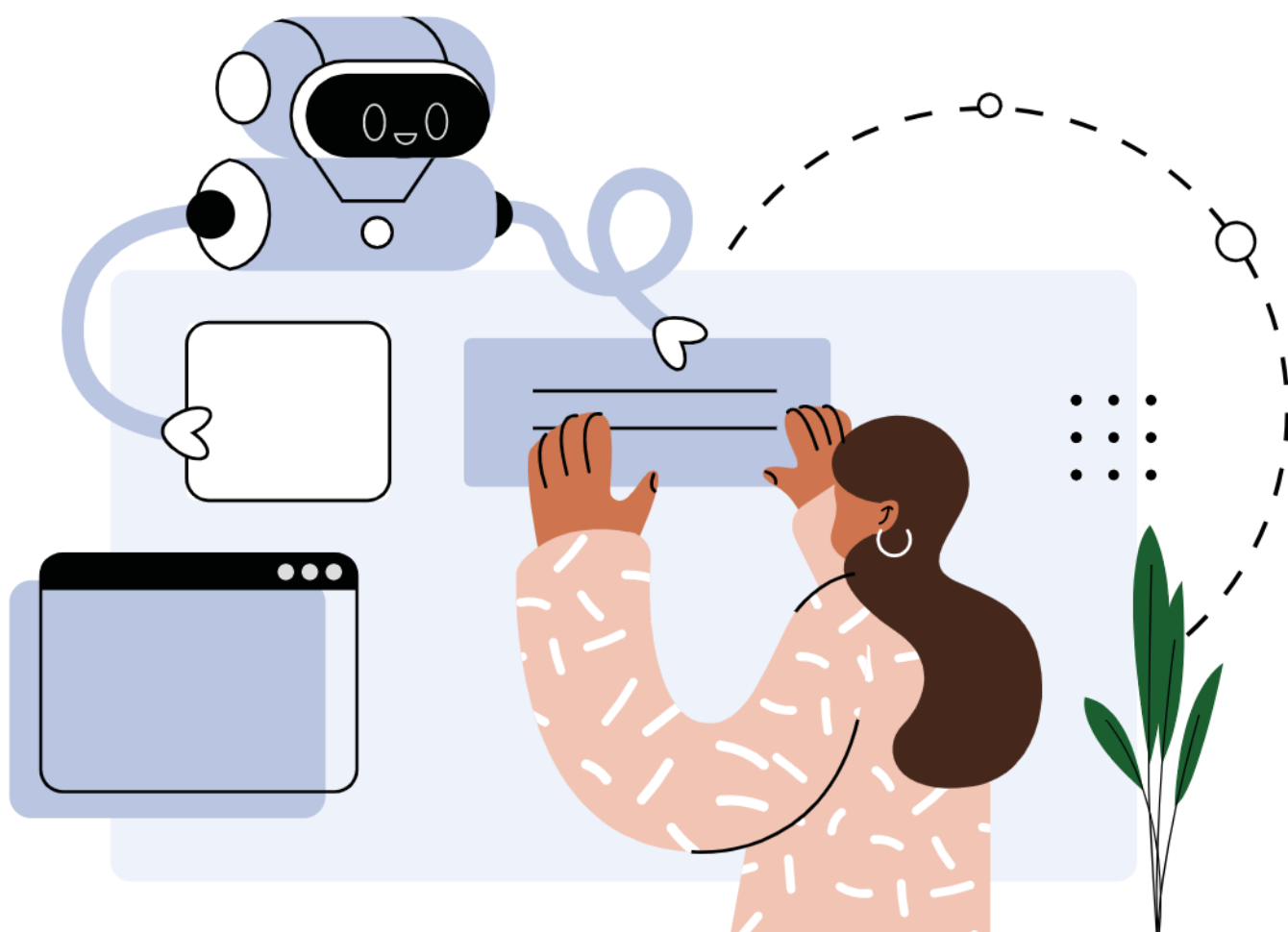


Guia para a IA generativa na educação e na pesquisa

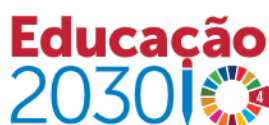


Setor de Educação da UNESCO

A educação é a prioridade máxima da UNESCO, porque se trata de um direito humano básico e é o alicerce para a construção da paz e a impulsão do desenvolvimento sustentável. A UNESCO é a agência especializada das Nações Unidas para a educação, e seu Setor de Educação assume liderança global e regional em educação, fortalece sistemas educacionais nacionais e responde a desafios globais contemporâneos por meio da educação com foco especial na igualdade de gênero e na África.

Agenda Mundial da Educação 2030

A UNESCO, no papel de agência especializada das Nações Unidas para a educação, está encarregada de liderar e coordenar a Agenda 2030 para a Educação, a qual faz parte de um movimento global para erradicar a pobreza por meio de 17 Objetivos de Desenvolvimento Sustentável (ODS) até 2030. A educação, essencial para o cumprimento de todos esses objetivos, tem seu próprio ODS, o de número 4, que visa a “*assegurar a educação inclusiva, equitativa e de qualidade, e promover oportunidades de aprendizagem ao longo da vida para todas e todos*”. O Marco de Ação da Educação 2030 fornece orientações para a implementação desses ambiciosos objetivos e compromissos.



Publicado em 2024 pela Organização das Nações Unidas para a Educação, a Ciência e a Cultura
7, place de Fontenoy, 75352 Paris 07 SP, França

© UNESCO 2024

ISBN: 978-92-3-700028-1



Esta publicação está disponível em acesso livre ao abrigo da licença Attribution-ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) (<http://creativecommons.org/licenses/by-sa/3.0/igo/>). Ao utilizar o conteúdo da presente publicação, os usuários aceitam os termos de uso do Repositório UNESCO de acesso livre (<https://www.unesco.org/pt/open-access/cc-sa>).

As imagens marcadas com um asterisco (*) não se enquadram na licença [CC-BY-SA](#) e não podem ser utilizadas ou reproduzidas sem a prévia autorização dos detentores dos direitos autorais.

Título original: *Guidance for generative AI in education and research*

Publicado em 2023 pela Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO)

As indicações de nomes e a apresentação do material ao longo deste livro não implicam a manifestação de qualquer opinião por parte da UNESCO a respeito da condição jurídica de qualquer país, território, cidade, região ou de suas autoridades, tampouco da delimitação de suas fronteiras ou limites.

As idéias e opiniões expressas nesta publicação são as dos autores e não refletem obrigatoriamente as da UNESCO nem comprometem a Organização.

Tradução:

Teresa Margarida Loureiro Cardoso, Universidade Aberta, LE@D, Rede WEIWER®, Portugal

Viviane Cristina Marques, UNICAMP, Brasil e Rede WEIWER®, LE@D, Universidade Aberta, Portugal

A tradução do Guia para o português contou com a revisão de Luiza Carvalho, do Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação sob os auspícios da UNESCO (Cetic.br), do Núcleo de Informação e Coordenação do Ponto BR (NIC.br) e Maria Luiza Monteiro Bueno e Silva, da Representação da UNESCO no Brasil.

Crédito da capa: Olexandra Simkina/Shutterstock.com*

Publicado e impresso pela UNESCO

Impresso em França

R E S U M O

Orientação para uma abordagem centrada no ser humano para o uso da IA generativa

As ferramentas de IA generativa (IAGen) disponíveis ao público estão surgindo rapidamente, e o lançamento de versões iterativas supera a adaptação das estruturas regulamentares nacionais. A ausência de regulamentações nacionais sobre a IAGen na maioria dos países deixa desprotegida a privacidade dos dados dos usuários, e as instituições de ensino não estão preparadas para validar essas ferramentas.

A primeira orientação global da UNESCO sobre a IAGen na educação tem como objetivo apoiar os países na implementação de ações imediatas, no planejamento de políticas a longo prazo e no desenvolvimento de capacidades humanas para garantir uma visão centrada no ser humano dessas novas tecnologias.

O Guia apresenta uma avaliação dos riscos potenciais que a IAGen pode representar para os valores humanos fundamentais, os quais promovem o agir humano, a inclusão, a equidade, a igualdade de gênero e as diversidades linguísticas e culturais, assim como as opiniões e expressões plurais.

Propõe passos fundamentais para que as autoridades governamentais possam regular a utilização das ferramentas de IAGen, incluindo a obrigatoriedade de proteção da privacidade dos dados e a consideração de um limite de idade para a sua utilização. Descreve os requisitos a serem cumpridos pelos fornecedores de IAGen para o uso ético e eficaz na educação.

O Guia enfatiza a necessidade de as instituições educacionais validarem os sistemas IAGen quanto à sua adequação ética e pedagógica para a educação. Convida a comunidade internacional a refletir sobre as suas implicações a longo prazo para o conhecimento, o ensino, a aprendizagem e a avaliação.

A publicação oferece recomendações concretas aos legisladores e às instituições de ensino sobre como o uso das ferramentas de IAGen pode ser orientado a proteger a agência humana e a beneficiar genuinamente os estudantes, os aprendizes e os pesquisadores.

Enquanto o
ChatGPT alcançou
100 milhões
de utilizadores ativos mensais
em janeiro de 2023, apenas
um país apresentou, em
julho, regulação voltada
à **IA generativa**.



unesco

"Uma vez que as guerras se iniciam nas mentes dos homens, é nas mentes dos homens que devem ser construídas as defesas da paz."

Guia para a IA generativa na educação e na pesquisa

Prefácio



© UNESCO

A Inteligência Artificial Generativa (IAGen) ganhou conhecimento público no final de 2022 com o lançamento do ChatGPT, o qual se tornou o aplicativo com crescimento mais rápido da história. Com o poder de imitar as capacidades humanas para produzir respostas como textos, imagens, vídeos, músicas e códigos de *software*, essas aplicações de IAGen causaram ampla comoção. Milhões de pessoas estão utilizando a IAGen em suas vidas, diariamente, e o potencial de adaptar os modelos para aplicações de domínios específicos com IA parecem ser ilimitadas.

As amplas capacidades para processar informação e produzir conhecimento têm implicações potencialmente enormes para a educação, uma vez que replicam o pensamento de ordem superior que constitui a base da aprendizagem humana. À medida que as ferramentas de IAGen conseguem automatizar cada vez mais alguns níveis básicos de redação e criação artística, elas estão forçando os formuladores de políticas e instituições educacionais a reavaliarem o porquê, o quê e como aprendemos. Essas considerações são cruciais para a educação nesta nova fase da Era Digital.

Essa publicação tem como objetivo apoiar o planejamento de regulamentações e políticas adequadas, além do desenvolvimento de capacidades humanas, garantindo que a IAGen se torne uma ferramenta que verdadeiramente beneficie e capacite professores, estudantes e pesquisadores.

Ela propõe passos-chaves para que as autoridades governamentais regulamentem a utilização da IA generativa. Apresenta, ainda, estruturas e exemplos concretos para a formulação de políticas e *design* instrucional que possibilitem a utilização ética e eficaz dessa tecnologia na educação. Por último, convoca a comunidade internacional a considerar as profundas implicações a longo prazo da IA generativa em como compreendemos o conhecimento e definimos conteúdos, métodos e resultados de aprendizagem e em como avaliamos e validamos a aprendizagem.

Baseando-se na “Recomendação sobre a Ética da Inteligência Artificial” da UNESCO, de 2021, este guia está fundamentado em uma abordagem humanística para a educação que promove a agência humana, inclusão, equidade, igualdade de gênero, diversidade cultural e linguística, assim como opiniões e expressões plurais. Além disso, responde ao relatório de 2021 da Comissão Internacional sobre os Futuros da Educação, “Reimaginar nossos futuros juntos: um novo contrato social para a educação”, que propõe redefinir nossa relação com a tecnologia, como uma parte integrante de nossos esforços para renovar o contrato social para a educação.

A IA não deve usurpar a inteligência humana. Pelo contrário, ela nos convida a reconsiderar as nossas compreensões estabelecidas sobre o conhecimento e a aprendizagem humana. A minha esperança é que este guia nos ajude a redefinir novos horizontes para a educação e embasar nosso pensamento coletivo e ações colaborativas que possam conduzir a futuros de aprendizagem digital centrada no ser humano para todos.

Stefania Giannini,
Diretora-Geral Adjunta de Educação da UNESCO

Agradecimentos

Sob a liderança de Stefania Gianini, Diretora-Adjunta de Educação, e com a orientação de Sobhi Tawil, Diretor da Divisão de Futuro da Aprendizagem e Inovação da Unesco, a elaboração foi conduzida por Fengchun Miao, Chefe da Unidade de Tecnologia e IA na Educação.

Agradecemos especialmente ao professor Wayne Holmes, Professor Associado da Universidade College London, que colaborou na redação de algumas seções da publicação.

Essa publicação é fruto de um esforço coletivo de líderes educacionais e especialistas no campo da IA e educação.

Ela contou com as percepções e contribuições de diversos especialistas incluindo: Mutlu Cukurova, Professor da Universidade College London; Colin de la Higuera, Cátedra da UNESCO em Tecnologias para a Formação de Professores com Recursos Educacionais Abertos na Universidade de Nantes; Shafika Isaacs, Pesquisadora Associada da Universidade de Johannesburg; Natalie Lao, Diretora Executiva da Fundação App Inventor; Qin Ni, Professor Associado da Universidade Normal de Shanghai; Catalina Nicolin, Especialista em TIC na Educação no European Digital Hub na Romênia; John Shaw-Taylor, Cátedra da UNESCO em IA e Professor de Estatísticas Computacionais e Aprendizado de Máquina na Universidade College London; Kelly Shirohira, Gerente Executiva na Jet Education Services; Ki-Sang Song, Professor na Universidade Nacional de Educação da Coreia; e Ilkka Tuomi, Cientista Chefe na Meaning Processing Ltd na Finlândia.

Vários colegas na UNESCO também contribuíram de diversas maneiras, incluindo: Dafna Feinholz, Chefe da Seção de Bioética e Ética da Ciência e Tecnologia; Francesc Pedró, Diretor do Instituto Internacional para o Ensino Superior na América Latina e no Caribe; Prateek Sibal, Especialista de Programa, Seção de Políticas Digitais e Transformação Digital; Saurabh Roy, Oficial Sênior de Projeto na Seção de Desenvolvimento de Professores, Divisão de Políticas e Sistemas de Aprendizado ao Longo da Vida; Benjamin Vergel De Dios, Especialista de Programa em TIC na Educação, Seção de Inovação Educacional e Desenvolvimento de Habilidades no Escritório de Bangcoc; os colegas na Entidade de Diversidade de Expressões Culturais no Setor de Cultura; e Mark West, Especialista de Programa, Divisão de Futuro da Aprendizagem e Inovação.

Agradecimentos também são devidos a Glen Hertelendy, Luisa Ferrara e Xianglei Zheng, Unidade de Tecnologia e IA na Educação, Futuro da Aprendizagem e Inovação, por coordenarem a produção da publicação.

Os agradecimentos também se estendem a Jenny Webster, por revisar e corrigir o texto, e a Ngoc-Thuy Tran, por criar o *layout*.

Índice

Prefácio	2
Agradecimentos	3
Lista de acrônimos e de abreviaturas	6
Introdução	7
1. O que é IA generativa e como ela funciona?	8
1.1 O que é IA generativa?	8
1.2 Como funciona a IA generativa?	8
1.2.1 Como funcionam os modelos de IAGen de texto?	9
1.2.2 Como funcionam os modelos de IAGen de imagem?	11
1.3 Engenharia de <i>prompts</i> para gerar resultados desejados	12
1.4 O surgimento do EdGPT e suas implicações	13
2. Controvérsias sobre a IA generativa e suas implicações para a educação	14
2.1 Agravamento da pobreza digital	14
2.2 Superar a adaptação regulatória nacional	14
2.3 Uso de conteúdo sem consentimento	15
2.4 Modelos inexplicáveis usados para gerar resultados	15
2.5 Conteúdo gerado por IA que polui a Internet	16
2.6 Falta de compreensão do mundo real	16
2.7 Reduzir a diversidade de opiniões e marginalizar ainda mais as vozes já marginalizadas	17
2.8 Gerar <i>deepfakes</i> mais avançados	17
3. Regular o uso da IA generativa na educação	18
3.1 Uma abordagem para a IA centrada no ser humano	18
3.2 Passos para regular a IAGen na educação	18
3.3 Regulamentações sobre a IAGen: elementos-chave	20
3.3.1 Agências governamentais reguladoras	20
3.3.2 Provedores de ferramentas de IAGen	22
3.3.3 Usuários institucionais	23
3.3.4 Usuários individuais	23
4. Buscar um marco referencial político para o uso de IAGen na educação e na pesquisa	24
4.1 Promover a inclusão, a equidade e a diversidade linguística e cultural	24
4.2 Proteger a agência humana	24
4.3 Monitorar e validar sistemas de IAGen para a educação	25
4.4 Desenvolver competências em IA, incluindo habilidades relacionadas à IAGen, para os estudantes	26

4.5 Desenvolver a capacidade de professores e pesquisadores para o uso adequado da IAGen.....	26
4.6 Promover opiniões diversas e a expressão plural de ideias	27
4.7 Testar modelos de aplicações localmente relevantes e construir uma base cumulativa de evidências	27
4.8 Analisar as implicações de longo prazo de maneira intersetorial e interdisciplinar.....	28
5. Facilitar o uso criativo da IAGen na educação e na pesquisa	29
5.1 Estratégias institucionais para facilitar o uso responsável e criativo da IAGen.....	29
5.2 Uma abordagem de “interação centrada no ser humano e pedagogicamente adequada”	30
5.3 Co-projetando o uso da IAGen na educação e na pesquisa	30
5.3.1 IA generativa para a pesquisa	30
5.3.2 IA generativa para facilitar o ensino	31
5.3.3 IA generativa como um treinador 1:1 para a aquisição de habilidades fundamentais a um ritmo pessoal.....	32
5.3.4 IA generativa para facilitar a aprendizagem baseada em pesquisas ou projetos	34
5.3.5 IA generativa como apoio aos estudantes com necessidades especiais.....	35
6. IAGen e o futuro da educação e da pesquisa	37
6.1 Questões éticas não exploradas	37
6.2 Direitos autorais e propriedade intelectual	37
6.3 Fontes de conteúdo e aprendizagem	37
6.4 Respostas homogeneizadas <i>versus</i> produções diversas e criativas	38
6.5 Repensando avaliações e resultados de aprendizagem	38
6.6 Processos de pensamento	39
Observações finais.....	39
Referências	40

Lista de tabelas

Tabela 1. Técnicas utilizadas em IA generativa	8
Tabela 2. GPTs de OpenIA	9
Tabela 3. Co-projetar o uso da IAGen para a pesquisa	31
Tabela 4. Co-projetar o uso da IAGen para apoiar professores e o ensino.....	32
Tabela 5. Co-projetar o uso da IAGen como uma instrutora individual para a aquisição autodirigida de habilidades básicas em linguagens e artes	33
Tabela 6. Co-projetar o uso da IAGen para facilitar a aprendizagem baseada em pesquisas ou projetos	34
Tabela 7. Co-projetar o uso da IAGen para apoiar estudantes com necessidades especiais.....	35

Lista de acrônimos e de abreviaturas

Conceitos e tecnologias

API	Interface de Programação de Aplicações
GAN	Redes Generativas Adversariais
GB	Gigabytes
GPT	Transformador Generativo Pré-treinado
IA	Inteligência Artificial
IAD	Inteligência Artificial Distribuída
IAG	Inteligência Artificial Geral
IAGen	Inteligência Artificial Generativa
LaMDA	Modelo de Linguagem para Aplicações de Diálogo
LLM	Modelo de Linguagem Grande
ML	Aprendizado de máquina
RGPD	Regulamento Geral sobre a Proteção de Dados
RNA	Rede Neural Artificial
TIC	Tecnologia da informação e comunicação
VAE	<i>Autoencoders</i> (Codificadores Automáticos) Variacionais

Organizações

AGCC	Cluster Governamental de IA na Nuvem (Cingapura)
CAC	Administração do Ciberespaço da China
OCDE	Organização para Cooperação e Desenvolvimento Econômico
UE	União Europeia
UNCTAD	Conferência das Nações Unidas sobre Comércio e Desenvolvimento
UNESCO	Organização das Nações Unidas para a Educação, a Ciência e a Cultura

Introdução

O lançamento do ChatGPT no final de 2022, como a primeira ferramenta de inteligência artificial generativa (IAGen) de fácil utilização disponibilizada amplamente ao público¹ e seguida por versões iterativas mais sofisticadas, gerou uma comoção ao redor do mundo e está impulsionando a corrida entre grandes empresas de tecnologia para se posicionarem no campo de desenvolvimento de modelos de IAGen.²

Em todo o mundo, a preocupação inicial na área educacional era de que o ChatGPT e ferramentas semelhantes de IAGen fossem utilizadas por estudantes para trapacear na realização de suas tarefas, minando assim o valor da avaliação de aprendizagem, certificação e qualificações (Anders, 2023). Enquanto algumas instituições de ensino proibiram o uso do ChatGPT, outras receberam a chegada da IAGen com cautela (Tlili, 2023). Muitas escolas e universidades, por exemplo, adotaram uma abordagem progressista acreditando que “em vez de proibir o seu uso, os estudantes e funcionários precisam de apoio na utilização eficaz, ética e transparente das ferramentas de IAGen” (Russell Group, 2023). Essa abordagem reconhece que a IAGen já está amplamente disponível, provavelmente se tornará mais sofisticada e possui tanto um potencial negativo específico quanto um potencial positivo único para a educação.

