

Ética e IA na Sala de Aula: Relato de Experiência com o Jogo AI-Audit em Disciplinas de Computação

Jhessica Silva^{1*}, Helena Maia^{1*}, Bruno Silva³, João Gondim¹, Gustavo Bonil², Gabriel Santos¹, Helio Pedrini¹, Sandra Avila¹

¹Instituto de Computação ²Instituto de Estudos da Linguagem

³Faculdade de Engenharia Elétrica e de Computação

Universidade Estadual de Campinas (UNICAMP)

Campinas – SP – Brasil

{jhessica.silva, helena.maia, sandra}@ic.unicamp.br

Abstract. *This paper presents an experience report on applying the game AI-AUDIT in three courses offered by the Institute of Computing at UNICAMP. The game presents examples of AI systems, the use of which has proliferated in society, along with the risks they can cause and examples of practices to mitigate them. The application of AI-AUDIT showed us that it is possible to stimulate discussions about ethics and AI in a ludic way. The game reinforced knowledge previously learned in the classroom. It contributed to teaching about risks that were still unknown, indicating its educational potential. However, the game's design and mechanics limitations impacted its effectiveness in building critical thinking on the subject.*

Resumo. *Este artigo apresenta um relato de experiência da aplicação do jogo AI-AUDIT em três disciplinas oferecidas pelo Instituto de Computação da UNICAMP. O jogo apresenta exemplos de sistemas de IA, cujo uso tem-se proliferado na sociedade, juntamente dos riscos que os mesmos podem causar e exemplos de práticas de mitigação desses riscos. A aplicação de AI-AUDIT nos mostrou que é possível estimular discussões sobre ética e IA de forma lúdica. O jogo reforçou conhecimentos aprendidos previamente em sala de aula e contribuiu no ensino de riscos ainda desconhecidos, indicando o seu potencial educativo. No entanto, limitações no design e na mecânica do jogo impactaram sua eficácia na construção de um pensamento crítico sobre o assunto.*

1. Introdução

Produtos baseados em modelos de Inteligência Artificial (IA) estão cada vez mais difundidos no cotidiano das pessoas [Google-Ipsos 2025]. Desde o lançamento do ChatGPT em 2022, *chatbots* se tornaram uma ferramenta popular de interação. Sistemas de vigilância e de segurança residencial utilizam modelos de reconhecimento facial e de outros dados biométricos, enquanto sistemas de recomendação, anúncios personalizados e recrutamento também incorporam modelos de IA [Rao et al. 2024, Stanford University 2025]. Apesar da significativa popularização dos modelos, seus possíveis impactos não foram difundidos na mesma velocidade, o que gera confiança

*Contribuição equivalente.

excessiva nesses sistemas e aplicação indevida. Muitos usuários desconhecem os riscos relacionados à exposição de dados e ao uso dessas ferramentas. Para desenvolvedores e responsáveis pela implementação, compreender as limitações, a aplicabilidade e os riscos da IA é fundamental [Silva 2024]. Dessa forma, a educação sobre os impactos da IA torna-se primordial para acompanhar seu frequente uso.

Na literatura, há um crescente interesse em incorporar o ensino de ética nos cursos superiores de computação. Um estudo sistemático sobre propostas e experiências relacionadas ao tema em cursos superiores de computação [Brown et al. 2024] mostrou que, do total de 100 artigos coletados considerando o período de 40 anos (1983–2022), quase metade foi publicada nos últimos cinco anos desse período. O estudo identificou um desafio recorrente: a predominância de conhecimentos técnicos sem a discussão dos impactos éticos, que recebe pouco espaço nos cursos. Diante dos alertas frequentes sobre os riscos do uso disseminado da IA [O’Neil 2017, Mehrabi et al. 2021, Silva 2022] e dos prejuízos já observados, essa situação precisa ser reconsiderada.