De fato, a IAGen tem uma infinidade de usos possíveis. Ela pode automatizar o processamento de informações e a apresentação de resultados em todas as principais representações simbólicas do pensamento humano. Além disso, ela possibilita a entrega de resultados, fornecendo produtos de conhecimento semiacabados. Ao libertar os humanos de algumas categorias de habilidades de pensamento de nível inferior, essa nova geração de ferramentas de IA pode ter implicações profundas para a compreensão da inteligência humana e da aprendizagem.

No entanto, a IAGen levanta inúmeras preocupações imediatas relacionadas a questões como segurança, privacidade de dados, direitos autorais e manipulação. Algumas dessas preocupações constituem riscos mais amplos relacionados à inteligência artificial que foram agravadas pela IAGen, enquanto outras surgiram recentemente com essa última geração de ferramentas. É imprescindível que cada uma dessas questões e preocupações sejam plenamente compreendidas e abordadas.

Este Guia foi elaborado para atender a essa necessidade urgente. Porém, um conjunto temático de orientações sobre IAGen para a educação não deve ser interpretado como uma afirmação de que a IAGen é a solução para os desafios fundamentais da educação. Apesar da hiperbolização midiática, é improvável que a IAGen, sozinha, resolva qualquer um dos problemas enfrentados pelos sistemas educacionais ao redor do mundo. Ao lidar com os problemas educacionais de longa data, é fundamental afirmar que a capacidade humana e a ação coletiva, e não a tecnologia, são os fatores determinantes para se chegar a soluções eficazes ante os desafios fundamentais enfrentados pela sociedade.

Este Guia tem como objetivo apoiar o planejamento de regulamentações, políticas e programas de desenvolvimento de capacidades humanas adequados, para garantir que a IAGen se torne uma ferramenta que verdadeiramente beneficie e capacite professores, estudantes e pesquisadores. Baseado na “Recomendação sobre a Ética da Inteligência Artificial” da UNESCO, o Guia está fundamentado em uma abordagem centrada no ser humano que promove a agência humana, inclusão, equidade, igualdade de gênero, diversidade cultural e linguística, bem como opiniões e expressões plurais.

O Guia primeiro explora o que é a IAGen e como ela funciona, apresentando as diversas tecnologias e modelos disponíveis (Seção 1) e, em seguida, identifica uma série de questões éticas e políticas controversas relacionadas tanto à IA em geral quanto à IAGen especificamente (Seção 2). Depois, discute as etapas e elementos-chave a serem examinados para que a IAGen seja regulamentada com base em uma abordagem centrada no ser humano - aquela que garante o uso ético, seguro, equitativo e significativo (Seção 3). A Seção 4 propõe medidas que podem ser adotadas para o desenvolvimento de marcos políticos coerentes e abrangentes para regular o uso da IAGen na educação e na pesquisa, enquanto a Seção 5 explora as possibilidades de uso criativo da IAGen no *design* de currículos, ensino, aprendizado e atividades de pesquisa. A Seção 6 conclui o Guia com considerações sobre as implicações de longo prazo da IAGen para a educação e a pesquisa.

1. O que é IA generativa e como ela funciona?

1.1 O que é IA generativa?

A Inteligência Artificial Generativa (IAGen) é uma tecnologia de inteligência artificial (IA) que gera conteúdo de forma automática em resposta a comandos escritos em interfaces de conversação em linguagem natural. Em vez de simplesmente fazer a curadoria de páginas da web, aproveitando o conteúdo existente, a IAGen na verdade produz novo conteúdo. Esses conteúdos podem aparecer em formatos que compreendem todas as representações simbólicas do pensamento humano: textos escritos em linguagem natural, imagens (incluindo fotografias, pinturas digitais e desenhos animados), vídeos, música e código de *software*. A IAGen é treinada usando dados coletados de páginas da web, conversas em mídias sociais e outros meios *online*. Ela gera seu conteúdo analisando estatisticamente as distribuições de palavras, *pixels* ou outros elementos nos dados que foram consumidos e identificando e repetindo padrões comuns (por exemplo, quais palavras geralmente aparecem seguidas de outras palavras).

Embora a IAGen possa produzir novo conteúdo, ela não pode gerar novas ideias ou soluções para desafios do mundo real, pois não compreende objetos ou relações sociais do mundo real que sustentam a linguagem. Além disso, apesar de sua produção fluida e impressionante, não se pode confiar na precisão da IAGen. Até mesmo o provedor do ChatGPT reconhece:

“embora ferramentas como o ChatGPT frequentemente possam gerar respostas que parecem razoáveis, não se pode confiar que sejam precisas” (OpenAI, 2023). Na maioria das vezes, os erros passam despercebidos, a menos que o usuário tenha um conhecimento sólido sobre o tópico em questão.

1.2 Como funciona a IA generativa?

As tecnologias específicas por trás da IAGen fazem parte da família de tecnologias de Inteligência Artificial chamada aprendizado de máquina (*machine learning* [ML]), que utiliza algoritmos para possibilitar a melhoria contínua e automática de seu desempenho a partir de dados. O tipo de ML que tem levado a muitos dos avanços em IA que temos visto nos últimos anos, como o uso de IA para reconhecimento facial, é o conhecido como redes neurais artificiais (RNAs), inspirado em como o cérebro humano funciona e em suas conexões sinápticas entre neurônios. Existem muitos tipos de RNAs.

Tanto as tecnologias de IA generativa de texto quanto as de imagem são baseadas em um conjunto de tecnologias de IA que estão disponíveis para pesquisadores há vários anos. O ChatGPT, por exemplo, utiliza um transformador generativo pré-treinado (GPT), enquanto a IAGen de imagem geralmente utiliza as chamadas redes generativas adversariais (GANs) (ver **Tabela 1**).³

Tabela 1. Técnicas utilizadas em IA generativa

Aprendizado de máquina (ML)		Um tipo de IA que utiliza dados para aprimorar automaticamente seu desempenho.
Rede neural artificial (RNA)		Um tipo de ML inspirado na estrutura e funcionamento do cérebro humano (por exemplo, as conexões sinápticas entre neurônios).
IA generativa de texto	Transformadores de propósito geral	Um tipo de RNA que é capaz de focar em diferentes partes de dados para determinar como elas se relacionam entre si.
	Modelos de linguagem grandes (LLM)	Um tipo de transformador de propósito geral que é treinado com grandes quantidades de dados textuais.
	Transformador generativo pré-treinado (GPT) ⁴	Um tipo de modelo de linguagem grande (LLM) que é pré-treinado em quantidades ainda maiores de dados, o que permite ao modelo capturar as nuances da linguagem e gerar textos coerentes e de acordo com o contexto.
IA generativa de imagem	Redes generativas adversariais (GANs)	Tipos de rede neurais usados para geração de imagens.
	Autoencoders variacionais (VAEs)	

1.2.1 Como funcionam os modelos de IAGen de texto?

A IA generativa de texto utiliza um tipo de RNA conhecido como transformador de propósito geral, e um tipo de transformador de propósito geral chamado modelo de linguagem grande. É por isso que os sistemas de IAGen de texto são frequentemente referidos como modelos de linguagem grande, ou LLMs. O tipo de LLM usado pela IAGen de texto é conhecido como transformador generativo pré-treinado, ou GPT (de onde vem o “GPT” em “ChatGPT”).

O ChatGPT é construído com o GPT-3, que foi desenvolvido pela OpenAI. Esta foi a terceira versão do GPT, sendo que a primeira foi lançada em 2018 e a mais recente, GPT-4, em março de 2023 (ver **Tabela 2**). Cada versão do OpenAI GPT melhorou de maneira progressiva em relação à anterior por meio de avanços em arquiteturas de IA, métodos de treinamento e técnicas de otimização. Um aspecto

reconhecido de seu progresso contínuo é a utilização de quantidades cada vez maiores de dados para treinar seu número exponencialmente crescente de “parâmetros”. Parâmetros podem ser pensados como botões metafóricos que podem ser ajustados para aperfeiçoar o desempenho do GPT. Eles incluem os “pesos” do modelo, que são parâmetros numéricos que determinam como o modelo processa a entrada e produz o resultado.

Além dos avanços na otimização de arquiteturas de IA e métodos de treinamento, essa rápida evolução também foi possível devido às enormes quantidades de dados⁵ e melhorias nas capacidades computacionais disponíveis para as grandes empresas. Desde 2012, as capacidades computacionais usadas para treinar modelos de IAGen têm dobrado a cada 3 ou 4 meses. Em comparação, a Lei de Moore tinha um período de duplicação de dois anos (OpenAI, 2018; Universidade de Stanford, 2019).

Tabela 2. GPTs de OpenAI

Modelo	Lançamento	Quantidade de dados treinados	Números de parâmetros	Características
GPT-1	2018	40 GB	117 milhões	Capaz de realizar tarefas de processamento de linguagem natural, como completar textos e responder a perguntas.
GPT-2	2019	40 GB	1.500 milhões	Capaz de realizar tarefas mais complexas de processamento de linguagem natural, como tradução automática e elaboração de resumos.
GPT-3	2020	17.000 GB	175.000 milhões	Capaz de realizar tarefas avançadas de processamento de linguagem natural, como escrever parágrafos coerentes e gerar artigos inteiros. Também é capaz de se adaptar a novas tarefas com apenas alguns exemplos.
GPT-4 ⁶	2023	1.000.000 GB (relatado, mas não confirmado)	170.000.000 milhões (relatado, mas não confirmado)	Confiabilidade aprimorada e é capaz de processar instruções mais complexas.

Uma vez que o GPT foi treinado, gerar uma resposta de texto para um estímulo envolve os seguintes passos:

1. O estímulo é dividido em unidades menores (chamadas de *tokens*) que são inseridas no GPT.
2. O GPT utiliza padrões estatísticos para prever quais palavras ou frases poderiam formar uma resposta coerente ao estímulo.
 - O GPT identifica padrões de palavras e frases que frequentemente coocorrem em seu modelo de dados pré-construído (que inclui textos extraídos da Internet e de outros lugares).
 - Usando esses padrões, o GPT estima a probabilidade de palavras ou frases específicas aparecerem em dado contexto.
 - Começando com uma previsão aleatória, o GPT utiliza essas probabilidades estimadas para prever a próxima palavra ou frase em sua resposta.
3. As palavras ou frases previstas são convertidas em texto legível.
4. O texto legível passa por filtros conhecidos como “guarda-corpos” (*guardrails*) para que seja removido qualquer conteúdo ofensivo.

5. Os passos de 2 a 4 são repetidos até que uma resposta seja concluída. A resposta é considerada concluída quando atinge um limite máximo de *tokens* ou atende a critérios de parada predefinidos.
6. A resposta é pós-processada para melhorar a legibilidade por meio da aplicação de formatação, pontuação e outros aprimoramentos (como começar a resposta com palavras que um humano poderia usar, como “claro”, “certamente” ou “desculpas”).

Apesar de os GPTs e a sua capacidade de gerar textos de forma automática estarem disponíveis para pesquisadores desde 2018, o que tornou o lançamento do ChatGPT tão inovador foi o acesso gratuito por meio de uma interface fácil de usar, o que significa que qualquer pessoa com acesso à Internet pode explorar a ferramenta. O lançamento do ChatGPT causou comoção ao redor do mundo e rapidamente levou outras empresas globais de tecnologia a se readequarem, juntamente com inúmeras *startups*, seja lançando seus próprios sistemas semelhantes ou construindo novas ferramentas com base nele.

Até julho de 2023, algumas das alternativas ao ChatGPT incluíam:

- **Alpaca:**⁷ uma versão aprimorada do Llama da Meta, da Universidade de Stanford, que tem como objetivo lidar com informações falsas, estereótipos sociais e linguagem tóxica em LLMs.
- **Bard:**⁸ um LLM da Google, baseado em seus sistemas LaMDA e PaLM 2, que possui acesso à Internet em tempo real, o que significa que ele pode fornecer informações atualizadas.
- **Chatsonic:**⁹ desenvolvido pela Writesonic, ele se baseia no ChatGPT enquanto também coleta dados diretamente do Google. Portanto, tem menos probabilidade de fornecer respostas factualmente incorretas.
- **Ernie** (também conhecido como **Wenxin Yiyao** 文心一言):¹⁰ um LLM bilíngue da Baidu, ainda em desenvolvimento, que integra amplo conhecimento com conjuntos massivos de dados para gerar texto e imagens.
- **Hugging Chat:**¹¹ desenvolvido pela Hugging Face, que enfatizou questões de ética e a transparência durante todo o seu desenvolvimento, treinamento e implementação. Além disso, todos os dados

utilizados para treinar seus modelos são de código aberto.

- **Jasper:**¹² um conjunto de ferramentas e APIs que pode, por exemplo, ser treinado para escrever no estilo preferido de um usuário específico. Também pode gerar imagens.
- **Llama:**¹³ um LLM de código aberto da Meta que requer menos poder computacional e menos recursos para testar novas abordagens, validar o trabalho de outros e explorar novos casos de uso.
- **Open Assistant:**¹⁴ um sistema de código aberto projetado para permitir que qualquer pessoa com experiência suficiente desenvolva seu próprio LLM. Foi construído com base em dados de treinamento selecionados por voluntários.
- **Tongyi Qianwen** (通义千问):¹⁵ um LLM da Alibaba que responde a comandos em inglês ou chinês. Está sendo integrado ao pacote de ferramentas de negócios da Alibaba.
- **YouChat:**¹⁶ um LLM que incorpora mecanismos de busca em tempo real para fornecer contextos e *insights* adicionais, visando gerar resultados mais precisos e confiáveis.

A maioria deles é de uso gratuito (com certas limitações), enquanto alguns são de código aberto. Vários outros produtos estão sendo lançados com base em algum desses LLMs. Alguns exemplos são:

- **ChatPDF:**¹⁷ resume e responde a perguntas sobre documentos enviados em formato PDF.
- **Elicit: The AI Research Assistant:**¹⁸ tem como objetivo automatizar partes dos fluxos de trabalho de pesquisadores, identificando artigos relevantes e resumindo informações-chave.
- **Perplexity:**¹⁹ fornece um “hub de conhecimento” para pessoas que buscam respostas rápidas e precisas adaptadas às suas necessidades.

Da mesma forma, ferramentas baseadas em LLM estão sendo incorporadas a outros produtos, como navegadores da web. Por exemplo, extensões para o navegador Chrome que foram construídas com base no ChatGPT incluem:

- **WebChatGPT:**²⁰ fornece ao ChatGPT acesso à Internet para possibilitar conversas mais precisas e atualizadas.
- **Compose IA:**²¹ autocompleta frases em *e-mails* e outros lugares.
- **TeamSmart IA:**²² fornece uma “equipe de assistentes virtuais”.
- **Wiseone:**²³ simplifica informações *online*.

Além disso, o ChatGPT foi incorporado em algumas ferramentas de busca²⁴ e está sendo implementado em grandes conjuntos de ferramentas de produtividade (por exemplo, Microsoft Word e Excel), tornando-o ainda mais acessível em escritórios e instituições de ensino em todo o mundo (Murphy Kelly, 2023).

Por fim, em uma transição interessante para a IAGen de imagem, o GPT mais recente da OpenAI, o GPT-4, é capaz de aceitar tanto imagens quanto textos em seus comandos. Nesse sentido, ele é multimodal. Por consequência, argumenta-se que o termo “modelo de linguagem grande” (LLM) está se tornando menos apropriado, razão pela qual os pesquisadores da Universidade de Stanford propuseram o termo “modelo de fundação” (Bommasani *et al.*, 2021). Essa alternativa ainda não foi amplamente adotada.

1.2.2 Como funcionam os modelos de IAGen de imagem?

A IAGen de imagem e a IAGen de música geralmente utilizam um tipo diferente de rede neural artificial (RNA) conhecido como redes generativas adversariais (GANs), que também podem ser combinadas com *autoencoders* variacionais (VAEs). As GANs têm duas partes (dois “adversários”): o “gerador” e o “discriminador”. No caso das GANs de imagem, o gerador cria uma imagem aleatória em resposta a um comando, e o discriminador tenta distinguir entre essa imagem gerada e imagens reais. O gerador, então, usa o resultado do discriminador para ajustar seus parâmetros, a fim de criar outra imagem. O processo é repetido, possivelmente milhares de vezes, com o gerador criando imagens cada vez mais realistas que o discriminador, por sua vez, tem cada vez mais dificuldade em distinguir de imagens reais. Por exemplo, uma GAN bem-sucedida, treinada em um conjunto de dados de milhares de fotografias de paisagens, pode gerar novas, ainda que irreais, imagens de paisagens quase indistinguíveis de fotografias reais. Enquanto isso, uma GAN treinada em um conjunto de dados de música popular (ou até mesmo de músicas de um único artista) pode gerar novas peças musicais

que seguem a estrutura e a complexidade da música original.

A partir de julho de 2023, os modelos de **IAGen de imagem** disponíveis incluem os listados abaixo, todos os quais geram imagens a partir de comandos de texto. A maioria é de uso gratuito, com algumas limitações:

- **Craiyon:**²⁵ anteriormente conhecido como mini DALL·E.
- **DALL·E 2:**²⁶ ferramenta de IAGen de imagem da OpenAI.
- **DreamStudio:**²⁷ ferramenta de IAGen de imagem da Stable Diffusion.
- **Fotor:**²⁸ incorpora a IAGen em uma variedade de ferramentas de edição de imagem.
- **Midjourney:**²⁹ uma ferramenta independente de IAGen de imagem.
- **NightCafe:**³⁰ interface para Stable Diffusion e DALL·E 2.
- **Photosonic:**³¹ gerador de arte de IA da WriteSonic.

Exemplos de **IAGen de vídeo** de fácil acesso incluem:

- **Elai:**³² pode converter apresentações, *sites* e texto em vídeos.
- **GliaCloud:**³³ pode gerar vídeos a partir de conteúdo de notícias, postagens em redes sociais, eventos esportivos ao vivo e dados estatísticos.
- **Pictory:**³⁴ pode criar vídeos curtos a partir de conteúdos extensos e de forma automática.
- **Runway:**³⁵ oferece uma variedade de ferramentas de geração e edição de vídeo (e imagem).

Finalmente, abaixo estão alguns exemplos de **IAGen de música** de fácil acesso:

- **Aiva:**³⁶ pode criar trilhas sonoras personalizadas de forma automática.
- **Boomy,**³⁷ **Soundraw,**³⁸ e **Voicemod:**³⁹ pode gerar músicas a partir de qualquer texto e não exige conhecimento em composição musical.

1.3 Engenharia de prompts para gerar resultados desejados

Embora usar a IAGen possa ser tão simples quanto digitar uma pergunta ou outro comando, a realidade é que ainda não é tão fácil para o usuário obter exatamente o resultado desejado. Por exemplo, a inovadora imagem de IA *Théâtre D'opéra Spatial*, que ganhou um prêmio na Feira Estadual do Colorado, nos Estados Unidos, precisou de semanas de redação de *prompts* e de ajuste de centenas de imagens para gerar a versão final (Roose, 2022). Um desafio semelhante, de escrever comandos eficazes para a IAGen de texto, tem levado ao aumento do número de empregos na área de “engenharia de *prompts*” aparecendo em sites de recrutamento (Popli, 2023). A “engenharia de *prompts*” refere-se aos processos e técnicas utilizados para compor entradas a fim de produzir um resultado de IAGen que se assemelhe mais à intenção original do usuário.

A engenharia de *prompts* tem mais êxito quando o comando articula uma cadeia de raciocínio coerente centrada em um problema específico ou uma sequência de pensamento em uma ordem lógica. Recomendações específicas incluem:

- Utilize uma linguagem **simples**, clara e direta que possa ser facilmente compreendida, evitando palavras complexas ou ambíguas.
- Inclua **exemplos** para ilustrar a resposta desejada ou o formato das conclusões geradas.
- Inclua o **contexto**, o qual é essencial para gerar conclusões relevantes e significativas.
- **Refine** e faça a iteração conforme necessário, experimentando com diferentes variações.
- Seja **ético**, evitando comandos que possam gerar conteúdo inadequado, tendencioso ou prejudicial.

Também é importante reconhecer desde já que os resultados gerados da IAGen não podem ser considerados confiáveis sem uma avaliação crítica. A OpenAI afirma o seguinte sobre o seu GPT mais sofisticado:⁴⁰



Apesar de suas capacidades, o GPT-4 possui limitações semelhantes aos modelos anteriores do GPT. Destaca-se que ele ainda não é totalmente confiável (‘alucina’

fatos e comete erros de raciocínio). Deve-se ter muito cuidado ao usar os resultados do modelo de linguagem, especialmente em contextos de alto risco, devendo-se seguir um protocolo exato (como realizando revisão humana, fundamentação com contexto adicional ou evitando usos de alto risco), adaptando-se às necessidades de um uso específico.”

Tendo em vista a qualidade dos resultados da IAGen, testes com usuários e avaliações de desempenho rigorosos devem ser conduzidos antes de validar as ferramentas para sua adoção em larga escala ou em contextos de alto risco. Tais exercícios devem ser pensados a partir de uma métrica de desempenho que seja a mais relevante para o tipo de tarefa para a qual os usuários solicitam que a IAGen forneça resultados. Por exemplo, para resolver problemas matemáticos, a “precisão” poderia ser usada como a principal métrica para quantificar com que frequência uma ferramenta de IAGen produz a resposta correta; para responder a perguntas sensíveis, a principal métrica para medir o desempenho poderia ser a “taxa de resposta” (com que frequência a IAGen responde diretamente a uma pergunta); para geração de códigos, a métrica poderia ser “a fração dos códigos gerados que são diretamente executáveis” (se o código gerado pode ser executado diretamente em um ambiente de programação e passar nos testes unitários); e para o raciocínio visual, a métrica poderia ser “correspondência exata” (se os objetos visuais gerados correspondem exatamente à realidade) (Chen *et al.*, 2023).



Implicações para a educação e a pesquisa:

Embora a IAGen possa ajudar professores e pesquisadores a gerar textos úteis e outros resultados para apoiar seu trabalho, não é necessariamente um processo direto. Podem ser necessárias várias iterações de um *prompt* antes que o resultado desejado seja alcançado. Uma preocupação é que os estudantes mais jovens, por definição menos experientes do que os professores, possam aceitar sem perceber e sem engajamento crítico um resultado da IAGen que seja superficial, impreciso ou até mesmo prejudicial.

Em resumo, em um nível superficial, a IAGen é fácil de usar; no entanto, resultados mais sofisticados requerem a intervenção de pessoas qualificadas e devem ser avaliados criticamente antes de serem utilizados.

1.4 O surgimento do EdGPT e suas implicações

Dado que os modelos da IAGen podem servir como base ou ponto de partida para o desenvolvimento de modelos mais especializados ou específicos para determinados domínios, alguns pesquisadores sugeriram que os GPTs deveriam ser renomeados como “modelos fundamentais” (Bommasani *et al.*, 2021). Na educação, desenvolvedores e pesquisadores começaram a ajustar um modelo fundamental para desenvolver o “EdGPT”.⁴¹ Modelos EdGPT são treinados com dados específicos para fins educacionais. Em outras palavras, o EdGPT tem como objetivo refinar o modelo derivado de grandes volumes de dados de treinamento gerais usando quantidades menores de dados educacionais, específicos da área e de alta qualidade.

Isso potencialmente confere ao EdGPT maior amplitude para apoiar a realização das transformações listadas na Seção 4.3. Por exemplo, os modelos EdGPT que visam a conceção conjunta de currículos podem permitir que educadores e alunos gerem materiais educativos apropriados, tais como planos de aula, questionários e atividades interativas que se alinhem estreitamente com uma abordagem pedagógica eficaz e com objetivos curriculares e níveis de desafio específicos para determinados alunos. Da mesma forma, no contexto de um instrutor individual de habilidades linguísticas, um modelo fundamental refinado com textos apropriados para um idioma específico pode ser usado para gerar sentenças, parágrafos ou conversas de exemplo a serem usados na prática. Quando os estudantes interagem com o modelo, ele pode responder com um texto relevante e gramaticalmente correto e que esteja no nível certo para eles. Em teoria, os resultados dos modelos EdGPT também

podem conter menos vieses gerais ou conteúdo questionável do que o GPT padrão, mas ainda assim podem gerar erros. É crucial observar que, a menos que os modelos subjacentes da IAGen e a abordagem mudem significativamente, o EdGPT ainda pode gerar erros e apresentar outras limitações. Portanto, ainda é importante que os principais usuários do EdGPT, especialmente professores e estudantes, adotem uma perspectiva crítica em relação a quaisquer resultados.

Atualmente, o aprimoramento de modelos fundamentais para um uso mais direcionado do GPT na educação está em estágio inicial. Entre os exemplos existentes está o EduChat, um modelo fundamental desenvolvido pela East China Normal University para fornecer serviços de ensino e aprendizado, cujos códigos, dados e parâmetros são de código aberto.⁴² Outro exemplo é o MathGPT, um LLM que está sendo desenvolvido pelo TAL Education Group e que se concentra na resolução de problemas relacionados à matemática e em fornecer explicações a usuários em todo o mundo.⁴³

No entanto, antes que progressos significativos sejam possíveis, é essencial direcionar esforços para aprimorar modelos fundamentais não apenas adicionando conhecimentos específicos de disciplinas e reduzindo vieses, numa lógica de *de-biasing*, mas também incorporando conhecimentos sobre métodos de aprendizagem relevantes e sobre como isso pode ser refletido na elaboração de algoritmos e modelos. O desafio é determinar em que medida os modelos EdGPT podem ir além do conhecimento específico da disciplina e focar também na pedagogia centrada no estudante e em interações positivas entre professor e estudante. Um desafio adicional é determinar em que medida os dados de estudantes e professores podem ser coletados e usados eticamente para informar um EdGPT. Por fim, também há a necessidade de se realizar investigações robustas para garantir que o EdGPT não comprometa os direitos humanos dos estudantes e nem desempodere os professores.

2. Controvérsias sobre a IA generativa e suas implicações para a educação

Após discutir sobre o que é a IAGen e como ela funciona, esta seção examina as controvérsias e riscos éticos suscitados por todos os sistemas de IAGen e considera algumas de suas implicações para a educação.

2.1 Agravamento da pobreza digital

Como mencionado anteriormente, a IAGen depende de enormes quantidades de dados e de um enorme poder computacional, bem como de suas inovações iterativas em arquiteturas de IA e métodos de treinamento, que na maioria das vezes estão disponíveis apenas para as maiores empresas de tecnologia internacionais e para algumas economias (principalmente os Estados Unidos, a República Popular da China e, em menor medida, a Europa). Isso significa que a possibilidade de criar e controlar a IAGen está fora do alcance da maioria das empresas e da maioria dos países, especialmente aqueles do Sul Global.

À medida que o acesso aos dados se torna cada vez mais essencial para o desenvolvimento econômico dos países e para as oportunidades digitais dos indivíduos, aqueles países e pessoas que não têm acesso a dados suficientes ou que não podem arcar com os custos ficam em uma situação de “pobreza de dados” (Marwala, 2023). A situação é semelhante em relação ao acesso ao poder computacional. A rápida difusão da IAGen em países e regiões tecnologicamente avançados acelerou exponencialmente a geração e o processamento de dados, intensificando simultaneamente a concentração

de riqueza em IA no Norte Global. Como consequência imediata, as regiões com falta de dados foram ainda mais excluídas e expostas ao risco, de longo prazo, de serem colonizadas pelos padrões incorporados nos modelos GPT. Os modelos atuais do ChatGPT são treinados com dados de usuários *online* que refletem os valores e as normas do Norte Global, tornando-os inadequados aos algoritmos de IA localmente relevantes em comunidades carentes de dados em várias partes do Sul Global ou em comunidades mais desfavorecidas do Norte Global.

2.2 Superar a adaptação regulatória nacional

Os principais provedores de IAGen também têm sido criticados por não permitirem que seus sistemas sejam sujeitos a revisões acadêmicas independentes e rigorosas (Dwivedi *et al.*, 2023).⁴⁴ As tecnologias fundamentais da IAGen de uma empresa tendem a ser protegidas como propriedade intelectual corporativa. Enquanto isso, muitas das empresas que estão começando a usar a IAGen estão enfrentando desafios cada vez maiores para conseguir manter a segurança de seus sistemas (Lin, 2023). Além disso, apesar dos apelos para que se tenha regulamentação, vindos da própria indústria de IA,⁴⁵ a elaboração de legislação sobre a criação e o uso de todo o tipo de IA, incluindo a IAGen, frequentemente encontra dificuldades de acompanhar o seu rápido ritmo de desenvolvimento. Isso explica, em parte, os desafios enfrentados por agências nacionais ou locais para compreender e governar as questões legais e éticas.⁴⁶

Embora a IAGen possa ampliar as capacidades humanas na conclusão de certas tarefas, o controle democrático das empresas promotoras da IAGen é limitado. Isso levanta a questão da regulamentação, especialmente no que diz respeito ao acesso e uso de dados domésticos, incluindo dados de instituições locais e indivíduos, bem como dados gerados no território dos países. É necessária uma legislação apropriada para que as agências governamentais locais possam obter algum controle sobre as ondas crescentes de IAGen, garantindo sua governança como um bem público.



Implicações para a educação e a pesquisa:

Pesquisadores, professores e estudantes devem adotar uma visão crítica das orientações acerca dos valores, padrões culturais e costumes sociais incorporados nos modelos de treinamento da IAGen. Os formuladores de políticas devem estar cientes e adotar medidas para enfrentar o agravamento das desigualdades causadas pela crescente lacuna existente no treinamento e no controle dos modelos IAGen.



Implicações para a educação e a pesquisa:

Pesquisadores, professores e estudantes devem estar cientes da falta de regulamentações adequadas para proteger a propriedade das instituições nacionais e dos indivíduos e os direitos dos usuários domésticos da IAGen, bem como para lidar com questões legislativas desencadeadas por ela.

2.3 Uso de conteúdo sem consentimento

Como mencionado anteriormente, os modelos de IAGen são construídos a partir de grandes volumes de dados (por exemplo, texto, sons, código e imagens), frequentemente retirados da Internet e geralmente sem permissão do proprietário. Muitos sistemas de IAGen de imagens, e alguns sistemas de IAGen de código, foram acusados, consequentemente, de violar direitos de propriedade intelectual. Na data de redação da presente publicação, existem vários casos jurídicos internacionais em andamento relacionados a esse problema.

Além disso, já foi apontado que os GPTs podem violar leis como o Regulamento Geral sobre a Proteção de Dados (GDPR) da União Europeia (2016), especialmente o direito ao esquecimento, visto que atualmente é impossível remover os dados de alguém (ou os resultados desses dados) de um modelo GPT uma vez que ele tenha sido treinado.



Implicações para a educação e a pesquisa:

- Pesquisadores, professores e estudantes precisam conhecer os direitos dos proprietários de dados e devem verificar se as ferramentas de IAGen que estão utilizando violam alguma regulamentação vigente.
- Pesquisadores, professores e estudantes também devem estar cientes de que as imagens ou códigos criados com a IAGen podem violar os direitos de propriedade intelectual de outra pessoa, e que as imagens, sons ou código que eles criam e compartilham na Internet podem ser explorados por outras IAGen.

2.4 Modelos inexplicáveis usados para gerar resultados

Não é de hoje que as redes neurais artificiais (RNAs) são reconhecidas como “caixas pretas”; ou seja, seu funcionamento interno não é aberto para inspeção. Como resultado, as RNAs acabam não sendo “transparentes” ou “explicáveis”, e não é possível determinar de que forma seus resultados foram obtidos.

Embora a abordagem geral, incluindo os algoritmos utilizados, normalmente seja explicável, os modelos específicos e seus parâmetros, incluindo os pesos do modelo, não são passíveis de inspeção, razão pela qual não se pode explicar um resultado específico gerado. Há bilhões de parâmetros/pesos em um modelo como o GPT-4 (ver **Tabela 2**), e são os pesos que, em conjunto, armazenam os padrões aprendidos que o modelo utiliza para gerar seus resultados. Como os parâmetros/pesos não são transparentes nas RNAs (**Tabela 1**), não é possível explicar exatamente como um resultado específico é criado por esses modelos.

A falta de transparência e explicabilidade da IAGen é cada vez mais problemática à medida que ela se torna mais complexa (ver **Tabela 2**), muitas vezes produzindo resultados inesperados ou indesejados. Além disso, os modelos de IAGen herdam e perpetuam vieses presentes nos dados usados para treiná-los, os quais, dada a natureza não transparente dos modelos, são difíceis de detectar e corrigir. Por fim, essa opacidade também é uma das principais causas de problemas de confiança relacionados à IAGen (Nazaretsky *et al.*, 2022a). Se os usuários não compreendem como um sistema de IAGen chegou a um resultado específico, eles são menos propensos a adotá-lo ou usá-lo (Nazaretsky *et al.*, 2022b).



Implicações para a educação e a pesquisa:

Pesquisadores, professores e estudantes devem estar cientes de que os sistemas de IAGen operam como caixas pretas e, consequentemente, é difícil, senão impossível, saber o porquê de um conteúdo específico ter sido criado. A falta de explicação sobre como os resultados são gerados tende a impor aos usuários a lógica definida pelos parâmetros projetados nos sistemas de IAGen. Esses parâmetros podem refletir valores e normas culturais ou comerciais específicas que distorcem implicitamente o conteúdo produzido.

2.5 Conteúdo gerado por IA que polui a Internet

Como os dados usados no treinamento do GPT são normalmente extraídos da Internet, a qual frequentemente contém linguagem discriminatória e outros tipos de linguagem considerados inaceitáveis, os desenvolvedores tiveram que implementar o que chamam de “guarda-corpos” (*guardrails*) para evitar que os resultados do GPT sejam ofensivos e/ou antiéticos. No entanto, devido à ausência de regulamentações rigorosas e de mecanismos eficazes de monitoramento, materiais tendenciosos gerados pela IAGen estão se espalhando cada vez mais pela Internet, poluindo uma das principais fontes de conteúdo ou conhecimento para a maioria dos estudantes em todo o mundo. Isso é especialmente importante porque o material gerado pela IAGen pode parecer bastante preciso e convincente, quando muitas vezes pode conter erros e ideias tendenciosas. Isso representa um alto risco para os jovens estudantes que não possuem conhecimentos prévios e sólidos sobre o tópico em questão. Também representa um risco recursivo para futuros modelos de GPT que serão treinados com textos extraídos da Internet, que os próprios modelos de GPT criaram e que também incluem seus vieses e erros.



Implicações para a educação e a pesquisa:

- Pesquisadores, professores e estudantes precisam estar cientes de que os sistemas de IAGen podem gerar materiais ofensivos e antiéticos.
- Também precisam estar cientes dos problemas de longo prazo que podem surgir em relação à confiabilidade do conhecimento, visto que os futuros modelos de GPT utilizarão como base os textos gerados por modelos de GPT anteriores.

2.6 Falta de compreensão do mundo real

Os GPTs de texto são, às vezes, pejorativamente chamados de “papagaios estocásticos” pois, como mencionado anteriormente, embora possam produzir textos que parecem convincentes, esses textos muitas vezes contêm erros e podem incluir afirmações prejudiciais (Bender *et al.*, 2021). Isso ocorre porque os GPTs apenas repetem padrões de

linguagem encontrados em seus dados de treinamento (geralmente textos extraídos da Internet), começando com padrões aleatórios (ou “estocásticos”), e sem entender seu significado – assim como um papagaio pode imitar sons sem realmente compreender o que está dizendo.

A desconexão entre os modelos de IAGen “parecerem” entender os textos que usam e geram, e a “realidade” de que não entendem a linguagem e o mundo real, pode levar professores e estudantes a depositarem nos resultados um nível de confiança não justificado. Isso representa sérios riscos para o futuro da educação. De fato, a IAGen não é informada com base em observações do mundo real ou em outros aspectos fundamentais do método científico, nem está alinhada com valores humanos ou sociais. Por essas razões, ela não consegue gerar conteúdo genuinamente novo sobre o mundo real, objetos e suas relações, pessoas e relações sociais, relações humano-objeto ou relações humano-tecnologia. Ainda é incerto dizer se os conteúdos aparentemente novos gerados pelos modelos de IAGen podem ser reconhecidos como conhecimento científico.

Como já mencionado, os GPTs frequentemente produzem textos imprecisos ou não confiáveis. Sabe-se, ainda, que os GPTs inventam coisas que não existem na vida real. Isso é, por vezes, chamado de “alucinação”, embora o uso desse termo seja criticado por ser antropomórfico e, portanto, enganoso. Esse fato é reconhecido pelas empresas que produzem a IAGen. Por exemplo, na parte inferior da interface pública do ChatGPT há a frase: “ChatGPT pode fornecer informações imprecisas sobre pessoas, lugares ou fatos”.²

Alguns defensores têm sugerido que a IAGen representa um passo significativo na jornada em direção à inteligência artificial geral (IAG), um termo que sugere uma classe de IA mais inteligente do que os humanos. No entanto, isso tem sido criticado há tempos, pois argumenta-se de que a IA nunca progredirá em direção à IAG, pelo menos até que, de alguma forma, una em simbiose tanto a IA baseada em conhecimento (também conhecida como IA simbólica ou baseada em regras) quanto a IA baseada em dados (também conhecida como aprendizado de máquina) (Marcus, 2022). As afirmações sobre IAG ou senciência também nos desviam de uma consideração mais cuidadosa dos danos atuais perpetrados com a IA, como a discriminação oculta contra grupos que já são discriminados (Metz, 2021).



Implicações para a educação e a pesquisa:

- O resultado de uma IAGen de texto pode parecer impressionantemente semelhante ao de um humano, como se ela entendesse o texto que gerou. No entanto, a IAGen não é capaz de entender. Em vez disso, essas ferramentas juntam as palavras de maneiras que são comuns na Internet. O texto gerado também pode estar incorreto.
- Pesquisadores, professores e estudantes precisam estar cientes de que um GPT não entende o texto que gera; que ele pode – e frequentemente o faz – gerar afirmações incorretas; e, portanto, é necessário adotar uma perspectiva crítica sobre tudo o que ele gera.

2.7 Reduzir a diversidade de opiniões e marginalizar ainda mais as vozes já marginalizadas

O ChatGPT e outras ferramentas semelhantes normalmente geram apenas respostas padrão que assumem os valores dos proprietários/criadores dos dados usados para treinar os modelos. Se uma sequência de palavras aparece com frequência nos dados de treinamento – como é o caso de tópicos comuns e não controversos e crenças predominantes ou dominantes – é provável que ela seja repetida pelo GPT em seus resultados.

Assim, há o risco de limitar e minar o desenvolvimento de opiniões e expressões plurais de ideias. Populações com pobreza de dados, incluindo comunidades marginalizadas no Norte Global, têm mínima ou limitada presença digital *online*. Suas vozes, portanto, não estão sendo ouvidas, e suas preocupações não estão representadas nos dados que são usados para treinar os GPTs, aparecendo raramente nos resultados. Por essas razões, dada a metodologia de pré-treinamento baseada em dados de páginas da web e conversas em mídias sociais, os modelos GPT podem marginalizar ainda mais pessoas já desfavorecidas.



Implicações para a educação e a pesquisa:

- Embora os desenvolvedores e fornecedores de modelos de IAGen sejam os principais responsáveis por abordar de forma contínua os vieses presentes nos conjuntos de dados e resultados desses modelos, os pesquisadores, professores e alunos, na posição de usuários, precisam estar cientes de que os resultados da IAGen de texto representam apenas a visão mais comum ou dominante de mundo à época em que seus dados de treinamento foram produzidos e que parte deles é problemático ou tendencioso (por exemplo, com papéis de gênero estereotipados).
- Estudantes, professores e pesquisadores não devem aceitar as informações fornecidas pela IAGen como verdade absoluta e devem sempre avaliá-las de maneira crítica.
- Pesquisadores, professores e estudantes também devem estar cientes de como as vozes de minorias podem ser negligenciadas, pois, por definição, são menos frequentes nos dados de treinamento.

2.8 Gerar deepfakes mais avançados

Além das controvérsias comuns a todas as IAGen, a IAGen de GANs pode ser usada para alterar ou manipular imagens ou vídeos existentes a fim de gerar falsificações que são difíceis de serem distinguidas das versões reais. A IAGen está facilitando cada vez a criação desses *deepfakes* e das chamadas *fake news*. Em outras palavras, a IAGen está facilitando que determinados atores cometam atos antiéticos, imorais e criminosos, como disseminar desinformação, promover discursos de ódio e incorporar o rosto de pessoas, sem o conhecimento ou consentimento delas, em materiais audiovisuais completamente falsos e, por vezes, comprometedores.



Implicações para a educação e a pesquisa:

Embora seja obrigação dos provedores de IAGen proteger os direitos autorais e de imagem dos usuários, pesquisadores, professores e estudantes também precisam estar cientes de que quaisquer imagens que compartilhem na Internet podem ser incorporadas aos dados de treinamento da IAGen e podem ser manipuladas e usadas de maneiras antiéticas.

3. Regular o uso da IA generativa na educação

Para lidar com as controvérsias em torno da IA generativa e aproveitar os seus benefícios potenciais na educação, é necessário, em primeiro lugar, regulamentá-la. A regulamentação da IAGen para fins educacionais requer uma série de etapas e medidas políticas baseadas em uma abordagem centrada no ser humano para garantir seu uso ético, seguro, equitativo e significativo.