Profissionais formados e em formação reconhecem os impactos dos modelos de IA e a importância da responsabilidade em seu desenvolvimento [Fort e Couillault 2016, Silva et al. 2024b]. No entanto, a avaliação e a discussão sobre os riscos desses modelos ainda não é prática recorrente [Silva et al. 2024a]. Pesquisadores destacam que essa conscientização reforça a necessidade de aprofundar as discussões sobre ética nos cursos de computação e áreas correlatas. Além disso, dúvidas quanto à confiabilidade e ao potencial mau uso dos sistemas de IA evidenciam a necessidade de estimular a reflexão sobre o uso dessas tecnologias na sociedade. Isso levaria à formação de profissionais mais capacitados para a criação e a aplicação de políticas e práticas de IA mais seguras.

No processo de aprendizagem, diversas abordagens podem ser exploradas, sendo o uso de jogos uma ferramenta educacional altamente engajante. O engajamento deve-se à sua estrutura baseada em desafios progressivos, em *feedback* contínuo e em recompensas claras, tornando o aprendizado mais gratificante [McGonigal 2017]. Segundo Miranda [Miranda 2001], “*o jogo é um tipo de linguagem carregado de elementos imagéticos, capaz de atrair a atenção dos estudantes e possibilitar a diversificação do ensino*”. Dessa forma, os jogos favorecem o desenvolvimento do raciocínio lógico, do pensamento crítico e das habilidades sociais, criando um ambiente dinâmico e interativo. Nesse contexto, Bell e Vahrenhold [Bell e Vahrenhold 2018] propuseram o *CS Unplugged*, uma coleção de atividades e jogos desconectados (*unplugged*) para ensinar ciência da computação sem a necessidade de computador. Apesar de sua popularidade entre educadores, a efetividade ainda carece de dados concretos. No entanto, as experiências de uso indicam melhores resultados quando combinadas com métodos formais de ensino.

Diretrizes para o ensino de IA na educação K-12¹ foram propostas pela iniciativa AI4K12 [Touretzky et al. 2019] e organizadas em cinco ideias envolvendo simplificações de conceitos técnicos e, em particular, os impactos positivos e negativos das tecnologias (ideia #5). Motivados pela quinta ideia, pesquisadores desenvolveram um jogo de cartas chamado AI-AUDIT [Ali et al. 2023], uma alternativa lúdica e capaz de gerar maior envolvimento dos participantes, com o intuito de promover reflexões éticas com os estudantes. Segundo Ali et al. [Ali et al. 2023], embora focado na educação K-12, o jogo

¹Educação K-12 é uma nomenclatura dos Estados Unidos e Canadá referente à educação do início do jardim de infância ao fim do ensino médio.

também se mostrou eficaz para estudantes de níveis mais avançados.

No presente trabalho, descrevemos e analisamos a aplicação do jogo AI-AUDIT em três disciplinas oferecidas pelo Instituto de Computação (IC) da Universidade Estadual de Campinas (UNICAMP). O relato da aplicação é estruturado em quatro aspectos: (1) entendimento das cartas, (2) *design* e acessibilidade do jogo, (3) mecanismo do jogo e (4) interação entre os alunos. As principais contribuições deste trabalho são:

- A extensão da avaliação do jogo considerando participantes de contextos cultural e geográfico distintos do original.
- A adaptação do jogo, originalmente em inglês, para o português brasileiro.
- A aplicação do jogo em turmas de diferentes níveis de formação, o que auxilia na análise “aproveitamento *versus* base teórica” entre os alunos.

Nossa experiência indica que o jogo auxiliou a fomentar discussões dos impactos dos sistemas de IA, estimulando o engajamento dos estudantes. Porém, aspectos da mecânica do jogo podem ter levado a conclusões indesejadas, por exemplo, de que questões éticas podem ser resolvidas tecnicamente. Além disso, frequentemente o objetivo de vencer o jogo superou a reflexão sobre os impactos causados pelos sistemas.

2. Jogo de Cartas AI-AUDIT

AI-AUDIT [Ali et al. 2023] foi proposto com o intuito de encorajar a crítica a sistemas de IA usados no cotidiano, fomentando a discussão de possíveis danos que cada IA pode ocasionar e das possíveis estratégias de mitigações desses danos. Trata-se de um jogo de cartas competitivo, onde os jogadores representam donos de negócios de tecnologia atuais, como carros autônomos e filtros faciais. Ao associar as IAs aos seus possíveis danos, o jogo auxilia na compreensão de quais riscos os desenvolvedores podem expor os usuários quando não elaboram e analisam seus sistemas adequadamente (e eticamente).