3.1 Uma abordagem para a IA centrada no ser humano

A “Recomendação sobre a Ética da Inteligência Artificial” de 2021 da UNESCO oferece o marco normativo necessário para começar a abordar as múltiplas controvérsias em torno da IA generativa, incluindo aquelas relacionadas à educação e à pesquisa. A recomendação se baseia em uma abordagem de IA centrada no ser humano, que defende que o uso da IA deve estar a serviço do desenvolvimento das capacidades humanas para um futuro inclusivo, justo e sustentável. Essa abordagem deve ser guiada pelos princípios dos direitos humanos e pela necessidade de proteger a dignidade humana e a diversidade cultural que define o conhecimento comum (*knowledge commons*). Em termos de governança, uma abordagem centrada no ser humano requer uma regulamentação adequada que possa garantir a agência humana, transparência e responsabilidade pública.

O “Consenso de Beijing sobre a inteligência artificial e a educação” de 2019 detalha o que uma abordagem centrada no ser humano implica para o uso da IA no contexto da educação. O Consenso afirma que o uso de tecnologias de IA na educação deve aprimorar as capacidades humanas para o desenvolvimento sustentável e a colaboração eficaz entre humanos e máquinas na vida e em contextos de aprendizado e trabalho. Também propõe que sejam adotadas ações adicionais com o objetivo de garantir acesso equitativo à IA para apoiar pessoas marginalizadas e abordar desigualdades, ao mesmo tempo em que promove a diversidade linguística e cultural. O Consenso sugere a adoção de abordagens governamentais integradas, intersetoriais e multissetoriais para o planejamento de políticas sobre IA na educação.

No guia “IA e Educação: Orientações para os Formuladores de Políticas (*AI and education: guidance for policy-makers*)” (UNESCO, 2022b) aprofunda-se ainda mais o debate sobre as implicações de uma abordagem centrada no ser humano, examinando os benefícios e riscos da IA na educação e o papel da educação como meio de desenvolver competências em IA. Propõe recomendações concretas para a formulação de políticas a fim de orientar o uso da IA para (i) possibilitar o acesso inclusivo a programas de aprendizado, especialmente para grupos vulneráveis, como alunos com deficiência; (ii) apoiar opções de aprendizado personalizadas e abertas; (iii) melhorar as provisões e a gestão baseadas em dados para expandir o acesso e aprimorar a qualidade do aprendizado; (iv) monitorar os processos de aprendizado e alertar os professores sobre riscos de falha; e (v) desenvolver a compreensão e as habilidades para o uso ético e significativo da IA.

3.2 Passos para regular a IAGen na educação

Antes do lançamento do ChatGPT, os governos vinham desenvolvendo ou adaptando marcos regulatórios para a coleta e o uso de dados e a adoção de sistemas de IA em diversos setores, incluindo a educação, o que proporcionou um contexto legislativo e político para a regulamentação de novas aplicações de IA que foram surgindo. Após o lançamento de vários modelos competitivos de IAGen a partir de novembro de 2022, os governos têm adotado diferentes respostas em suas políticas - desde a proibição da IAGen até a avaliação da necessidade de adaptação de estruturas existentes ou a formulação urgente de novas regulamentações.

Estratégias governamentais para regular e facilitar o uso criativo da IAGen foram mapeadas e revisadas em abril de 2023 (UNESCO, 2023b).⁴⁷ A revisão sugere uma lista de sete passos que as agências governamentais podem dar para regular a IA generativa e reafirmar o controle público, a fim de aproveitar seu potencial em diversos setores, incluindo a educação.

Passo 1: Endossar regulamentações gerais de proteção de dados internacionais ou regionais ou desenvolver regulamentações nacionais.

O treinamento de modelos de IAGen envolve a coleta e o processamento de dados *online* de cidadãos em vários países. O uso, pelos modelos de IAGen, de dados e conteúdos sem consentimento está evidenciando ainda mais os problemas relacionados à proteção de dados.

Regulamentos gerais de proteção de dados, dos quais o GDPR da UE, promulgado em 2018, é um dos exemplos precursores, fornecem o marco legal necessário para regular a coleta e o processamento de dados pessoais pelos desenvolvedores de IAGen. De acordo com o portal *Data Protection and Privacy Legislation Worldwide* da Conferência das Nações Unidas sobre Comércio e Desenvolvimento (UNCTAD), 137 de 194 países estabeleceram legislação voltada a assegurar a proteção de dados e a privacidade.⁴⁸

No entanto, ainda é incerto dizer em que medida os *frameworks* estão sendo implementados nesses países. Por isso, é cada vez mais importante garantir que sejam devidamente implementados, o que inclui o monitoramento regular das operações dos sistemas de IAGen. Também é urgente que países que ainda não possuem leis gerais de proteção de dados as desenvolvam.

Passo 2: Adotar/revisar e financiar estratégias sobre IA abrangentes para todo o setor governamental.

A regulamentação da IA generativa deve ser parte integrante de estratégias nacionais mais amplas de IA que garantam o uso seguro e equitativo da IA em diversos setores, incluindo a educação. A formulação, a aprovação, o financiamento e a implementação de estratégias nacionais de IA exigem uma abordagem que envolva todo setor governamental. Somente uma abordagem com essas características pode garantir a coordenação de ações intersetoriais necessárias para abordar de forma integrada os desafios emergentes.

Até o início de 2023, cerca de 67 países⁴⁹ haviam desenvolvido ou planejado estratégias nacionais sobre IA, das quais 61 eram estratégias independentes de IA, enquanto 7 eram capítulos sobre IA integrados em estratégias nacionais mais amplas sobre TIC ou digitalização. Compreensivelmente, dado o seu caráter inédito, nenhuma dessas estratégias nacionais havia abordado a IA generativa como uma questão específica até a data em que o presente informe estava sendo redigido.

É fundamental que os países revisem suas estratégias nacionais existentes de IA ou as desenvolvam, garantindo disposições para regular o uso ético da IA em diversos setores, incluindo a educação.

Passo 3: Consolidar e implementar regulamentações específicas sobre a ética da IA.

Para abordar as dimensões éticas apresentadas pelo uso da IA, são necessárias regulamentações específicas.

A revisão das estratégias nacionais existentes de IA realizada pela UNESCO em 2023 indica que a identificação de tais questões éticas e a formulação de princípios orientadores são comuns a aproximadamente 40 estratégias nacionais de IA.⁵⁰ E mesmo nesses casos, os princípios éticos precisariam ser traduzidos em leis ou regulamentos aplicáveis, o que raramente acontece. Na verdade, apenas cerca de 20 países definiram regulamentações claras sobre a ética da IA, incluindo no contexto educacional, seja como parte das estratégias nacionais de IA ou de outras formas. Curiosamente, enquanto a educação é destacada como um domínio político em cerca de 45 estratégias nacionais de IA,⁵¹ as referências à educação são articuladas mais em termos de habilidades em IA e desenvolvimento de talentos necessários para apoiar a competitividade nacional, e menos em termos de questões éticas.

Países que ainda não possuem regulamentações sobre ética em IA devem formulá-las e implementá-las urgentemente.

Passo 4: Ajustar ou garantir o cumprimento das leis de direitos autorais existentes para regular o conteúdo gerado por IA.

O uso cada vez mais difundido de IAGen trouxe novos desafios relacionados à questão dos direitos autorais, tanto no que diz respeito aos conteúdos ou obras protegidos por direitos autorais sendo usados para treinar modelos, quanto ao *status* dos resultados de conhecimento “não humanos” que produzem.

Até o momento, apenas a China, os países da União Europeia e os Estados Unidos adaptaram as suas leis sobre direitos autorais para que elas contemplem as implicações da IA generativa. O Escritório de Direitos Autorais dos EUA, por exemplo, determinou que os resultados de sistemas de IAGen, como o ChatGPT, não são protegíveis sob a lei de direitos autorais dos EUA, argumentando que “os direitos autorais podem proteger apenas o material que é produto da criatividade humana” (Escritório de Direitos Autorais

dos EUA, 2023). Enquanto isso, na União Europeia, a Lei de IA proposta exige que os desenvolvedores de ferramentas de IA divulguem os materiais protegidos por direitos autorais que utilizaram na construção de seus sistemas (Comissão Europeia, 2021). A China, por meio de sua regulamentação sobre a IAGen divulgada em julho de 2023, exige a rotulagem dos resultados da IAGen como conteúdo gerado por IA, reconhecendo-os apenas como resultados de síntese digital.

Regulamentar o uso de materiais protegidos por direitos autorais no treinamento de modelos de IAGen e definir o *status* de direitos autorais dos resultados da IAGen estão se tornando novas responsabilidades das leis de direitos autorais. É urgente que as leis existentes sejam adaptadas para considerar essa questão.

Passo 5: Elaborar marcos regulatórios sobre a IA generativa.

O rápido desenvolvimento das tecnologias de IA está forçando as agências de governança nacionais/locais a acelerar a atualização de suas regulamentações. Até julho de 2023, apenas um país, a China, havia divulgado regulamentações específicas sobre a IAGen. As “Regulamentações Provisórias sobre a Prestação de Serviços de IA Generativa”, divulgadas em 13 de julho de 2023 (Administração do Ciberespaço da China, 2023a), exigem que os fornecedores de sistemas de IAGen rotulem conteúdos, imagens e vídeos gerados por IA de maneira apropriada e legal, em conformidade com o seu atual *Regulamento sobre Síntese Profunda no Âmbito dos Serviços de Informação Online*. É preciso desenvolver mais marcos nacionais voltados à IAGen com base em uma avaliação das lacunas presentes nas regulamentações e leis locais vigentes.

Passo 6: Desenvolver capacidades para o uso adequado de IAGen na educação e na pesquisa.

Escolas e outras instituições educacionais devem desenvolver habilidades para compreender os possíveis benefícios e riscos da IA, incluindo a IAGen, para a educação. Somente com base nessa compreensão será possível validar a adoção de ferramentas de IA. Além disso, é preciso apoiar professores e pesquisadores no fortalecimento de suas habilidades no uso adequado da IAGen, incluindo treinamento e orientação contínua. Alguns países lançaram programas de desenvolvimento de capacidades nesse sentido, incluindo Cingapura, que oferece uma plataforma dedicada ao desenvolvimento de capacidades de IA em instituições educacionais por meio de seu *Cluster Governamental de IA na Nuvem*, que inclui um repositório exclusivo de modelos GPT (Ocampo, 2023).

Passo 7: Refletir sobre as implicações de longo prazo da IAGen para a educação e a pesquisa.

O impacto das versões atuais da IAGen está apenas começando a se desdobrar, e seus efeitos na educação ainda estão sendo plenamente explorados e compreendidos. Enquanto isso, versões mais avançadas da IAGen e outras classes de IA continuam a ser desenvolvidas e implementadas. No entanto, continuam surgindo questões cruciais sobre as implicações da IAGen para a criação, transmissão e validação do conhecimento - para o ensino e aprendizado, para o *design* e a avaliação de currículos e para a pesquisa e os direitos autorais. A maioria dos países está no estágio inicial de adoção da IAGen na educação, mesmo que os impactos de longo prazo ainda não tenham sido totalmente compreendidos. Para garantir um uso da IA centrado no ser humano, se torna urgente conduzir debates públicos abertos e diálogos no âmbito político sobre as implicações de longo prazo. Debates inclusivos envolvendo governo, setor privado e outros parceiros devem servir para reunir ideias e contribuições para a atualização progressiva de regulamentações e políticas.

3.3 Regulamentações sobre a IAGen: elementos-chave

É crucial que todos os países estabeleçam regulamentações adequadas para a IAGen, garantindo que ela promova o desenvolvimento na educação e em outros contextos. Esta seção propõe ações em torno de elementos-chave que podem ser adotadas por: (1) agências reguladoras governamentais, (2) desenvolvedores de ferramentas de IA, (3) usuários institucionais e (4) usuários individuais. Embora vários elementos dos marcos regulatórios sejam de natureza transnacional, todos eles devem considerar também o contexto local, ou seja, os sistemas educacionais específicos do país e os marcos regulatórios gerais em vigor.

3.3.1 Agências governamentais reguladoras

Para coordenar o *design*, o alinhamento e a implementação de regulamentações sobre a IAGen, é necessário adotar uma abordagem governamental integrada. São recomendados sete elementos e ações chave:

- **Coordenação interseccional:** estabelecer um órgão nacional para liderar a integração governamental em relação à IAGen e coordenar a cooperação entre os setores.

- **Alinhamento da legislação:** alinhar o marco aos contextos legislativos e regulatórios pertinentes de cada país – incluindo, por exemplo, leis gerais de proteção de dados, regulamentações sobre segurança na Internet, leis sobre a segurança de dados produzidos a partir dos cidadãos ou usados para servi-los, bem como outras legislações e práticas habituais relevantes; e avaliar a adequação das regulamentações existentes e realizar as adaptações necessárias em resposta a novas questões levantadas pela IAGen.
- **Equilíbrio entre a regulamentação da IAGen e a promoção da inovação em IA:** promover a cooperação interseccional entre empresas, organizações, instituições de ensino e pesquisa, bem como órgãos públicos relevantes, para desenvolver conjuntamente modelos confiáveis; incentivar a construção de ecossistemas de código aberto para promover o compartilhamento de recursos de supercomputação e conjuntos de dados de pré-treinamento de alta qualidade; e estimular a aplicação prática da IAGen em diversos setores e a criação de conteúdo de alta qualidade visando o bem público.
- **Avaliação e classificação dos riscos potenciais da IA:** estabelecer princípios e um processo para a avaliação e categorização da eficácia, segurança e proteção dos serviços da IAGen antes de sua implementação e ao longo do ciclo de vida do sistema; considerar mecanismos de categorização com base nos níveis de risco que a IAGen pode implicar para os cidadãos; e classificá-los sob regulamentações rigorosas (ou seja, proibindo aplicativos ou sistemas habilitados por IA com riscos inaceitáveis), regulamentações específicas para aplicações de alto risco e regulamentações gerais para aplicações que não são consideradas como de alto risco. Ver o projeto de Lei de IA da UE como exemplo dessa abordagem.
- **Proteção da privacidade de dados:** considerar o fato de que o uso da IAGen quase sempre implica o compartilhamento de dados de usuários com os desenvolvedores da IAGen; exigir a elaboração e implementação de leis para a proteção das informações pessoais dos usuários; e identificar e combater as práticas ilegais de armazenamento, perfilagem e compartilhamento de dados.

- **Definição e aplicação de restrição de idade para o uso da IAGen:** a maioria das aplicações de IAGen é projetada com foco principal em usuários adultos. Essas aplicações frequentemente apresentam riscos significativos para crianças, incluindo a exposição a conteúdo inadequado, bem como o potencial de manipulação. Diante desses riscos e dada a considerável incerteza que continua a envolver aplicações iterativas de IAGen, é altamente recomendável impor restrições de idade para tecnologias de IA de uso geral, visando proteger os direitos e o bem-estar das crianças.

Atualmente, os termos de uso do ChatGPT exigem que os usuários tenham pelo menos 13 anos, e os usuários menores de 18 anos devem ter a permissão de um dos pais ou do responsável legal para utilizar os serviços.⁵² Essas restrições de idade são derivadas da Lei de Proteção à Privacidade Online de Crianças dos Estados Unidos da América (Comissão Federal de Comércio, 1998). Aprovada em 1998, antes do uso generalizado das redes sociais e muito antes da criação de aplicativos de IAGen fáceis de usar e potentes como o ChatGPT, a lei dos EUA especifica que organizações ou provedores individuais de redes sociais não estão autorizados a fornecer serviços para crianças menores de 13 anos sem a permissão dos pais. Muitos analistas consideram essa restrição de idade muito baixa e defendem que a legislação a eleve para 16 anos. O Regulamento Geral sobre a Proteção de Dados (GDPR) da União Europeia (2016) especifica que os usuários devem ter pelo menos 16 anos para utilizar os serviços de redes sociais sem a permissão dos pais.

O surgimento de tantos *chatbots* de IAGen exige que os países considerem de maneira cuidadosa – e deliberem publicamente – a restrição de idade apropriada para manter conversas independentes com plataformas de IAGen. O limite mínimo deve ser de 13 anos. Os países também precisarão analisar se a autodeclaração de idade continua sendo um meio apropriado de verificação. Também deverão impor aos provedores de IAGen a responsabilidade pela verificação de idade, e aos pais ou responsáveis, de monitorar as conversas independentes no caso de crianças menores de idade.

- **Propriedade nacional de dados e o risco de pobreza de dados:** adotar medidas legislativas para proteger a propriedade nacional de dados e regular os provedores de IAGen que operam dentro de suas fronteiras. Para conjuntos de dados gerados por cidadãos e utilizados para fins comerciais, estabelecer regulamentações para promover a cooperação mutuamente benéfica, de modo a evitar que essa categoria de dados seja levada para fora do país para ser explorada exclusivamente por grandes empresas de tecnologia.

3.3.2 Provedores de ferramentas de IAGen

Os provedores de IAGen incluem organizações e indivíduos responsáveis pelo desenvolvimento e disponibilização de ferramentas de IAGen, e/ou que utilizam tecnologias de IAGen para fornecer serviços, incluindo por meio de interfaces de programação de aplicações (APIs) programáveis. A maioria dos provedores influentes de ferramentas de IAGen são empresas com sólidos recursos financeiros. Deve ficar claro para os provedores de IAGen que são eles os responsáveis pelas questões éticas desde a sua concepção, inclusive pela implementação dos princípios éticos estipulados nas regulamentações. Devem ser cumpridas as dez categorias de responsabilidade listadas a seguir:

- **Responsabilidades humanas:** os provedores de IAGen devem ser responsáveis por garantir a adesão a valores fundamentais e propósitos legítimos, respeitando a propriedade intelectual e preservando as práticas éticas, ao mesmo tempo em que devem evitar a disseminação de desinformação e discursos de ódio.
- **Dados e modelos confiáveis:** deve-se exigir que os provedores de IAGen comprovem a confiabilidade e ética das fontes de dados, assim como dos métodos utilizados por seus modelos e para gerar resultados. Eles devem ser instruídos a adotar dados e modelos fundamentais com fontes legais comprovadas e a obedecer às leis de propriedade intelectual correspondentes (por exemplo, caso os dados estiverem protegidos por direitos de propriedade intelectual). Além disso, quando os modelos precisarem utilizar informações pessoais, a coleta dessas informações deve ocorrer apenas com o consentimento informado e explícito dos proprietários.
- **Geração de conteúdo não discriminatório:** os provedores de IAGen devem proibir a criação e a implementação de sistemas de IAGen que gerem conteúdo tendencioso ou discriminatório com base em raça, nacionalidade, gênero ou outras características protegidas. Eles devem garantir a existência de “guarda-corpos” (*guardrails*) robustos para evitar que a IAGen produza conteúdo ofensivo, tendencioso ou falso, bem como assegurar que as pessoas envolvidas nas informações resguardadas pelos “guarda-corpos” estejam protegidas e não sejam exploradas.
- **Explicabilidade e transparência dos modelos de IAGen:** os provedores devem submeter às agências públicas de governança explicações sobre as fontes, a escala e os tipos de dados utilizados pelos modelos; suas regras para rotular os dados no pré-treinamento; os métodos ou algoritmos que seus modelos utilizam para gerar conteúdos ou respostas; e os serviços fornecidos por suas ferramentas de IAGen. Quando necessário, eles devem oferecer suporte para que as agências de governança possam entender a tecnologia e os dados. Deve ser transparente aos usuários a propensão com que a IAGen pode gerar conteúdos com erros e respostas contestáveis.
- **Rotulagem do conteúdo de IAGen:** de acordo com as leis ou regulamentações pertinentes sobre a síntese de informações *online* assistida por IA, os provedores precisam rotular os trabalhos, relatórios, imagens e vídeos gerados pela IAGen de forma adequada e legal. Por exemplo, o resultado da IAGen deve ser claramente rotulado como sendo produzido por uma máquina.
- **Princípios de segurança e proteção:** os provedores de IAGen devem garantir um serviço seguro, robusto e sustentável ao longo de todo o ciclo de vida de um sistema de IAGen.
- **Especificações sobre adequação para acesso e uso:** os provedores de IAGen devem fornecer especificações claras sobre o público apropriado, cenários de uso e propósitos de seus serviços, ajudando os usuários das ferramentas de IAGen a tomar decisões racionais e responsáveis.

- **Reconhecimento das limitações e prevenção de riscos previsíveis:** os provedores de IAGen devem divulgar de forma clara as limitações dos métodos utilizados pelos sistemas e seus resultados. Eles precisam desenvolver tecnologias para garantir que os dados de entrada, métodos e resultados não causem danos previsíveis aos usuários, juntamente com protocolos para mitigar danos imprevisíveis quando estes ocorrem. Eles também devem fornecer orientações para ajudar os usuários a entender o conteúdo gerado pela IAGen com base em princípios éticos e prevenir a dependência excessiva e o vício nos conteúdos gerados.
- **Mecanismos para reclamações e soluções:** os provedores de IAGen precisam estabelecer mecanismos e canais para a coleta de reclamações de usuários e do público em geral, tomando ações oportunas para aceitar e processar essas reclamações.
- **Monitoramento e denúncia de uso ilegal:** os provedores devem cooperar com organismos públicos de governança para facilitar o monitoramento e a denúncia de usos ilegais. Isso inclui casos em que pessoas utilizam produtos de IAGen de maneiras ilegais ou violam valores éticos ou sociais, como promover a desinformação ou o discurso de ódio, gerar *spam* ou criar *malware*.

3.3.3 Usuários institucionais

Usuários institucionais incluem autoridades educacionais e instituições como universidades e escolas, que têm a responsabilidade de decidir se a IAGen deve ser adotada e quais tipos de ferramentas de IAGen devem ser adquiridas e implementadas dentro da instituição.

- **Auditoria institucional de algoritmos, dados e resultados da IAGen:** implementar mecanismos para monitorar da melhor forma possível os algoritmos e dados utilizados pelas ferramentas de IAGen e os resultados que elas geram. Isso inclui auditorias e avaliações regulares, proteção de dados dos usuários e filtragem automática de conteúdo inadequado.
- **Validação da proporcionalidade e proteção do bem-estar dos usuários:** implementar mecanismos de classificação nacionais ou criar uma política institucional para categorizar

e validar sistemas e aplicativos de IAGen. Garantir que os sistemas de IAGen adotados pela instituição estejam alinhados com marcos éticos validados localmente e não causem danos previsíveis aos usuários-alvo da instituição, especialmente crianças e grupos vulneráveis.

- **Revisão e abordagem de impactos de longo prazo:** com o tempo, depender de ferramentas ou conteúdos de IAGen na educação pode ter efeitos profundos no desenvolvimento de capacidades humanas, como nas habilidades relacionadas ao pensamento crítico e à criatividade. Esses possíveis efeitos devem ser avaliados e tratados.
- **Adequação à idade:** considerar a implementação de restrições mínimas de idade para o uso independente de IAGen na instituição.

3.3.4 Usuários individuais

Usuários individuais incluem potencialmente a todas as pessoas do mundo que têm acesso à Internet e a pelo menos um tipo de ferramenta de IAGen. O termo “usuários individuais”, utilizado aqui, refere-se principalmente a professores, pesquisadores e estudantes individuais em instituições educacionais formais ou aqueles que participam de programas de estudo não formais.

- **Conhecimento dos termos de referência sobre o uso de IAGen:** ao assinar ou expressar consentimento com os acordos de serviço, os usuários devem estar cientes da obrigação de cumprir com os termos de referência estipulados no acordo e nas leis ou regulamentações que o respaldam.
- **Uso ético de aplicações de IAGen:** os usuários devem utilizar a IAGen de forma responsável e evitar explorá-la de maneiras que possam prejudicar a reputação e os direitos legítimos de outras pessoas.
- **Monitoramento e denúncia de aplicações ilegais de IAGen:** ao descobrir aplicações de IAGen que violam uma ou mais regulamentações, os usuários devem notificar as agências governamentais reguladoras.

4. Buscar um marco referencial político para o uso de IAGen na educação e na pesquisa

Regulamentar a IAGen para aproveitar os seus benefícios potenciais para a educação e a pesquisa requer o desenvolvimento de políticas adequadas. Os dados da pesquisa de 2023 citada anteriormente indicam que apenas alguns países adotaram políticas ou planos específicos para o uso de IA na educação. A seção anterior apresentou uma visão, os passos necessários e os elementos e ações chave que podem ser adotados por vários interessados. Esta seção apresenta medidas que podem ser aplicadas no desenvolvimento de marcos políticos coerentes e abrangentes para regulamentar o uso de IAGen na educação e na pesquisa.

Um ponto de partida nesse sentido é o guia “IA e Educação: Orientações para os Formuladores de Políticas (*AI and education: guidance for policy-makers*)” (UNESCO, 2022b). Ele propõe um conjunto abrangente de recomendações para orientar os governos no desenvolvimento e implementação de políticas setoriais sobre IA e educação, com foco na promoção da educação de qualidade, equidade social e inclusão. A maioria das recomendações permanece aplicável até hoje e pode ser adaptada para orientar a formulação de políticas específicas sobre a IAGen na educação. A seguir são propostas oito medidas específicas para o planejamento de políticas sobre a IAGen na educação e na pesquisa, complementando a orientação existente.

4.1 Promover a inclusão, a equidade e a diversidade linguística e cultural

A urgente importância da inclusão deve ser reconhecida e abordada ao longo do ciclo de vida da IAGen. Mais especificamente, as ferramentas de IAGen não serão capazes de ajudar a enfrentar os desafios fundamentais da educação ou o cumprimento dos compromissos do ODS 4, a menos que elas sejam acessíveis em termos de inclusão (independentemente de gênero, etnia, necessidades educacionais especiais, *status* socioeconômico, localização geográfica, situação de deslocamento, etc.) e se promoverem, desde a sua concepção, a equidade, a diversidade linguística e o pluralismo cultural. Para alcançar isso, a seguir são recomendadas três medidas políticas:

- Identificar aqueles que não têm ou não podem arcar com a conexão à Internet ou o acesso a dados e tomar medidas para promover a conectividade universal e as competências digitais, a fim de reduzir as barreiras ao acesso equitativo e inclusivo a aplicações de IA. Estabelecer mecanismos de financiamento sustentáveis para o desenvolvimento e fornecimento de ferramentas habilitadas de IA para alunos com deficiência ou necessidades especiais. Promover o uso de IAGen para apoiar, ao longo de toda a vida, estudantes de todas as idades, localidades e origens.
- Desenvolver critérios para a validação de sistemas de IAGen para garantir que não haja viés de gênero, discriminação contra grupos marginalizados ou discurso de ódio incorporados em dados ou algoritmos.
- Desenvolver e implementar especificações inclusivas para sistemas de IAGen e adotar medidas institucionais para proteger a diversidade linguística e cultural ao implementar IAGen em grande escala na educação e na pesquisa. As especificações devem exigir que os provedores de IAGen incluam, no treinamento dos modelos, dados em vários idiomas, especialmente em idiomas locais ou indígenas, aprimorando a capacidade da IAGen de responder e gerar texto multilíngue. As especificações e medidas institucionais devem proibir integralmente os provedores de IAGen de realizar qualquer remoção de idiomas minoritários, de forma intencionada ou não, ou discriminar falantes de línguas indígenas. Também devem exigir que os provedores impeçam os sistemas de promover apenas idiomas ou normas culturais dominantes.

4.2 Proteger a agência humana

À medida que a IAGen se torna cada vez mais sofisticada, um dos principais riscos relacionados diz respeito ao seu potencial de minar a agência humana. Conforme mais indivíduos utilizam a IAGen para escrever ou para outras atividades criativas,

é possível que involuntariamente passem a depender dela. Isso pode comprometer o desenvolvimento de habilidades intelectuais. Embora a IAGen possa ser usada para desafiar e ampliar o pensamento humano, não deve permitir-se que ela o usurpe. A proteção e o aprimoramento da agência humana devem sempre ser considerações centrais ao projetar e adotar a IAGen, sob as seguintes sete perspectivas:

- Informar aos estudantes a respeito dos tipos de dados sobre eles que a IAGen pode coletar, como esses dados são utilizados e o impacto que isso pode ter em sua educação e em suas vidas em geral.
- Proteger a motivação intrínseca dos estudantes para crescerem e aprenderem como indivíduos. Reforçar a autonomia humana em relação aos seus próprios enfoques de pesquisa, ensino e aprendizagem em um contexto de uso de sistemas de IAGen cada vez mais sofisticados.
- Evitar o uso da IAGen quando este privar os estudantes de oportunidades para desenvolver habilidades cognitivas e sociais por meio de observações do mundo real, práticas empíricas como experimentos, discussões com outros humanos e raciocínio lógico independente.
- Garantir interação social suficiente e exposição adequada à produção criativa feita por humanos e evitar que os estudantes se tornem viciados ou dependentes da IAGen.
- Utilizar ferramentas de IAGen para minimizar a pressão de tarefas e provas, em vez de agravá-la.
- Consultar a opinião de pesquisadores, professores e estudantes sobre a IAGen e utilizar esse *feedback* para decidir se determinadas ferramentas de IAGen devem ser implementadas em escala institucional, e de que maneira. Incentivar estudantes, professores e pesquisadores a criticar e questionar as metodologias por trás dos sistemas de IA, a precisão do conteúdo gerado e as normas ou pedagogias que podem ser impostas.
- Evitar transferir a responsabilidade humana aos sistemas de IAGen para a tomada de decisões de alto impacto.

4.3 Monitorar e validar sistemas de IAGen para a educação

Como observado, o desenvolvimento e a implementação da IAGen devem ser éticos desde a sua concepção. Posteriormente, uma vez que a IAGen está em uso e ao longo de seu ciclo de vida, ela precisa ser monitorada e validada de forma cuidadosa quanto aos seus riscos éticos, sua adequação e rigor pedagógico e seu impacto nos estudantes, professores e nas relações em sala de aula/escola. Nesse sentido, recomenda-se as cinco ações a seguir:

- Construir mecanismos de validação para testar se os sistemas de IAGen usados na educação e na pesquisa não possuem vieses, especialmente de gênero, e se são treinados com dados representativos quanto à diversidade (em termos de gênero, deficiência, *status* social e econômico, origem étnica e cultural e localização geográfica).
- Abordar a complexa questão do consentimento informado, especialmente em contextos nos quais crianças ou outros estudantes mais vulneráveis não são capazes de dar um consentimento verdadeiramente informado.
- Verificar se os resultados da IAGen incluem imagens *deepfake*, notícias falsas (imprecisas ou falaciosas) ou discurso de ódio. Se for constatado que a IAGen está gerando conteúdo inadequado, as instituições e educadores devem estar dispostos e aptos a tomar ações rápidas e robustas para mitigar ou eliminar o problema.
- Exercer uma rigorosa validação ética das aplicações de IAGen antes que elas sejam oficialmente adotadas em instituições educacionais ou de pesquisa (ou seja, adotar uma abordagem ética desde a sua concepção).
- Antes de tomar decisões sobre a adoção institucional, garantir que as aplicações de IAGen em questão não causem danos previsíveis aos estudantes, sejam eficazes do ponto de vista educacional, pertinentes considerando as idades e habilidades dos aprendizes-alvo e alinhadas com princípios pedagógicos sólidos (ou seja, baseadas nos domínios relevantes do conhecimento e nos resultados de aprendizagem e desenvolvimento de valores esperados).

4.4 Desenvolver competências em IA, incluindo habilidades relacionadas à IAGen, para os estudantes

O desenvolvimento de competências em IA entre os estudantes é fundamental para o uso seguro, ético e significativo da IA na educação e em outras áreas. No entanto, de acordo com dados da UNESCO, no início de 2022 apenas cerca de 15 países haviam desenvolvido e implementado, ou estavam em processo de desenvolver, currículos de IA oficialmente endossados pelo governo nas escolas (UNESCO, 2022c). Os últimos avanços da IAGen reforçaram ainda mais a necessidade urgente de que todos alcancem um nível apropriado de alfabetização nas dimensões humana e tecnológica da IA, compreendendo seu funcionamento em termos amplos, bem como o impacto específico da IAGen. Para tanto, as cinco ações listadas a seguir são urgentemente necessárias:

- Comprometer-se com a oferta de currículos de IA oficialmente endossados pelo governo para a educação escolar e a educação técnica e profissionalizante, bem como para a aprendizagem ao longo da vida. Os currículos de IA devem abordar o impacto da IA em nossas vidas, incluindo as questões éticas que ela suscita, assim como uma compreensão sobre algoritmos e dados apropriada para cada idade e habilidades para o uso adequado e criativo de ferramentas de IA, incluindo aplicações de IAGen;
- Apoiar instituições de ensino superior e de pesquisa para aprimorar programas de desenvolvimento de habilidades em IA de forma local;
- Promover a igualdade de gênero no desenvolvimento de competências avançadas em IA e criar um grupo de profissionais equilibrado em termos de gênero;
- Desenvolver previsões interseccionais sobre as mudanças nacionais e globais em relação a empregos causadas pela automação mais recente da IAGen e aprimorar habilidades para o futuro em todos os níveis de sistemas educacionais e de aprendizado ao longo da vida, com base em possíveis mudanças na demanda; e
- Oferecer programas especiais para trabalhadores de idades mais avançadas e para cidadãos que talvez precisem aprender novas habilidades e se adaptar a novos ambientes.

4.5 Desenvolver a capacidade de professores e pesquisadores para o uso adequado da IAGen

De acordo com dados de pesquisa de 2023 sobre o uso governamental de IA na educação (UNESCO, 2023c), apenas sete países (China, Finlândia, Geórgia, Catar, Espanha, Tailândia e Turquia) relataram ter desenvolvido ou estarem desenvolvendo estruturas ou programas de treinamento sobre IA para professores. Apenas o Ministério da Educação de Cingapura informou ter construído um repositório *online* centrado no uso do ChatGPT para o ensino e a aprendizagem. Isso evidencia que, na maioria dos países, os professores não têm acesso à capacitação bem estruturada sobre o uso de IA na educação, especialmente sobre o uso de IAGen.

Para preparar os professores para o uso responsável e efetivo da IAGen, os países precisam adotar as quatro medidas a seguir:

- Formular ou ajustar orientações com base em experiências locais para ajudar pesquisadores e professores a explorar as ferramentas de IAGen já amplamente disponíveis e orientar a elaboração de novas aplicações de IA específicas para cada domínio.
- Proteger os direitos dos professores e pesquisadores, bem como o valor de suas práticas ao utilizar a IAGen. Isto é, compreender o papel único dos professores em facilitar o pensamento de ordem superior, organizar a interação humana e promover valores humanos.
- Definir uma orientação a respeito dos valores, conhecimentos e habilidades necessárias para que os professores compreendam e utilizem os sistemas de IAGen de forma eficaz e ética. Capacitar os professores para que possam criar ferramentas específicas baseadas em IAGen para facilitar a aprendizagem na sala de aula e em seu próprio desenvolvimento profissional.
- Revisar, de forma dinâmica, as competências necessárias para que os professores compreendam e utilizem a IA no ensino, na aprendizagem e em seu desenvolvimento profissional, e integrar conjuntos emergentes de valores, compreensão e habilidades em IA nos marcos de competência e programas de capacitação para professores em exercício e em formação.

4.6 Promover opiniões diversas e a expressão plural de ideias

Como mencionado anteriormente, a IAGen não compreende nem os comandos e nem as respostas. Em vez disso, suas respostas são baseadas em probabilidades de padrões de linguagem encontrados em dados (da Internet) que foram incorporados durante o treinamento do seu modelo. Para lidar com alguns dos principais problemas encontrados em seus resultados, novos métodos estão sendo pesquisados, como conectar a IAGen a bancos de dados de conhecimento e motores de raciocínio. No entanto, devido à sua forma de funcionamento, aos seus materiais de origem e às perspectivas tácitas de seus desenvolvedores, a IAGen, por definição, reproduz visões de mundo dominantes em seus resultados e compromete as opiniões minoritárias e plurais. Portanto, para que as civilizações humanas prosperem, é essencial reconhecer que a IAGen não pode ser uma fonte autoritária de conhecimento sobre qualquer tema que ela aborde.

Como consequência, os usuários precisam analisar criticamente os resultados da IAGen. Em particular:

- Compreender o papel da IAGen como uma fonte de informação rápida, mas muitas vezes não confiável. Embora alguns *plugins* e ferramentas baseadas em LLM mencionadas anteriormente sejam projetados para atender à necessidade de acessar informações validadas e atualizadas, há poucas evidências robustas até o momento de que eles são eficazes.
- Incentivar estudantes e pesquisadores a analisar criticamente as respostas fornecidas pela IAGen. Reconhecer que a IAGen geralmente apenas repete opiniões já estabelecidas ou padrão, minando, assim, opiniões diversas e minoritárias, bem como a expressão plural de ideias.
- Oferecer aos estudantes oportunidades suficientes de aprender por meio de tentativas e erros, experimentos empíricos e observações do mundo real.

4.7 Testar modelos de aplicações localmente relevantes e construir uma base cumulativa de evidências

Até o momento, os modelos de IAGen são predominantemente compostos por informações do Norte Global, sub-representando vozes do Sul Global e de comunidades indígenas. Somente por meio de determinados esforços, como o uso de dados sintéticos (Marwala, 2023), as ferramentas de IAGen poderão se tornar sensíveis ao contexto e às necessidades das comunidades locais, especialmente as do Sul Global. Para que seja possível explorar abordagens relevantes para as necessidades locais, colaborando de maneira mais ampla, são recomendadas as seguintes oito ações:

- Garantir que a elaboração e a adoção da IAGen sejam planejadas de forma estratégica, em vez de facilitar um processo de implementação de caráter passivo e não crítico.
- Incentivar os criadores de IAGen a focar em opções de aprendizagem abertas, exploratórias e diversas.
- Testar e ampliar casos de uso de aplicações de IA baseados em evidência na educação e na pesquisa de acordo com as prioridades educacionais, e não apenas por serem novidade, mito ou popularizadas.
- Orientar o uso da IAGen para incentivar a inovação na pesquisa, inclusive aproveitando as capacidades computacionais, os grandes conjuntos de dados e os resultados da IAGen para informar e inspirar a melhoria das metodologias de pesquisa.
- Rever as implicações sociais e éticas de incorporar a IAGen nos processos de pesquisa.
- Estabelecer critérios específicos com base em pesquisas e metodologias pedagógicas comprovadas e construir uma base de evidências sobre a eficácia da IAGen no apoio à oferta de oportunidades de aprendizado inclusivas, ao alcance de objetivos de aprendizado e pesquisa e à promoção da diversidade linguística e cultural.
- Adotar medidas progressivas para fortalecer as evidências sobre o impacto social e ético da IAGen.
- Analisar os custos ambientais do uso em larga escala de tecnologias de IA (por exemplo, a energia e os recursos necessários

para treinar modelos GPT) e desenvolver metas sustentáveis a serem atingidas pelos provedores de IA, a fim de evitar contribuir com as mudanças climáticas.

4.8 Analisar as implicações de longo prazo de maneira intersetorial e interdisciplinar

Abordagens intersetoriais e interdisciplinares são essenciais para o uso eficaz e ético da IAGen na educação e na pesquisa. Somente ao contar com uma variedade de conhecimentos e reunir múltiplos interessados será possível identificar os desafios-chave de maneira rápida e abordá-los de maneira eficaz, minimizando as implicações negativas de longo prazo e aproveitando os benefícios contínuos e cumulativos. Portanto, são recomendadas as três ações a seguir:

- Colaborar com provedores de IA, educadores, pesquisadores e representantes de pais e estudantes para planejar ajustes em todo o sistema de currículos e metodologias de avaliação, a fim de aproveitar plenamente o potencial e mitigar os riscos da IAGen para a educação e a pesquisa.
- Reunir especialistas interdisciplinares e intersetoriais, incluindo educadores, pesquisadores, cientistas da aprendizagem, engenheiros de IA e representantes de outras partes interessadas, para analisar as implicações de longo prazo da IAGen para a aprendizagem e a produção de conhecimento; a pesquisa e os direitos autorais; os currículos e a avaliação; e a colaboração humana e as dinâmicas sociais.
- Fornecer orientações oportunas para informar as atualizações progressivas de regulamentações e políticas.

5. Facilitar o uso criativo da IAGen na educação e na pesquisa

Quando o ChatGPT foi lançado pela primeira vez, educadores em todo o mundo expressaram preocupação sobre seu potencial para produzir trabalhos escolares e sobre como isso poderia ajudar os estudantes a trapacear. Mais recentemente, diversas pessoas e organizações, incluindo algumas das principais universidades do mundo, têm argumentado que “o gênio está fora da garrafa”, e ferramentas como o ChatGPT estão aqui para ficar e podem ser usadas de forma produtiva em ambientes educacionais. Enquanto isso, na Internet, há várias sugestões para o uso da IAGen na educação e na pesquisa. Isso inclui usá-la para inspirar novas ideias, gerar exemplos com múltiplas perspectivas, desenvolver planos de aula e apresentações, resumir materiais existentes e estimular a criação de imagens. Embora novas ideias apareçam na Internet quase todos os dias, pesquisadores e educadores ainda estão descobrindo exatamente a implicação da IAGen para o ensino, o aprendizado e a pesquisa. Em particular, as pessoas que estão por trás de muitos dos usos propostos podem não ter considerado adequadamente os princípios éticos, enquanto outros são impulsionados pelos potenciais tecnológicos da IAGen, em vez das necessidades de pesquisadores, professores ou alunos. Esta seção destaca maneiras de facilitar o uso criativo da IAGen na educação e na pesquisa.