O jogo é composto de três tipos de cartas: *Negócios (Business)*, *Danos (Harm)* e *Recursos (Feature)* (Figura 1). Os jogadores podem configurar um novo negócio — o que seria o equivalente ao lançamento de um novo sistema de IA por empresas de tecnologia — jogando uma carta de *Negócios*.

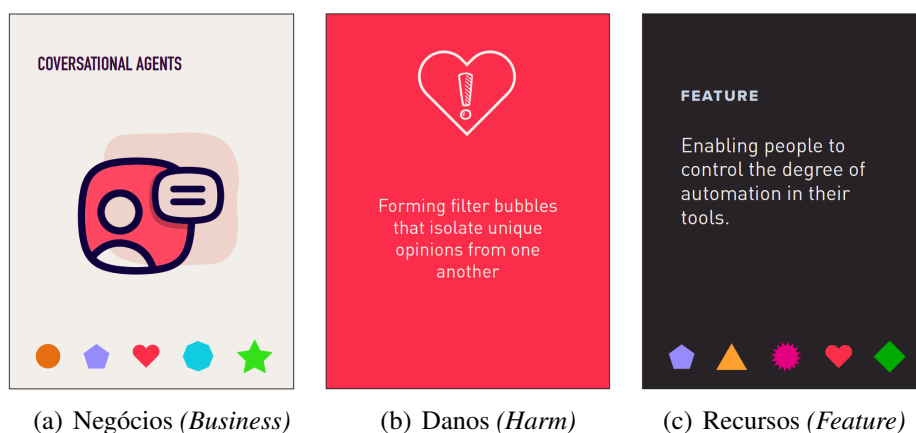


Figura 1. Exemplo das cartas de AI-AUDIT.

Os jogadores também podem desafiar os colegas usando cartas de *Danos*. Para a definição dos danos, Ali et al. [Ali et al. 2023] consultaram materiais sobre ética, adap-

tando os danos apontados aos negócios definidos, sempre levando em consideração o público-alvo. O jogador desafiado pode então utilizar uma carta de *Recursos* para defender seu negócio. Tanto para o desafio, quanto para a defesa, é esperado que os jogadores apresentem argumentos sobre por que o sistema é suscetível àquele risco e como a estratégia usada na defesa pode mitigar o dano. Através de cartas coringas (*Wild Dano*, *Wild Recurso*), os jogadores ainda podem criar seus próprios danos contra um negócio ou seus próprios recursos que mitiguem um dano jogado contra o seu negócio.

3. Metodologia

Esta seção apresenta a metodologia adotada para a aplicação de AI-AUDIT em três disciplinas oferecidas pelo IC/UNICAMP. Para torná-lo possível de ser aplicado, foi necessário adaptá-lo para a língua portuguesa e adequar determinadas cartas. A Seção 3.1 apresenta as adaptações realizadas, a Seção 3.2 apresenta o escopo das disciplinas e o nível de formação dos alunos e a Seção 3.3 apresenta o processo para a aplicação do jogo.

3.1. Adaptações Realizadas no Jogo

Para a aplicação do jogo AI-AUDIT, foi necessário adaptá-lo para a língua portuguesa. Essa adaptação é importante pois as disciplinas são ministradas em português. Dessa forma, não pode ser exigido dos alunos um conhecimento prévio do inglês — língua na qual o jogo foi originalmente publicado². Todas as cartas passaram por um processo de tradução manual do inglês para o português. Para isso, foi necessário refazer as cartas do jogo em tamanho 8,5cm×5,5cm. Neste processo, buscou-se manter a essência do jogo original, sem modificar consideravelmente o conteúdo e o visual das cartas. As informações textuais foram traduzidas para o português; a disposição e o tamanho das ilustrações e dos textos foram reajustados de acordo ao tamanho da impressão (Figura 2).