5.1 Estratégias institucionais para facilitar o uso responsável e criativo da IAGen

Conforme mencionado anteriormente, instituições educacionais e de pesquisa devem desenvolver, implementar e validar estratégias e marcos éticos adequados para orientar o uso responsável e ético de sistemas e aplicativos de IAGen, atendendo às necessidades de ensino, aprendizagem e pesquisa. Isso pode ser alcançado por meio das quatro estratégias a seguir:

- **Orientação e capacitação:** fornecer orientação e capacitação para pesquisadores, professores e alunos sobre ferramentas de IAGen, garantindo que compreendam as questões éticas, como vieses na rotulagem de dados e nos algoritmos, e que cumpram as regulamentações pertinentes de privacidade de dados e propriedade intelectual.
- **Desenvolvimento de capacidades de engenharia de *prompts* de IAGen:** além do conhecimento específico de cada disciplina, pesquisadores e professores também precisarão ter capacidades em engenharia e avaliação crítica de *prompts* gerados por IAGen. Dado que os desafios apresentados pela IAGen são complexos, é essencial que pesquisadores e professores recebam capacitação e apoio de alta qualidade para isso.
- **Detectação de plágio por meio de IAGen em trabalhos escritos:** a IAGen pode permitir que estudantes apresentem textos que não escreveram como de sua própria autoria, um novo tipo de plágio. Os provedores de IAGen devem rotular seus resultados com marcas d'água com os dizeres “gerado por IA”, enquanto ferramentas ainda estão sendo desenvolvidas para identificar materiais produzidos por IA. No entanto, há poucas evidências de que essas medidas ou ferramentas sejam eficazes. A estratégia institucional a ser adotada de forma imediata é manter a integridade acadêmica e reforçar a responsabilidade por meio de detecção rigorosa realizada por humanos. Já na estratégia de longo prazo, as instituições e os educadores precisam repensar a elaboração de tarefas, de modo que não sejam usadas como forma de avaliação daquelas que as ferramentas de IAGen podem realizar melhor do que os estudantes. Em vez disso, devem abordar aquilo que os humanos podem fazer e que as ferramentas de IAGen e outras ferramentas de IA não podem, incluindo aplicar valores humanos, como compaixão e criatividade, a desafios complexos do mundo real.
- **Implementação institucional de princípios éticos:** garantir que pesquisadores, professores e alunos utilizem ferramentas de IAGen de maneira responsável e ética, lidando de forma crítica com a precisão e a validade dos resultados gerados.

5.2 Uma abordagem de “interação centrada no ser humano e pedagogicamente adequada”

Pesquisadores e educadores devem priorizar a agência humana e a interação responsável e pedagogicamente adequada entre humanos e ferramentas de IA ao decidir se vão usar a IAGen e de que forma. Isso inclui as cinco considerações a seguir:

- o uso da(s) ferramenta(s) deve contribuir para atender às necessidades humanas e tornar a aprendizagem ou a pesquisa mais eficaz do que com uma abordagem sem tecnologia ou outra alternativa;
- o uso da(s) ferramenta(s) por educadores e estudantes deve responder à sua motivação intrínseca;
- o processo de uso da(s) ferramenta(s) deve ser controlado pelos educadores, estudantes ou pesquisadores humanos;
- a escolha e organização da(s) ferramenta(s) e do conteúdo gerado devem ser proporcionais, baseadas na faixa etária dos estudantes, nos resultados esperados e no tipo de conhecimento que se espera desenvolver (por exemplo, factual, conceitual, procedimental ou metacognitivo) ou problema-alvo (por exemplo, bem estruturado ou mal estruturado); e
- os processos de utilização devem garantir o envolvimento interativo dos humanos com a IAGen e o pensamento de ordem superior, bem como a responsabilidade humana pelas decisões relacionadas à veracidade do conteúdo gerado por IA, as estratégias de ensino ou pesquisa e o seu impacto nos comportamentos humanos.

5.3 Co-projetar o uso da IAGen na educação e na pesquisa

O uso da IAGen na educação e na pesquisa não deve ser imposto por uma abordagem “de cima para baixo”, nem impulsionado por exageros comerciais. Em vez disso, seu uso seguro e eficaz deve ser co-projetado por professores, estudantes e pesquisadores. Também é necessário um processo robusto de teste e avaliação para examinar a eficácia e o impacto de longo prazo de diferentes usos.

Para facilitar a co-elaboração recomendada, este Guia propõe uma estrutura composta por seis perspectivas para consolidar interações pedagogicamente apropriadas e priorizar a agência humana:

- domínios de conhecimento ou problemas apropriados;
- resultados esperados;
- ferramentas apropriadas de IAGen e avanços comparativos;
- requisitos para usuários;
- métodos pedagógicos humanos necessários e exemplos de *prompts*; e
- riscos éticos.

Esta seção fornece exemplos de como um processo de co-elaboração no uso de IAGen pode informar práticas de pesquisa, auxiliar no ensino, fornecer orientação para a aquisição autodirigida de habilidades fundamentais, facilitar o pensamento de ordem superior e apoiar estudantes com necessidades especiais. Esses exemplos representam apenas a ponta do *iceberg* do crescente número de domínios nos quais o IAGen tem potencial.

5.3.1 IA generativa para a pesquisa

Os modelos de IAGen têm demonstrado potencial para expandir visões sobre esboços de pesquisa e enriquecer a exploração de dados, bem como a revisão de literatura (ver a **Tabela 3**). Embora uma variedade mais ampla de casos de uso possa surgir, é necessário realizar novas pesquisas para definir o alcance potencial de problemas de pesquisa e resultados esperados, demonstrar a eficácia e a precisão e garantir que a agência humana na compreensão do mundo real por meio da pesquisa não seja prejudicada pelo uso de ferramentas de IA.

Tabela 3. Co-projetar o uso da IAGen para a pesquisa

Usos potenciais, mas não comprovados	Domínios de conhecimento ou problemas apropriados	Resultados esperados	Ferramentas apropriadas de IAGen e avanços comparativos	Requisitos para usuários	Métodos pedagógicos humanos necessários e exemplos de prompts	Possíveis riscos
Consultor de IA para esboço de pesquisa.	Pode ser útil em domínios de pesquisa bem estruturados.	Desenvolvimento e resposta a perguntas de pesquisa, sugerindo metodologias apropriadas. Potencial de transformação: instrutor individual para planejamento da pesquisa.	A partir da lista apresentada na Seção 1.2, avaliar se as ferramentas de IAGen são localmente acessíveis, de código aberto e se foram rigorosamente testadas ou validadas por autoridades. Considerar, ainda, as vantagens e os desafios de determinada ferramenta de IAGen, garantindo que ela atenda adequadamente às necessidades humanas específicas.	O pesquisador deve ter um entendimento básico do(s) tópico(s). Ele deve desenvolver a habilidade de verificar informações e, especialmente, ser capaz de detectar citações de artigos de pesquisa que não existem.	Ideias básicas para a definição de problemas de pesquisa (por exemplo, público-alvo, temas, contexto), bem como metodologias, resultados esperados e formatos. Exemplo de prompt: <i>Escreva 10 possíveis questões de pesquisa para [tópico x] e classifique-as por importância para [o campo de pesquisa y].</i>	É preciso ter atenção ao alto risco de a IAGen inventar informações (como publicações de pesquisa inexistentes) e de usuários serem tentados a copiar e colar esboços de pesquisa gerados por IA, o que pode reduzir as oportunidades de aprendizado por meio de tentativas e erros para pesquisadores em início de carreira.
Explorador generativo de dados e revisor bibliográfico.	Pode ser útil em domínios de pesquisa pouco estruturados.	Coleta automática de informações, exploração de uma ampla gama de dados, elaboração de rascunhos de revisões de literatura e automação de partes da interpretação de dados. Potencial de transformação: treinadores de IA para exploração de dados e revisões de literatura.	A partir da lista apresentada na Seção 1.2, avaliar se as ferramentas de IAGen são localmente acessíveis, de código aberto e se foram rigorosamente testadas ou validadas por autoridades. Considerar, ainda, as vantagens e os desafios de determinada ferramenta de IAGen, garantindo que ela atenda adequadamente às necessidades humanas específicas.	Os pesquisadores devem ter um conhecimento sólido de metodologias e técnicas para a análise de dados.	Definições progressivas dos problemas; alcance dos dados e fontes bibliográficas; metodologias utilizadas para a exploração dos dados e revisões da literatura; e resultados esperados e seus formatos.	É preciso desconfiar das informações fabricadas pela IAGen, do tratamento impróprio de dados, de possíveis violações de privacidade, perfilagem não autorizadas e vieses de gênero. É preciso estar atento à propagação de normas dominantes e à ameaça às normas alternativas e opiniões plurais.

5.3.2 IA generativa para facilitar o ensino

O uso de plataformas gerais e ferramentas específicas de IAGen para educação deve ser projetado para aprimorar a compreensão dos professores sobre sua disciplina, bem como seu conhecimento em metodologias de ensino, inclusive por meio da criação conjunta, por parte de professores e da IAGen, de planos de aula, conjuntos de disciplinas e cursos, ou de um currículo inteiro. Os assistentes conversacionais de professores assistidos por IAGen ou “gêmeos generativos de assistentes de ensino”,⁵³ que são

pré-treinados com base em dados de professores experientes e bibliotecas, foram testados em algumas instituições educacionais e podem ter um potencial desconhecido, bem como um risco ético inexplorado. Os processos de aplicação prática e as futuras iterações desses modelos ainda precisam ser cuidadosamente auditados por meio do marco recomendado neste Guia e supervisionados por humanos, conforme exemplificado na **Tabela 4**.

Tabela 4. Co-projetar o uso da IAGen para apoiar professores e o ensino

Usos potenciais, mas não comprovados	Domínios de conhecimento ou problemas apropriados	Resultados esperados	Ferramentas apropriadas de IAGen e avanços comparativos	Requisitos para usuários	Métodos pedagógicos humanos necessários e exemplos de prompts	Possíveis riscos
Currículo ou curso co-projetado.	Conhecimento conceitual sobre determinados tópicos de ensino e conhecimento procedimental sobre metodologias de ensino.	Auxiliar no processo de elaboração do currículo e das aulas, incluindo a elaboração ou expansão de pontos de vista sobre as principais áreas do tema em questão e a definição da estrutura do currículo. Também pode ajudar os professores a preparar provas e exames, oferecendo exemplos de perguntas e critérios para avaliação. Potencial de transformação: currículo gerado por IA.	A partir da lista apresentada na Seção 1.2, avaliar se as ferramentas de IAGen são localmente acessíveis, de código aberto e se foram rigorosamente testadas ou validadas por autoridades. Considerar, ainda, as vantagens e os desafios de determinada ferramenta de IAGen, garantindo que ela atenda adequadamente às necessidades humanas específicas.	Os professores precisam compreender e especificar cuidadosamente o que desejam que o currículo, os cursos, as aulas ou as avaliações abranjam e alcancem, se desejam abordar conhecimentos procedimentais ou conceituais e qual teoria de ensino desejam aplicar.	Perguntas à IAGen sobre sugestões de estrutura e exemplos de conhecimento factual sobre temas, sugestões de métodos e processos de ensino para temas ou problemas, ou criação de conjuntos de cursos ou planos de aula com base em temas e formatos. Os criadores humanos de currículos precisam verificar os conhecimentos factuais e avaliar a adequação dos conjuntos de cursos sugeridos.	O risco de a IAGen impor normas e métodos pedagógicos dominantes é elevado. Ela pode inadvertidamente perpetuar práticas excludentes em favor dos grupos que já são ricos em dados e reforçar desigualdades no acesso a oportunidades educacionais relevantes e de alta qualidade, prejudicando grupos com pobreza de dados.
Chatbot generativo como assistente de ensino.	Conhecimento conceitual de vários domínios em problemas bem estruturados.	Fornecer suporte individualizado, responder perguntas e identificar recursos. Potencial de transformação: gêmeos generativos de assistentes de professores.	A partir da lista apresentada na Seção 1.2, avaliar se as ferramentas de IAGen são localmente acessíveis, de código aberto e se foram rigorosamente testadas ou validadas por autoridades. Considerar, ainda, as vantagens e desafios de determinada ferramenta de IAGen, garantindo que ela atenda adequadamente às necessidades humanas específicas.	Ele oferece suporte aos professores, mas direciona-se diretamente aos estudantes, o que exige que eles tenham conhecimentos prévios, capacidades e habilidades metacognitivas suficientes para verificar os resultados da IAGen e perceber caso haja desinformações. Portanto, pode ser mais apropriado para os estudantes de ensino superior.	Exige que os professores compreendam claramente os problemas, monitorem a conversa e ajudem os estudantes a verificar respostas duvidosas fornecidas pela IAGen.	Com base nas capacidades atuais dos modelos de IAGen, as instituições educacionais precisam garantir a supervisão humana das respostas fornecidas pelas ferramentas, estando atentas ao risco de desinformação. Isso também pode limitar o acesso dos estudantes à orientação e ao apoio humanos, prejudicando o desenvolvimento de uma estreita relação entre professor e estudante, o que é especialmente preocupante no caso das crianças.

5.3.3 IA generativa como um treinador 1:1 para a aquisição de habilidades fundamentais a um ritmo pessoal

Embora cada vez mais atenção é dada ao pensamento de ordem superior e à criatividade ao definir os resultados de aprendizagem, não há dúvidas sobre a importância das habilidades básicas no desenvolvimento psicológico e na progressão das competências das crianças. Entre uma ampla gama de capacidades, essas habilidades fundamentais

incluem ouvir, pronunciar e escrever em língua materna ou estrangeira, além de habilidades básicas de matemática, arte e programação. O método pedagógico de “prática intensiva” não deve ser considerado obsoleto; em vez disso, deve ser revitalizado e aprimorado com as tecnologias de IAGen para incentivar a prática individualizada das habilidades básicas pelos estudantes. Se orientadas por princípios éticos e pedagógicos, as ferramentas de IAGen têm o potencial de se tornar instrutoras individuais para essa prática autodirigida, conforme ilustrado na **Tabela 5**.

Tabela 5. Co-projetar o uso da IAGen como uma instrutora individual para a aquisição autodirigida de habilidades básicas em linguagens e artes

Usos potenciais, mas não comprovados	Domínios de conhecimento ou problemas apropriados	Resultados esperados	Ferramentas apropriadas de IAGen e avanços comparativos	Requisitos para usuários	Métodos pedagógicos humanos necessários e exemplos de prompts	Possíveis riscos
Instrutor individual de habilidades linguísticas.	Aprendizado de idiomas, incluindo a prática de conversação.	Engajar os estudantes na prática de conversação para ajudá-los a aprimorar as habilidades para ouvir, falar e escrever, oferecendo <i>feedback</i> , correções e modelos da língua materna ou de línguas estrangeiras. Auxiliar os estudantes a melhorarem suas habilidades de escrita. Potencial de transformação: tutoriais individuais de idiomas no nível iniciante.	A partir da lista apresentada na Seção 1.2, avaliar se as ferramentas de IAGen são localmente acessíveis, de código aberto e se foram rigorosamente testadas ou validadas por autoridades. Considerar, ainda, as vantagens e desafios de determinada ferramenta de IAGen, garantindo que ela atenda adequadamente às necessidades humanas específicas.	Pode ser estabelecida uma idade limite para as conversas independentes, considerando a possibilidade de os sistemas de IAGen gerarem resultados culturalmente insensíveis ou inadequados para a idade. O estudante deve ter uma motivação intrínseca inicial para iniciar uma conversa com um sistema de IA. O estudante deve ser capaz de adotar uma abordagem crítica às sugestões da IAGen e de verificar sua precisão.	Ao usar plataformas de IAGen em geral, os professores podem orientar os estudantes a interagirem com as ferramentas de IAGen e solicitarem <i>feedback</i> para aprimoramento, correção da pronúncia ou exemplos de escrita. Por exemplo: <i>Inicie uma conversa no idioma [x] e me ajude a melhorar de forma contínua.</i> <i>Sugira algumas ideias para me ajudar a escrever sobre [tema x].</i>	É preciso estar atento à linguagem culturalmente insensível ou contextualmente imprecisa, bem como à inadvertida perpetuação de estereótipos ou preconceitos culturais. Sem estratégias pedagógicas adequadas que simulem as motivações intrínsecas dos estudantes, a criatividade e a originalidade das crianças podem acabar sendo limitadas, levando a uma escrita formulista. Isso também pode limitar as oportunidades de interações na vida real, as opiniões e expressões plurais e o pensamento crítico.
Instrutor individual de artes.	Habilidades técnicas em áreas artísticas como música e desenho.	Fornecer sugestões de técnicas de arte (por exemplo, dicas sobre perspectiva e cor) ou de composição musical (por exemplo, melodia e progressão de acordes). Potencial de transformação: professor individual de artes em nível introdutório.	A partir da lista apresentada na Seção 1.2, avaliar se as ferramentas de IAGen são localmente acessíveis, de código aberto e se foram rigorosamente testadas ou validadas por autoridades. Considerar, ainda, as vantagens e desafios de determinada ferramenta de IAGen, garantindo que ela atenda adequadamente às necessidades humanas específicas.	Os estudantes devem ter objetivos iniciais para criar uma arte ou música, uma compreensão fundamental dos elementos-chave do domínio da arte ou da música e habilidades básicas para analisar as obras de arte ou composições musicais.	Professores humanos devem pedir aos estudantes que comparem as técnicas artísticas das ferramentas de IA com suas próprias obras de arte. Professores ou orientadores humanos devem incentivar os estudantes a desenvolverem e aplicarem sua imaginação e criatividade, algo que a IAGen não pode substituir. Exemplo de <i>prompt</i> : <i>Sugira algumas ideias para me inspirar a criar uma imagem sobre [temas/ ideias].</i>	Pode expor as crianças a conteúdos inadequados ou ofensivos, o que pode violar seu direito à proteção e ao bem-estar. As ferramentas de IAGen aumentam o risco de impedir que os estudantes desenvolvam sua imaginação e criatividade.
Instrutor individual de codificação ou aritmética.	Conhecimentos e habilidades conceituais de programação em nível introdutório. Pode também se aplicar à aprendizagem de matemática básica.	Apoiar a aprendizagem autodirigida de conhecimentos e habilidades básicas de codificação, identificar erros no código dos estudantes e fornecer <i>feedback</i> imediato, e adaptar respostas a perguntas. Potencial de transformação: professor de codificação individualizado em nível introdutório.	A partir da lista apresentada na Seção 1.3, avaliar se as ferramentas de IAGen são localmente acessíveis, de código aberto e se foram rigorosamente testadas ou validadas por autoridades. Considerar, ainda, as vantagens e desafios de determinada ferramenta de IAGen, garantindo que ela atenda adequadamente às necessidades humanas específicas.	Encontrar e definir um problema, assim como projetar algoritmos para resolvê-lo, continuando sendo aspectos centrais do aprendizado de codificação e programação. Os estudantes devem ter uma motivação intrínseca para usar a codificação, junto com conhecimentos e habilidades básicos no uso da linguagem de programação.	Professores e orientadores humanos devem ensinar conhecimentos e habilidades básicas, além de inspirar os estudantes a aplicar o pensamento computacional e a programação para resolver problemas, inclusive por meio de codificação colaborativa. Exemplo de <i>prompt</i> : <i>Proponha algumas ideias não usuais de codificação.</i>	A precisão dos <i>feedbacks</i> e das sugestões continua sendo uma questão problemática, já que a IAGen nem sempre estará correta. Há um alto risco de que as ferramentas de IAGen impeçam os estudantes de desenvolverem habilidades de pensamento computacional e a capacidade de encontrar e definir problemas significativos de codificação.

5.3.4 IA generativa para facilitar a aprendizagem baseada em pesquisas ou projetos

Quando não utilizadas com o propósito de facilitar o pensamento de ordem superior ou a criatividade, as ferramentas de IAGen tendem a incentivar a prática de plágio ou a gerar resultados superficiais, como “papagaios estocásticos”. Porém, dado que os modelos

de IAGen têm sido treinados com dados em grande escala, elas têm o potencial de atuar como oponentes em diálogos Socráticos ou como assistentes de pesquisa na aprendizagem baseada em projetos. No entanto, esses potenciais só podem ser aproveitados por meio de processos de desenho instrucional/de aprendizagem que visem estimular o pensamento de ordem superior, como exemplificado na **Tabela 6**.

Tabela 6. Co-projetar o uso de IAGen para facilitar a aprendizagem baseada em pesquisas ou projetos

Usos potenciais, mas não comprovados	Domínios de conhecimento ou problemas apropriados	Resultados esperados	Ferramentas apropriadas de IAGen e avanços comparativos	Requisitos para usuários	Métodos pedagógicos humanos necessários e exemplos de <i>prompts</i>	Possíveis riscos
Oponente Socrático.	Problemas pouco estruturados.	Envolver os estudantes em um diálogo remanescente do questionamento Socrático do conhecimento prévio, levando à descoberta de novos conhecimentos ou à uma compreensão mais profunda. Potencial de transformação: opponente Socrático individual.	A partir da lista apresentada na Seção 1.3, avaliar se as ferramentas de IAGen são localmente acessíveis, de código aberto e se foram rigorosamente testadas e validadas por autoridades. Considerar, ainda, as vantagens e desafios de determinada ferramenta de IAGen, garantindo que ela atenda dequadamente às necessidades humanas específicas.	O estudante deve ter atingido a idade mínima permitida para realizar conversas independentes com ferramentas de IAGen. Ele deve ter conhecimentos e habilidades prévios para verificar se os argumentos e as informações apresentados são precisos.	Professores humanos podem ajudar na preparação de uma lista de perguntas, que sejam cada vez mais profundas, em forma de exemplos para os estudantes adaptarem em seus <i>prompts</i> . Os estudantes também podem iniciar com um <i>prompt</i> mais amplo, como “Inicie um diálogo Socrático para me ajudar a ter uma perspectiva crítica sobre [tópico x]” e, em seguida, ir aprofundando o diálogo por meio de <i>prompts</i> cada vez mais refinados.	As atuais ferramentas de IAGen podem gerar respostas semelhantes ou padronizadas, limitando a exposição dos estudantes a pontos de vista diversos e perspectivas alternativas, levando a um efeito de câmara de eco e prejudicando o desenvolvimento do pensamento independente.
Assessor para uma aprendizagem baseada em projetos.	Problemas de pesquisa pouco estruturados em ciências ou estudos sociais.	Apoiar a criação de conhecimento, ajudando os estudantes a realizar uma aprendizagem baseada em projetos. Isso implica a IAGen desempenhando um papel semelhante ao do orientador de pesquisa descrito na Tabela 3. Potencial de transformação: instrutor individual de aprendizagem baseada em projetos.	A partir da lista apresentada na Seção 1.3, avaliar se as ferramentas de IAGen são localmente acessíveis, de código aberto e se foram rigorosamente testadas ou validadas por autoridades. Considerar, ainda, as vantagens e desafios de determinada ferramenta de IAGen, garantindo que ela atenda dequadamente às necessidades humanas específicas.	Os estudantes podem atuar como pesquisadores juniores no planejamento e implementação da aprendizagem baseada em projetos. Os estudantes devem ter idade suficiente para o uso independente de plataformas de IAGen. Eles devem ter motivações e habilidades para participar de atividades autodirigidas de aprendizagem baseadas em projetos, para que não sejam tentados a copiar e colar passivamente as respostas fornecidas pelas ferramentas de IAGen.	Professores humanos orientam os estudantes a solicitar à IAGen ideias básicas para a definição de problemas de pesquisa, conforme sugerido na Seção 5.3.1. Estudantes, sozinhos ou em grupo, utilizam ferramentas IAGen para realizar revisões bibliográficas, coletar e processar dados e criar relatórios.	Estudantes sem conhecimentos prévios e sólidos e sem a capacidade necessária para verificar a precisão das respostas podem ser induzidos ao erro pelas informações fornecidas pelas ferramentas de IAGen. Isso também pode limitar as discussões e interações entre os estudantes e reduzir as oportunidades de aprendizagem colaborativa, prejudicando potencialmente seu desenvolvimento social.

5.3.5 IA generativa como apoio aos estudantes com necessidades especiais

Em teoria, os modelos de IAGen têm o potencial de auxiliar os estudantes com deficiências auditivas ou visuais. As práticas emergentes incluem legendas ou legendas descritivas ativadas por IAGen para estudantes surdos ou com dificuldades auditivas, bem como a descrição de áudio gerada por IAGen para os estudantes com deficiência visual. Os modelos de IAGen também podem converter texto em fala e fala em texto para permitir que pessoas com deficiências visuais, auditivas ou de fala acessem conteúdos, façam perguntas e se comuniquem com seus colegas. No entanto, essa função ainda não foi aproveitada em larga escala. De acordo com a pesquisa mencionada anteriormente, realizada pela UNESCO em 2023 sobre o uso de IA pelos governos na educação, apenas quatro países (Catar, China, Jordânia e Malásia) relataram que seus órgãos governamentais validaram e recomendaram ferramentas assistidas por IA para apoiar o acesso inclusivo de estudantes com deficiência (UNESCO, 2023c).

Também há uma tendência de iterações de modelos de IAGen sendo treinados para apoiar os estudantes a usarem suas próprias línguas, incluindo as minoritárias

e as indígenas, para aprender e se comunicar. Por exemplo, o PaLM 2, o LLM de nova geração do Google, é treinado com dados paralelos que abrangem centenas de idiomas, na forma de pares de texto de origem e destino. A inclusão de dados paralelos multilíngues é projetada para aprimorar ainda mais a capacidade do modelo de entender e gerar texto multilíngue (Google, 2023b).

Ao fornecer tradução, parafraseamento e correção automática em tempo real, as ferramentas de IAGen têm o potencial de auxiliar os estudantes que utilizam línguas minoritárias a comunicar ideias e aprimorar sua colaboração com colegas de diferentes origens linguísticas. No entanto, isso não ocorrerá de forma natural em grande escala. Somente com um desenho intencional pode-se aproveitar esse potencial para amplificar as vozes de grupos marginalizados.

Finalmente, também tem sido sugerido que os sistemas de IAGen têm o potencial de realizar diagnósticos com base em conversas, identificando problemas psicológicos, socioemocionais e dificuldades de aprendizagem. No entanto, há poucas evidências de que essa abordagem seja eficaz ou segura, e quaisquer diagnósticos exigiriam interpretação de profissionais qualificados.

Tabela 7. Co-projetar o uso da IAGen para apoiar estudantes com necessidades especiais

Usos potenciais, mas não comprovados	Domínios de conhecimento ou problemas apropriados	Resultados esperados	Ferramentas apropriadas de IAGen e avanços comparativos	Requisitos para usuários	Métodos pedagógicos humanos necessários e exemplos de <i>prompts</i>	Possíveis riscos
Diagnóstico conversacional de dificuldades de aprendizagem.	Pode ser útil para os estudantes que estão enfrentando dificuldades de aprendizagem causadas por problemas psicológicos, sociais ou emocionais.	Usar a linguagem natural para identificar as necessidades de estudantes que têm problemas psicológicos, sociais ou emocionais ou dificuldades de aprendizagem, a fim de fornecer a eles suporte ou instrução adequados. Potencial de transformação: instrutor inicial e individual para os estudantes com problemas sociais ou emocionais ou dificuldades de aprendizagem.	Além das ferramentas gerais de IAGen, buscar por <i>chatbots</i> controlados por IAGen. Avaliar se eles são localmente acessíveis, de código aberto e se foram rigorosamente testados ou validados por autoridades. Considerar, ainda, as vantagens e desafios de determinada ferramenta de IAGen, garantindo que ela atenda adequadamente às necessidades humanas específicas.	Professores ou especialistas que trabalham com esse grupo de estudantes precisarão garantir que os conselhos iniciais sugeridos pelo sistema de IAGen sejam precisos.	Professores ou facilitadores precisam criar ambientes confortáveis para envolver o estudante em uma conversa, a fim de diagnosticar problemas psicológicos, sociais, emocionais ou dificuldades de aprendizado.	Pode diagnosticar inadvertidamente de maneira incorreta as dificuldades específicas do estudante, levando ao fornecimento de suporte inadequado.

Usos potenciais, mas não comprovados	Domínios de conhecimento ou problemas apropriados	Resultados esperados	Ferramentas apropriadas de IAGen e avanços comparativos	Requisitos para usuários	Métodos pedagógicos humanos necessários e exemplos de <i>prompts</i>	Possíveis riscos
Ferramentas de acessibilidade assistidas por IA.	Essas ferramentas possibilitam que os estudantes com deficiência auditiva ou visual acessem uma gama mais ampla de conteúdo, melhorando assim a qualidade de sua aprendizagem.	Atender às necessidades de acesso dos estudantes e apoiar a aquisição de conhecimentos específicos das disciplinas, fornecendo legendas ou interpretação em linguagem de sinais por IA para conteúdos de áudio ou vídeo e descrições de áudio para textos ou outros materiais visuais. Potencial de transformação: apoio individual e personalizado à linguagem assistido por IA.	Além das ferramentas gerais de IAGen, buscar geradores de legendas e descrições de áudio alimentados por IA que sejam relevantes e confiáveis. Avaliar se são localmente acessíveis, de código aberto e se foram rigorosamente testados ou validados por autoridades. Considerar, ainda, as vantagens e desafios de determinada ferramenta de IAGen, garantindo que ela atenda adequadamente às necessidades humanas específicas.	Os educadores ou facilitadores devem ajudar os estudantes a acessar e aprender a operar as ferramentas de IAGen. Eles também precisam garantir que os resultados das ferramentas realmente apoiem esses estudantes e não reforcem os desafios e preconceitos que eles enfrentam.	É necessário testar a acessibilidade de plataformas ou ferramentas para identificar e corrigir problemas de acessibilidade antes de utilizá-las. As ferramentas de IAGen podem apenas proporcionar acesso ao conteúdo, então educadores e facilitadores devem focar em aprimorar a qualidade da aprendizagem e o bem-estar social. É preciso que ensinem aos estudantes sobre como criar comandos de voz ou texto de acordo com as suas habilidades.	As legendas ou descrições de áudio produzidas por plataformas de IAGen que não foram projetadas especificamente para apoiar a visão ou audição frequentemente são imprecisas e podem induzir os estudantes com necessidades especiais ao erro. Essas ferramentas podem inadvertidamente reforçar preconceitos existentes.
Amplificador generativo para estudantes marginalizados.	Pode ser útil para estudantes de origens linguísticas ou culturais minoritárias expressarem e ampliarem suas vozes, participarem <i>online</i> e realizarem estudos sociais colaborativos.	Oferecer tradução, parafraseamento e correção automática da escrita em tempo real para apoiar os estudantes de grupos marginalizados a usar suas próprias línguas para se comunicar com colegas de diferentes origens linguísticas. Potencial de transformação: LLMs inclusivos para estudantes marginalizados.	Um exemplo específico a ser considerado é o PaLM 2. Avaliar se as ferramentas de IAGen são localmente acessíveis, de código aberto e se foram rigorosamente testadas ou validadas por autoridades. Considerar, ainda, as vantagens e desafios de determinada ferramenta de IAGen, garantindo que ela atenda adequadamente às necessidades humanas específicas.	Os estudantes devem ter conhecimento ou opiniões substanciais sobre o tema da conversa ou estudo colaborativo. Eles precisam ser capazes de fazer contribuições responsáveis e não discriminatórias, evitando discursos de ódio.	Professores ou educadores devem projetar estudos e tarefas escritas para os estudantes sobre temáticas sociais ou culturais, ou organizar seminários <i>online</i> ou colaborações interculturais para estimular os estudantes a gerar ideias e compartilhar opiniões.	É preciso identificar e corrigir os erros nas traduções e parafraseamentos feitos por IA que possam causar mal-entendidos interculturais. Este uso pode proporcionar oportunidades para que os estudantes marginalizados ampliem suas vozes, mas não afetará a causa raiz da pobreza de dados e, portanto, não será capaz de descolonizar as ferramentas de IA.

6. IAGen e o futuro da educação e da pesquisa

As tecnologias de IAGen ainda estão evoluindo rapidamente e têm o potencial de ter um impacto profundo na educação e na pesquisa, visto que elas ainda não são totalmente compreendidas. Portanto, suas possíveis implicações de longo prazo para a educação e a pesquisa exigem atenção imediata e uma revisão mais aprofundada.

6.1 Questões éticas não exploradas

As ferramentas de IAGen, cada vez mais sofisticadas, levantarão preocupações éticas adicionais que precisam ser examinadas em detalhe. Além das Seções 2 e 3, análises mais aprofundadas e voltadas ao futuro são necessárias para revelar e abordar questões éticas ainda não exploradas a partir de ao menos cinco perspectivas, que foram listadas a seguir:

- **Acesso e equidade:** sistemas de IAGen na educação podem agravar disparidades existentes no acesso à tecnologia e aos recursos educacionais, aprofundando ainda mais as desigualdades.
- **Conexão humana:** sistemas de IAGen na educação podem reduzir a interação humana e os aspectos socioemocionais essenciais da aprendizagem.
- **Desenvolvimento intelectual humano:** sistemas de IAGen na educação podem limitar a autonomia e agência dos estudantes ao fornecer soluções pré-determinadas ou reduzir a variedade de experiências de aprendizagem possíveis. Seu impacto a longo prazo no desenvolvimento intelectual de jovens estudantes precisa ser investigado.
- **Impacto psicológico:** sistemas de IAGen que imitam interações humanas podem ter efeitos psicológicos desconhecidos nos estudantes, levantando preocupações sobre seu desenvolvimento cognitivo e bem-estar emocional, bem como sobre o potencial de manipulação.
- **Viés e discriminação ocultos:** à medida que sistemas de IAGen mais sofisticados

são desenvolvidos e aplicados na educação, é provável que eles gerem novos vieses e formas de discriminação com base nos dados de treinamento e nos métodos usados pelos modelos, o que pode acarretar resultados desconhecidos e potencialmente prejudiciais.

6.2 Direitos autorais e propriedade intelectual

O surgimento da IAGen está mudando rapidamente a maneira como obras científicas, artísticas e literárias são criadas, distribuídas e consumidas. A cópia, a distribuição ou o uso não autorizado de obras protegidas por direitos autorais sem permissão do proprietário viola seus direitos exclusivos e pode resultar em consequências legais. Por exemplo, há acusações relacionadas ao fato de os treinamentos de modelos de IAGen estarem infringindo direitos autorais. Em um caso recente, uma música gerada por IA de Drake e The Weeknd (Abel Tesfaye) alcançou milhões de ouvintes antes de ser retirada da Internet devido a uma disputa de direitos autorais (Coscarelli, 2023). Embora os marcos regulatórios emergentes pretendam exigir que os provedores de IAGen reconheçam e protejam a propriedade intelectual dos proprietários do conteúdo usado pelo modelo, está se tornando cada vez mais desafiador determinar a propriedade e originalidade da quantidade avassaladora de obras geradas. Essa falta de rastreabilidade não apenas levanta preocupações sobre a proteção dos direitos dos criadores e a garantia de uma compensação justa por suas contribuições intelectuais, mas também introduz desafios no contexto educacional sobre como os resultados das ferramentas de IAGen podem ser usados de maneira responsável. Isso pode ter implicações profundas para o sistema de pesquisa.

6.3 Fontes de conteúdo e aprendizagem

Ferramentas de IAGen estão mudando a forma de gerar e fornecer conteúdo de ensino e aprendizagem. No futuro, o conteúdo gerado por meio de conversas entre humanos e IA pode se tornar uma das principais fontes

de produção de conhecimento. Isso provavelmente minará ainda mais o envolvimento direto dos estudantes com conteúdo educacional baseado em recursos, livros didáticos e currículos criados e validados por humanos. A aparência fidedigna do texto gerado por IAGen pode confundir jovens estudantes que não têm conhecimento prévio suficiente para reconhecer imprecisões ou questioná-las efetivamente. Reconhecer ou não como “aprendizagem” o engajamento dos estudantes com conteúdo não validado também é algo contestável.

A consequente concentração em informações agregadas de segunda mão também pode reduzir as oportunidades dos estudantes de construir conhecimento por meio de métodos comprovados, como perceber e experimentar diretamente o mundo real, aprender com tentativas e erros, realizar experimentos empíricos e desenvolver o senso comum. Isso também pode ameaçar a construção social do conhecimento e o fomento de valores sociais por meio de práticas colaborativas em sala de aula.

6.4 Respostas homogeneizadas versus produções diversas e criativas

A IAGen restringe narrativas plurais, pois os resultados gerados tendem a representar e reforçar pontos de vista dominantes. Como resultado, a homogeneização do conhecimento limita o pensamento pluralista e criativo. A crescente dependência de professores e estudantes em ferramentas de IAGen para buscar sugestões pode levar à padronização e conformidade de respostas, enfraquecendo o valor do pensamento independente e da pesquisa autônoma. A potencial homogeneização da expressão em textos e obras de arte pode limitar a imaginação, a criatividade e as perspectivas alternativas dos estudantes.

Os fornecedores de IAGen e educadores precisam considerar até que ponto o EdGPT pode ser desenvolvido e usado para promover a criatividade, a colaboração, o pensamento crítico e outras habilidades de pensamento de ordem superior.

6.5 Repensando avaliações e resultados de aprendizagem

As implicações da IAGen para a avaliação vão muito além das preocupações imediatas sobre os estudantes trapacearem em seus trabalhos escritos. Precisamos lidar com o fato de que a IAGen pode produzir

trabalhos e redações relativamente bem-organizados e impressionantes obras de arte, além de ser capaz de passar em exames baseados em conhecimento em determinadas áreas. Portanto, é necessário repensar o que exatamente deve ser aprendido e com que finalidade, bem como a forma que a aprendizagem deve ser avaliada e validada.

A discussão crítica entre educadores, formuladores de políticas, estudantes e outros interessados deve considerar as seguintes quatro categorias de resultados de aprendizagem:

Valores: os valores necessários para garantir a elaboração e o uso da tecnologia centrados no ser humano são centrais para repensar os resultados de aprendizagem e sua avaliação na era digital. Ao revisitar o propósito da educação, os valores que fundamentam a maneira como a tecnologia se relaciona com a educação devem ser explicitados. É por meio dessa perspectiva normativa que os resultados de aprendizagem e sua avaliação e validação precisam ser iterativamente atualizados para responder ao uso cada vez mais pervasivo das tecnologias, incluindo a IA, na sociedade.

Conhecimentos e habilidades fundamentais: mesmo nos domínios de competências nos quais as ferramentas de IA podem superar os humanos, os estudantes ainda precisarão de conhecimentos e habilidades fundamentais robustas. As habilidades fundamentais de alfabetização, numeracia e de literacia científica básica continuarão sendo essenciais para a educação no futuro. O escopo e a natureza dessas habilidades precisarão ser revisados regularmente para refletir os ambientes cada vez mais ricos em IA em que vivemos.

Habilidades de pensamento de ordem superior: os resultados de aprendizagem precisarão incluir habilidades necessárias para apoiar o pensamento de ordem superior e a resolução de problemas com base na colaboração entre humanos e IA, bem como no uso de resultados gerados pela IA. Isso pode incluir a compreensão dos papéis do conhecimento factual e conceitual na fundamentação do pensamento de ordem superior e a avaliação crítica do conteúdo gerado por IA.

Habilidades profissionais necessárias para trabalhar com IA: nos domínios em que a IA pode superar os humanos e está automatizando unidades de tarefas, os estudantes necessitam

de adquirir novas habilidades que lhes permitam desenvolver, operar e trabalhar com ferramentas de IA. A reformulação dos resultados de aprendizagem e da avaliação educacional precisará refletir as habilidades profissionais necessárias para os novos empregos criados pela IA.

6.6 Processos de pensamento

A perspectiva mais fundamental das implicações de longo prazo da IAGen para a educação e a pesquisa segue sendo a relação complementar entre a agência humana e as máquinas. Uma das principais questões é se os humanos podem delegar níveis básicos de processos de pensamento e aquisição de habilidades para a IA e, em vez disso, concentrar-se em habilidades de pensamento de ordem superior com base nos resultados fornecidos pela IA.

A escrita, por exemplo, está normalmente associada à estruturação do pensamento. Com a IAGen, em vez de começar do zero para planejar os objetivos, o escopo e o esboço de um conjunto de ideias, os humanos agora podem começar já com um esboço bem estruturado fornecido pela IAGen. Alguns especialistas caracterizaram o uso da IAGen para gerar textos dessa forma como “escrever sem pensar” (Chayka, 2023). À medida que essas novas práticas assistidas por IAGen se tornam mais amplamente adotadas, os métodos estabelecidos para a aquisição e avaliação de habilidades de escrita precisarão ser adaptados. Uma opção para o futuro é que a aprendizagem da escrita possa se concentrar no desenvolvimento de habilidades de planejamento e composição de *prompts*, na avaliação crítica dos resultados da IAGen e nas habilidades de pensamento de ordem superior, bem como na co-escrita com base nos esboços feitos por IAGen.

Observações finais

A partir da perspectiva de uma abordagem centrada no ser humano, as ferramentas de IA devem ser projetadas para ampliar ou aprimorar as habilidades intelectuais e sociais humanas, e não para miná-las, conflitá-las ou usurpá-las. Há muito se espera que as ferramentas de IA possam ser integradas como parte dos recursos disponíveis para os seres humanos, a fim de apoiar análises e ações para futuros mais inclusivos e sustentáveis.

Para que a IA seja um elemento confiável da colaboração entre seres humanos e máquinas – em níveis individual, institucional e sistêmico – a abordagem centrada no ser humano, da “Recomendação sobre a Ética da Inteligência Artificial” da UNESCO de 2021, precisa ser mais especificada e implementada de acordo com as características específicas de tecnologias emergentes, como a IAGen. Somente dessa forma podemos garantir que a IAGen se torne uma ferramenta confiável para pesquisadores, professores e alunos.

Embora a IAGen deva ser utilizada para servir à educação e à pesquisa, todos nós precisamos estar cientes de que ela também pode alterar os sistemas estabelecidos e seus fundamentos nesses âmbitos. A transformação da educação e da pesquisa desencadeada pela IAGen, caso ocorra, deve ser rigorosamente revisada e orientada por uma abordagem centrada no ser humano. Somente assim podemos garantir que o potencial da IA, em particular, bem como de todas as outras categorias de tecnologias utilizadas na educação de maneira mais ampla, aprimore as capacidades humanas para construir futuros digitais inclusivos para todos.