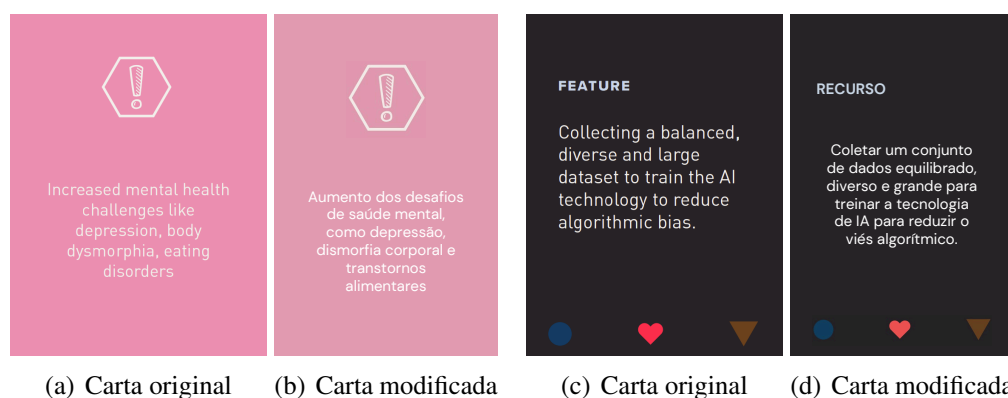


Figura 2. Exemplo das alterações realizadas nas cartas do AI-AUDIT.

Algumas cartas de *Negócios* foram disponibilizadas sem uma descrição textual do que a mesma se referia. Para facilitar a jogabilidade aos alunos, uma pequena descrição foi adicionada a essas cartas (Figura 3). A seguir, são apresentadas as descrições adicionadas no esquema: [*Carta original*]: [**Carta traduzida**] + [nova adição de descrição]

- *Self-driving cars*: **Carros autônomos** que podem se deslocar sem a necessidade de intervenção humana direta.

²A versão utilizada do jogo pode ser acessada em: <https://safinahali.com/aiaudit.pdf>.

- *Conversational agents*: **Agentes de conversação** que podem simular conversas realistas com seres humanos em chats.
- *Generative AI art magazine*: **Revista de arte por IA generativa** que apresenta artes geradas com o uso de sistemas de IA.
- *Socially interactive robot*: **Robô socialmente interativo** que se relaciona de maneira dinâmica com humanos em diferentes tarefas.
- *Language translation algorithm*: **Algoritmo de tradução de idiomas** que realiza a tradução de uma frase entre línguas diferentes.



Figura 3. Exemplo da adição de descrição nas cartas de Negócios do AI-AUDIT.

No processo de adaptação das cartas, notou-se que algumas cartas de *Negócios* foram disponibilizadas com símbolos faltantes ou que não faziam tanto sentido para os pesquisadores deste relato. Com o objetivo de tornar o jogo mais informativo, as cartas com maiores incoerências foram alteradas pela adição ou remoção de danos associados (Figura 4). A seguir, são apresentadas as alterações realizadas:

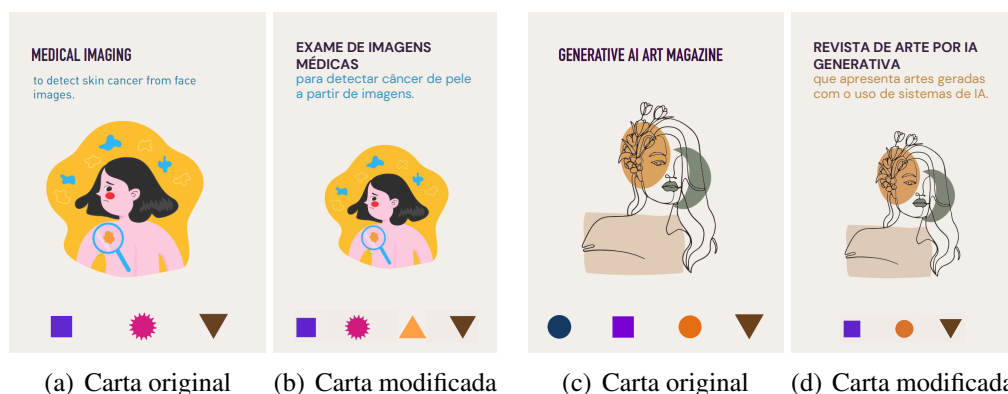


Figura 4. Exemplo da adição ou remoção de danos nas cartas de AI-AUDIT.