Referências

- Administração do Ciberespaço da China. 2023a. 国家互联网信息办公室关于《生成式人工智能服务管理办法（征求意见稿）》公开征求意见的通知 [Notificação da Administração do Ciberespaço da China sobre Comentários Públicos a respeito das “Medidas Administrativas para Serviços de Inteligência Artificial Generativa (Projeto para comentários)”]. Administração do Ciberespaço da China (CAC), Pequim. (Em chinês.) Disponível em: http://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm (Acessado em 19 de julho de 2023.)
- Administração do Ciberespaço da China. 2023b. 生成式人工智能服务管理暂行办法 [Medidas Provisórias para o Gerenciamento de Serviços de Inteligência Artificial Generativa]. Administração do Ciberespaço da China (CAC), Pequim. (Em chinês.) Disponível em: http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm (Acessado em 19 de julho de 2023.)
- Anders, B. A. 2023. *Is using ChatGPT cheating, plagiarism, both, neither, or forward thinking?* Cambridge, Cell Press. Disponível em: <https://doi.org/10.1016/j.patter.2023.100694> (Acessado em 23 de junho de 2023.)
- Bass, D.; Metz, R. 2023. *OpenAI's Sam Altman Urges Congress to Regulate Powerful New Technology*. Nova York, Bloomberg. Disponível em: <https://www.bloomberg.com/news/newsletters/2023-05-17/openai-sam-altman-urges-congress-to-regulate-powerful-new-ai-technology> (Acessado em 23 de junho de 2023.)
- Bender, E. M.; Gebru, T.; McMillan-Major, A.; Shmitchell, S. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. Nova York, Association for Computing Machinery. Disponível em: <https://doi.org/10.1145/3442188.3445922> (Acessado em 23 de junho de 2023.)
- Bommasani, R. et al. 2021. *On the Opportunities and Risks of Foundation Models*. Stanford, Universidade de Stanford. Disponível em: <https://crfm.stanford.edu/report.html> (Acessado em 23 de junho de 2023.)
- Bove, T. 2023. *Big tech is making big AI promises in earnings calls as ChatGPT disrupts the industry: 'You're going to see a lot from us in the coming few months'*. Nova York, Fortune. Disponível em: <https://fortune.com/2023/02/03/google-meta-apple-ai-promises-chatgpt-earnings> (Acessado em 3 de julho de 2023.)
- Chayka, K. 2023. *My A.I. Writing Report*. Nova York, The New Yorker. Disponível em: <https://www.newyorker.com/culture/infinite-scroll/my-ai-writing-robot> (Acessado em 1 de agosto de 2023.)
- Chen, L.; Zaharia, M.; Zou, J. 2023. *How Is ChatGPT's Behavior Changing over Time?* Ithaca, arXiv. Disponível em: <https://arxiv.org/pdf/2307.09009> (Acessado em 31 de julho de 2023.)
- Comissão Europeia. 2021. *Laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts*. Bruxelas, Comissão Europeia. Disponível em: <https://artificialintelligenceact.eu> (Acessado em 23 de junho de 2023.)
- Comissão Federal de Comércio. 1998. *Children's Online Privacy Protection Act of 1998*. Washington DC, Comissão Federal de Comércio. Disponível em: <https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa> (Acessado em 4 de setembro de 2023.)
- Coscarelli, J. 2023. *An A.I. Hit of Fake 'Drake' and 'The Weeknd' Rattles the Music World*. Nova York, New York Times. Disponível em: <https://www.nytimes.com/2023/04/19/arts/music/ai-drake-the-weeknd-fake.html> (Acessado em 30 de agosto de 2023.)
- Dwivedi, Y. K.; Kshetri, N.; Hughes, L.; Slade, E. L.; Jeyaraj, A.; Kar, A. K.; Baabdullah, A. M.; Koohang, A.; Raghavan, V.; Ahuja, M.; Albanna, H.; Albashrawi, M. A.; Al-Busaidi, A. S.; Balakrishnan, J.; Barlette, Y.; Basu, S.; Bose, I.; Brooks, L.; Buhalis, D.; Carter, L.; Chowdhury, S.; Crick, T.; Cunningham, S. W.; Davies, G. H.; Davison, R. M.; Dé, R.; Dennehy, D.; Duan, Y.; Dubey, R.; Dwivedi, R.; Edwards, J. S.; Flavián, C.; Gauld, R.; Grover, V.; Hu, M.-C.; Janssen, M.; Jones, P.; Junglas, I.; Khorana, S.; Kraus, S.; Larsen, K. R.; Latreille, P.; Laumer, S.; Malik, F. T.; Mardani, A.; Mariani, M.; Mithas, S.; Mogaji, E.; Horn Nord, J.; O'Connor, S.; Okumus, F.; Pagani, M.; Pandey, N.; Papagiannidis, S.; Pappas, I. O.; Pathak, N.; Pries-Heje, J.; Raman, R.; Rana, N. P.; Rehm, S.-V.; Ribeiro-

- Navarrete, S.; Richter, A.; Rowe, F.; Sarker, S.; Stahl, B. C.; Tiwari, M. K.; van der Aalst, W.; Venkatesh, V.; Viglia, G.; Wade, M.; Walton, P.; Wirtz, J.; Wright, R. 2023. Artigo de opinião: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, Vol. 71. Amsterdã, Elsevier, p. 102642. Disponível em: <https://doi.org/10.1016/j.ijinfomgt.2023.102642> (Acessado em 25 de agosto de 2023.)
- E2Analyst. 2023. *GPT-4: Everything you want to know about OpenAI's new AI model*. San Francisco, Medium. Disponível em: <https://medium.com/predict/gpt-4-everything-you-want-to-know-about-openais-new-ai-model-a5977b42e495> (Acessado em 1 de agosto de 2023.)
- Escritório de Direitos Autorais dos EUA. 2023. Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence. *Federal Register*, Vol. 88, No. 51. Washington DC, Escritório de Direitos Autorais dos EUA, Biblioteca do Congresso, pp. 16190-16194. Disponível em: <https://www.federalregister.gov/d/2023-05321> (Acessado em 3 de julho de 2023.)
- Giannini, S. 2023. *Generative AI and the Future of Education*. Paris, UNESCO. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000385877> (Acessado em 29 de agosto de 2023.)
- Google. 2023a. *Recommendations for Regulating AI*. Mountain View, Google. Disponível em: <https://ai.google/static/documents/recommendations-for-regulating-ai.pdf> (Acessado em 23 de junho de 2023.)
- Google. 2023b. *PaLM 2 Technical Report*. Mountain View, Google. Disponível em: <https://doi.org/10.48550/arXiv.2305.10403> (Acessado em 20 de julho de 2023.)
- Lin, B. 2023. *AI Is Generating Security Risks Faster Than Companies Can Keep Up*. Nova York, The Wall Street Journal. Disponível em: <https://www.wsj.com/articles/ai-is-generating-security-risks-faster-than-companies-can-keep-up-a2bdedd4> (Acessado em 25 de agosto de 2023.)
- Marcus, G. 2022. Hoping for the Best as AI Evolves. *Communications of the ACM*, Vol. 66, No. 4. Nova York, Association for Computing Machinery. Disponível em: <https://doi.org/10.1145/3583078> (Acessado em 23 de junho de 2023.)
- Marwala, T. 2023. *Algorithm Bias — Synthetic Data Should Be Option of Last Resort When Training AI Systems*. Universidade das Nações Unidas. Disponível em: <https://unu.edu/article/algorithm-bias-synthetic-data-should-be-option-last-resort-when-training-ai-systems> (Acessado em 31 de julho de 2023.)
- Metz, C. 2021. *Who Is Making Sure the A.I. Machines Aren't Racist?* Nova York, The New York Times. Disponível em: <https://www.nytimes.com/2021/03/15/technology/artificial-intelligence-google-bias.html> (Acessado em 23 de junho de 2023.)
- Murphy Kelly, S. 2023. *Microsoft is bringing ChatGPT technology to Word, Excel and Outlook*. Atlanta, CNN. Disponível em: <https://edition.cnn.com/2023/03/16/tech/openai-gpt-microsoft-365/index.html> (Acessado em 25 de agosto de 2023.)
- Nazaretsky, T.; Cukurova, M.; Alexandron, G. 2022a. An Instrument for Measuring Teachers' Trust in AI-Based Educational Technology. *LAK22: LAK22: 12th International Learning Analytics and Knowledge Conference*. Vancouver, Association for Computing Machinery, pp. 55-66.
- Nazaretsky, T.; Ariely, M.; Cukurova, M.; Alexandron, G. 2022b. Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British Journal of Educational Technology*, Vol. 53, No. 4. Hoboken, NJ, Wiley, pp. 914-931. Disponível em: <https://doi.org/10.1111/bjet.13232> (Acessado em 1 de agosto de 2023.)
- Ocampo, Y. 2023. *Singapore Unveils AI Government Cloud Cluster*. Cingapura, OpenGov Asia. Disponível em: <https://opengovasia.com/singapore-unveils-ai-government-cloud-cluster> (Acessado em 25 de agosto de 2023.)
- OpenAI. 2018. *AI and compute*. São Francisco, OpenAI. Disponível em: <https://openai.com/research/ai-and-compute> (Acessado em 23 de junho de 2023.)
- OpenAI. 2023. *Educator considerations for ChatGPT*. São Francisco, OpenAI. Disponível em: <https://platform.openai.com/docs/chatgpt-education> (Acessado em 23 de junho de 2023.)
- Popli, N. 2023. *The AI Job That Pays Up to \$335K—and You Don't Need a Computer Engineering Background*. Nova York, TIME USA. Disponível em: <https://time.com/6272103/ai-prompt-engineer-job> (Acessado em 23 de junho de 2023.)

Roose, K. 2022. *An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy*. Nova York, The New York Times. Disponível em: <https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html> (Acessado em 23 de junho de 2023.)

Russell Group, 2023. *Russell Group principles on the use of generative AI tools in education*. Cambridge, Russell Group. Disponível em: https://russellgroup.ac.uk/media/6137/rg_ai_principles-final.pdf (Acessado em 25 de agosto de 2023.)

The Verge. 2023a. *OpenAI co-founder on company's past approach to openly sharing research: 'We were wrong'*. Washington DC, Vox Media. Disponível em: <https://www.theverge.com/2023/3/15/23640180/openai-gpt-4-launch-closed-research-ilya-sutskever-interview> (Acessado em 1 de agosto de 2023.)

The Verge. 2023b. *OpenAI CEO Sam Altman on GPT-4: 'people are begging to be disappointed and they will be'*. Washington DC, Vox Media. Disponível em: <https://www.theverge.com/23560328/openai-gpt-4-rumor-release-date-sam-altman-interview> (Acessado em 1 de agosto de 2023.)

Tlili, A.; Shehata, B.; Agyemang Adarkwah, M.; Bozkurt, A.; Hickey, D. T.; Huang, R.; Agyemang, B. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, Vol. 10, No. 15. Berlim, Springer. Disponível em: <https://doi.org/10.1186/s40561-023-00237-x> (Acessado em 23 de junho de 2023.)

UNESCO. 2019. *Consenso de Beijing sobre a inteligência artificial e a educação*. Paris, UNESCO. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000372249> (Acessado em 3 de julho de 2023.)

UNESCO. 2022a. *Recomendação sobre a Ética da Inteligência Artificial*. Paris, UNESCO. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000381137_por (Acessado em 22 de março de 2024.)

UNESCO. 2022b. *AI and education: guidance for policy-makers*. Paris, UNESCO. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000376709> (Acessado em 23 de junho de 2023.)

UNESCO. 2022c. *Currículos de IA para a educação básica: um mapeamento de currículos de IA aprovados pelos governos*. Paris, UNESCO. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000380602_por (Acessado em 20 de julho de 2023.)

UNESCO. 2022d. *Guidelines for ICT in education policies and masterplans*. Paris, UNESCO. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000380926> (Acessado em 31 de julho de 2023.)

UNESCO. 2023a. *Artificial Intelligence: UNESCO calls on all Governments to implement Global Ethical Framework without delay*. Paris, UNESCO. Disponível em: <https://www.unesco.org/en/articles/artificial-intelligence-unesco-calls-all-governments-implement-global-ethical-framework-without> (Acessado em 3 de julho de 2023.)

UNESCO. 2023b. *Mapping and analysis of governmental strategies for regulating and facilitating the creative use of GenAI*. Não publicado.

UNESCO. 2023c. *Survey for the governmental use of AI as a public good for education*. Não publicado (Submetido para a UNESCO).

UNESCO. 2023. *Technology in Education: A tool on whose terms?* Paris, Global Education Monitoring Report Team. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000385723> (Acessado em 25 de agosto de 2023.)

UNESCO. 2023. *ChatGPT e inteligência artificial na educação superior: guia de início rápido*. Caracas, UNESCO International Institute for Higher Education in Latin America and the Caribbean. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000385146_por (Acessado em 22 de março de 2024.)

União Europeia. 2016. *Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016 relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral sobre a Proteção de Dados)*. Bruxelas, Diário Oficial da União Europeia. Disponível em: <http://data.europa.eu/eli/reg/2016/679/oj> (Acessado em 22 de março de 2024.)

Universidade de Stanford. 2019. *Artificial Intelligence Index Report*. Stanford, Universidade de Stanford. Disponível em: <https://hai.stanford.edu/ai-index-2019> (Acessado em 23 de junho de 2023.)

Universidade de Stanford. 2023. *Artificial Intelligence Index Report*. Stanford, Universidade de Stanford. Disponível em: <https://hai.stanford.edu/research/ai-index-2023> (Acessado em 23 de junho de 2023.)

Notas de fim

- 1 Modelos da IAGen foram disponibilizados a pesquisadores e outras partes interessadas muito antes do ChatGPT. Por exemplo, em 2015, o Google lançou o 'DeepDream' (<https://en.wikipedia.org/wiki/DeepDream>).
- 2 Ver <https://chat.openai.com>
- 3 Para uma explicação das técnicas e tecnologias de inteligência artificial e suas relações, consulte UNESCO, 2022b, pp. 8-10.
- 4 É importante observar que, como a IAGen ainda é relativamente nova, é comum que diferentes empresas utilizem esses termos de maneiras distintas e, por vezes, empreguem palavras diferentes para se referirem ao mesmo conceito.
- 5 Há uma preocupação com o fato de que os dados utilizados para treinar futuras iterações do OpenAI GPT podem incluir quantidades substanciais de texto gerado por versões anteriores do GPT. Esse ciclo auto-referencial pode contaminar os dados de treinamento e, portanto, comprometer as capacidades dos futuros modelos de GPT.
- 6 Observação Importante: A NB OpenIA, empresa que desenvolveu os GPTs nesta tabela, não divulgou publicamente informações detalhadas sobre o GPT-4 (The Verge, 2023a). Na verdade, o número de parâmetros foi refutado pelo CEO da OpenIA (The Verge, 2023b). No entanto, as cifras incluídas aqui foram relatadas por diversas fontes (por exemplo, consulte E2Analyst, 2023). De qualquer forma, a principal conclusão é que o GPT-4 é construído com base em um conjunto de dados massivamente maior e utiliza um número massivamente maior de parâmetros do que o GPT-3.
- 7 Ver <https://crfm.stanford.edu/2023/03/13/alpaca.html>
- 8 Ver <https://bard.google.com>
- 9 Ver <https://writesonic.com/chat>
- 10 Ver <https://yiyao.baidu.com/welcome>
- 11 Ver <https://huggingface.co/chat>
- 12 Ver <https://www.jasper.ai>
- 13 Ver <https://ai.facebook.com/blog/large-language-model-llama-meta-ai>
- 14 Ver <https://open-assistant.io>
- 15 Ver <https://www.alizila.com/alibaba-cloud-debuts-generative-ai-model-for-corporate-users>
- 16 Ver <https://you.com>
- 17 Ver <https://www.chatpdf.com>
- 18 Ver <https://elicit.org>
- 19 Ver <https://www.perplexity.ai>
- 20 Veja <https://tools.zmo.ai/webchatgpt>
- 21 Veja <https://www.compose.ai>
- 22 Ver <https://www.teamsmart.ai>
- 23 Ver <https://wiseone.io>
- 24 Ver <https://www.microsoft.com/en-us/bing>
- 25 Ver <https://www.craiyon.com>
- 26 Ver <https://openai.com/product/dall-e-2>
- 27 Ver <https://dream.ai/create>
- 28 Ver <https://www.fotor.com/features/ai-image-generator>
- 29 Ver <https://www.midjourney.com>
- 30 Ver <https://creator.nightcafe.studio>
- 31 Ver <https://writesonic.com/photosonic-ai-art-generator>
- 32 Ver <https://elai.io>
- 33 Ver <https://www.gliacloud.com>
- 34 Ver <https://pictory.ai>
- 35 Ver <https://runwayml.com>
- 36 Ver <https://www.aiva.ai>
- 37 Ver <https://boomy.com>
- 38 Ver <https://soundraw.io>
- 39 Ver <https://www.voicemod.net/text-to-song>

- 40 Ver <https://openai.com/research/gpt-4>
- 41 Ver <https://www.educhat.top> e <https://www.mathgpt.com>
- 42 Ver <https://www.educhat.top>
- 43 Ver <https://www.mathgpt.com>
- 44 Existem algumas exceções, como a Hugging Face, um grupo dedicado ao desenvolvimento de IA de código aberto.
- 45 Ver, por exemplo, os apelos da Google (2023a) e da OpenAI (Bass; Metz, 2023).
- 46 Para conhecer um projeto de regulamentação de IA, consulte o Projeto de Lei de IA da Comissão Europeia (2021).
- 47 A revisão foi feita com base nos dados coletados de uma pesquisa da UNESCO, que contemplou seus 193 Estados-Membros, sobre o uso governamental de IA na educação (UNESCO, 2023c); no Observatório de Políticas de IA da OCDE; no Relatório do Índice de IA da Universidade de Stanford (Universidade de Stanford, 2023); e nas informações obtidas em primeira mão de um grupo de especialistas internacionais.
- 48 Ver <https://unctad.org/page/data-protection-and-privacy-legislation-worldwide>
- 49 A partir do mapeamento, até abril de 2023, os seguintes países haviam publicado estratégias nacionais sobre IA: Argentina, Austrália, Áustria, Bélgica, Benin, Brasil, Canadá, Bulgária, Chile, China, Colômbia, Chipre, República Tcheca, Dinamarca, Egito, Estônia, Finlândia, França, Alemanha, Hungria, Islândia, Índia, Indonésia, Irlanda, Itália, Japão, Jordânia, Letônia, Lituânia, Luxemburgo, Malásia, Malta, Maurício, México, Países Baixos, Noruega, Nova Zelândia, Omã, Peru, Polônia, Portugal, Filipinas, Catar, República da Coreia, Romênia, Federação Russa, Arábia Saudita, Sérvia, Cingapura, Eslovênia, Espanha, Suécia, Tailândia, Turquia, Tunísia, Emirados Árabes Unidos, Reino Unido, Estados Unidos, Uruguai e Vietnã. Além disso, alguns países incorporaram suas estratégias de IA em estratégias mais amplas de TIC ou digitais, incluindo Argélia, Botswana, Cazaquistão, Kenya, Serra Leoa, Eslováquia, Suíça e Uganda.
- 50 De acordo com uma revisão rápida de todas as estratégias nacionais de IA (UNESCO, 2023b), mais de 40 estratégias têm seções dedicadas à questão da ética.
- 51 De acordo com uma revisão rápida de todas as estratégias nacionais de IA (UNESCO, 2023b), cerca de 45 estratégias têm seções dedicadas à questão da educação.
- 52 Ver <https://openai.com/policies/terms-of-use>
- 53 Em alguns países, um professor pode ter um assistente de ensino (AE), cujo papel é dedicar tempo para responder às perguntas dos alunos sobre o material do curso. A IAGen pode ser utilizada para desenvolver um “gêmeo generativo” de um AE, que pode oferecer suporte aos alunos e outros professores; porém, ele também pode apresentar alguns problemas (por exemplo, no que se refere às relações sociais na sala de aula).



unesco

Organização
das Nações Unidas
para a Educação,
a Ciência e a Cultura

Guia para a IA generativa na educação e na pesquisa

Este guia tem como objetivo apoiar o planejamento de regulamentações, políticas e programas de desenvolvimento de capacidades humanas adequados para garantir que a inteligência artificial generativa (IAGen) se torne uma ferramenta que genuinamente beneficie e capacite professores, alunos e pesquisadores. Ele explica as técnicas de IA usadas pela IAGen e lista modelos GPT disponibilizados ao público, especialmente aqueles sob licença de código aberto. Também inicia uma discussão sobre o surgimento do EdGPT - modelos de IAGen treinados com dados específicos para fins educacionais. Além disso, resume algumas das principais controvérsias em torno da IAGen, desde o agravamento da pobreza digital até a homogeneização de opiniões, e desde os *deepfakes* mais avançados até questões de direitos autorais. Com base em uma visão humanista, a orientação propõe etapas-chave para a regulamentação de ferramentas de IAGen, incluindo a obrigação de proteger a privacidade dos dados e estabelecer um limite de idade para conversas independentes com plataformas de IAGen. Para orientar o uso adequado das ferramentas na educação e na pesquisa, este guia propõe uma abordagem centrada no humano e apropriada para a idade nos processos de validação ética e *design* pedagógico.