- A carta de *Negócios* 'Exame de Imagens Médicas para detectar câncer de pele a partir de imagens' não continha o dano 'Diagnosticar incorretamente a doença de um paciente', sendo que ambos são altamente correlacionadas. Dessa forma, o dano foi adicionado à carta.
- A carta de *Negócios* 'Ferramenta de previsão de crimes que pode prever crimes futuros com uma semana de antecedência e com cerca de 90% de precisão' não continha o dano 'Levar a prisões injustas de pessoas', sendo que ambos são altamente correlacionadas. Dessa forma, o dano foi adicionado à carta.

- A carta de *Negócios* ‘Revista de Arte por IA generativa que apresenta artes geradas com o uso de sistemas de IA’ continha o dano ‘Levar a prisões injustas de pessoas’. Como não conseguimos mapear diretamente o motivo pelo qual esse cenário aconteceria, o dano foi removido da carta.

Nesse sentido, outras alterações ainda poderiam ser feitas, pois entendemos que algumas cartas não abordam danos que poderiam estar presentes. Entretanto, não fizemos essas modificações para não alterarmos drasticamente a forma como o jogo foi proposto.

Por fim, para manter a correlação dos símbolos das cartas de *Danos* com as de *Negócios* e de *Recursos*, foi realizada a troca do círculo laranja ● para uma superelipse (quadrado com cantos arredondados) laranja ■ referente ao dano ‘Espalhar informações erradas ou notícias falsas’ (Figura 5).

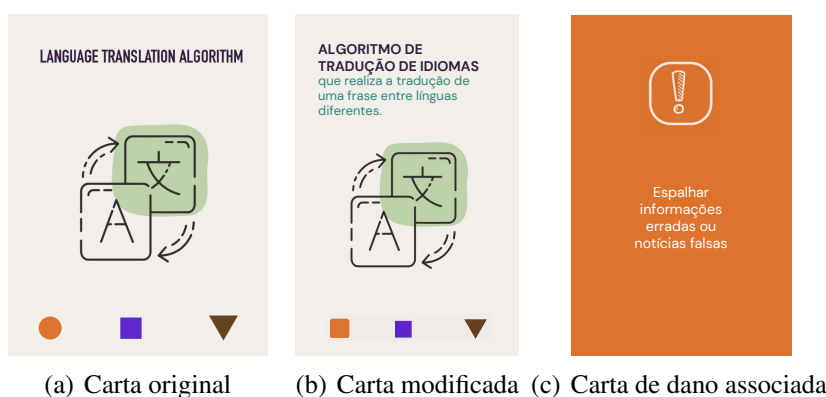


Figura 5. Exemplo da alteração do símbolo de dano nas cartas de AI-AUDIT.

3.2. Caracterização e Formação das Disciplinas

O jogo foi aplicado em três disciplinas oferecidas pelo IC/UNICAMP: Aprendizado de Máquina, Computação e Sociedade e Tecnologias da Informação. A Tabela 1 oferece uma visão geral dessas disciplinas. A seguir, é apresentado o conteúdo abordado em cada uma delas.

Aprendizado de Máquina: Disciplina eletiva oferecida a alunos de graduação e pós-graduação. Aborda técnicas clássicas de aprendizado supervisionado e não supervisionado, além de métodos avançados de aprendizado profundo (*deep learning*). A ética no desenvolvimento de modelos de aprendizado de máquina é tema central do curso, abordada nas quatro primeiras aulas e retomada ao longo do semestre em diversas discussões. A carga horária total é de 60 horas, distribuídas em duas aulas semanais de 2 horas cada. No semestre de aplicação do jogo, a disciplina foi ministrada pela Prof^a. Sandra Avila.

Computação e Sociedade: Disciplina obrigatória para os alunos de graduação de Ciência da Computação e eletiva para os demais cursos. Aborda tópicos sobre problemas éticos decorrentes do uso da tecnologia de informação, oportunidades multidisciplinares do uso da computação, tendências de mercado e globalização, aspectos da política de informática no Brasil e no mundo, entre outros tópicos. A carga horária total é de 30 horas, distribuídas em aulas semanais de 1 hora. No semestre de aplicação do jogo, a disciplina foi ministrada pelo Prof. André Santanchè.

Tecnologias da Informação: Disciplina obrigatória para os alunos do Programa de Formação Interdisciplinar Superior (ProFIS), curso de dois anos da UNICAMP desti-

nado aos estudantes que concluíram o Ensino Médio em escolas públicas de Campinas. O ProFIS tem como objetivo proporcionar aos alunos uma visão abrangente dos diferentes cursos oferecidos pela UNICAMP. Após a conclusão do ProFIS, os alunos podem ingressar diretamente em qualquer curso de graduação da UNICAMP, sem necessidade de prestar vestibular. No contexto do curso de Ciência da Computação, a disciplina abrange, de forma ampla, áreas fundamentais que incluem: noções básicas de programação, arquitetura de computadores, engenharia de *software*, redes de computadores, banco de dados, interação humano-computador e inteligência artificial. A carga horária total é de 60 horas, distribuídas em aulas semanais de 4 horas. No semestre de aplicação do jogo, a disciplina foi ministrada pela Prof^a. Sandra Avila.

Tabela 1. Visão geral das disciplinas.

Disciplina	Nível de formação dos alunos	Número de alunos
Aprendizado de Máquina	A partir do 4º semestre da graduação	32 (Graduação)
	e pós-graduação	31 (Pós-graduação)
Computação e Sociedade	A partir do 1º semestre da graduação	99 (Graduação)
Tecnologias da Informação	Recém-formados do Ensino Médio	107 (ProFIS)

3.3. Aplicação do Jogo

Previamente à aula de aplicação do jogo AI-AUDIT nas três disciplinas, um texto informativo sobre ética em Inteligência Artificial foi enviado aos alunos, juntamente com as regras do jogo³. Esse texto foi adaptado do capítulo “Questões éticas em IA e PLN” do livro das Brasileiras em PLN [Nunes et al. 2024], que oferece uma descrição amigável e introdutória ao assunto, além de ser correlacionado à proposta do jogo.

No começo das aulas de aplicação, as regras do jogo foram apresentadas aos alunos através de uma apresentação de slides ilustrativa⁴. A aplicação foi realizada em dias diferentes para cada uma das disciplinas — nos mesmos dias e horários das aulas de cada uma delas. Em todas as aplicações, os alunos foram divididos aleatoriamente em duas salas de aula, onde as mesas foram organizadas para formar os grupos de jogadores.

Para lidar com a grande quantidade de alunos em cada disciplina, a aplicação do jogo foi realizada em grupos de até 5 duplas — diferente do proposto pelo artigo que o introduz que considerou um grupo de 7 pessoas jogando individualmente. Nesse formato de aplicação, os integrantes de cada dupla jogaram cooperativamente entre si e competitivamente contra os demais. Para todos grupos de alunos, um guia rápido foi oferecido com as principais regras do jogo resumidas⁵. O tempo médio de cada partida foi de aproximadamente 40 minutos.

Durante a aplicação, os pesquisadores envolvidos ficaram à disposição dos alunos para sanar dúvidas e observar como jogavam. A aplicação foi realizada sob supervisão da Prof^a. Sandra Avila com o apoio de 1 pesquisadora de pós-doutorado e 3 alunos de doutorado em Ciência da Computação, 1 aluno de doutorado em Engenharia Elétrica e 1 aluno de iniciação científica do curso de Letras (pesquisadores coautores deste relato).

³Texto introdutório disponibilizado aos alunos junto das regras do jogo.

⁴Apresentação de slides com as regras do jogo em português.

⁵Guia rápido das regras do jogo.

Após a aplicação do jogo, foram coletados relatos orais dos alunos em sala de aula, a fim de compreendermos suas percepções sobre a experiência, como jogaram e quais aprendizados adquiriram ou reforçaram em relação aos conteúdos já aprendidos.

4. Relato de Experiência

A aplicação de AI-AUDIT nas disciplinas apresentadas proporcionou uma experiência extremamente rica, com diversas discussões sobre os impactos éticos da IA no cotidiano e com promoção de engajamento entre os alunos. No entanto, algumas limitações foram observadas ao longo da aplicação, tanto em termos de *design* quanto em termos de mecânica do jogo, impactando em sua eficácia no auxílio à construção (e reflexão) de um pensamento crítico sobre o assunto.

Sobre o entendimento das cartas: Uma das principais dificuldades identificadas por nós, pesquisadores deste estudo, foi a falta de clareza e coerência em algumas cartas. Durante a adaptação para o português, como relatado na Seção 3.1, observamos inconsistências em cartas de *Recursos* e *Danos* atribuídas a algumas cartas de *Negócios*, como no caso da ‘Revista de Arte por IA generativa’, que apresentava o dano de ‘Levar a prisões injustas de pessoas’ sem uma justificativa clara, enquanto a carta de ‘Ferramenta de previsão de crime’ não continha esse dano, apesar de serem altamente correlacionados. A falta de um material de apoio com descrição detalhada das cartas também foi um desafio para uma melhor compreensão do jogo. Esse material, segundo o artigo que o propôs, foi preparado para acompanhar o jogo, porém não foi disponibilizado.

Apesar disso, o jogo gerou muitas discussões entre os alunos sobre o que seria o *Negócio*, o porquê dele ser suscetível a determinados *Danos* e o porquê do *Recurso* ser uma boa defesa, trazendo uma abertura para a reflexão sobre os conceitos abordados.

Sobre o design e acessibilidade do jogo: Na primeira turma onde o jogo foi aplicado, o esquema de símbolos das cartas apresentou problemas de jogabilidade, dificultando a associação intuitiva entre as cartas de *Negócios*, *Recursos* e *Danos*. Isso se deu, pois os alunos não conheciam todas as cartas do jogo. Dessa forma, apenas ver o símbolo do dano sem saber o seu significado, tornou a dinâmica um tanto complicada. Para contornar esse problema, nas aplicações seguintes, criamos e disponibilizamos um guia online para os alunos com os símbolos e seus significados⁶. Apesar dessa limitação, o esquema de cores associado aos símbolos de diferentes formatos mostrou-se também essencial, dado que recebemos o relato de uma pessoa daltônica de que essa opção a deixou mais confortável durante o jogo, dando-lhe mais autonomia.

Enfrentamos, também, uma complicação durante a aplicação da atividade, uma vez que não soubemos com antecedência que uma pessoa participaria remotamente da aula. Dessa forma, não houve um planejamento adequado para sua interação, o que dificultou sua inclusão durante a atividade, na qual ela apenas assistiu a uma equipe jogar.

Sobre o mecanismo do jogo: A mecânica do jogo apresentou limitações, especialmente na quantidade de danos nas cartas de *Negócios* e nas de *Recursos*. Algumas cartas de *Negócios* continham um número significativamente maior de danos do que outras, o que levava a um desbalanceamento. Ou seja, *Negócios* com mais danos foram mais atacados pelos demais jogadores, enquanto os com menos danos possuíam uma alta chance de

⁶Dicionário de símbolos dos danos e seus significados.

vencerem a partida. Esse ponto fez também com que o jogo transferisse aos alunos uma falsa impressão de que algumas IAs são menos perigosas e prejudiciais do que outras.

Adicionalmente, algumas cartas de *Negócios* não representaram adequadamente os desafios éticos associados às mesmas. Como os *Negócios* eram defendidos com *Recursos* que carregavam os mesmos símbolos de danos associados, algumas jogadas se tornaram inconsistentes. Por exemplo, na nossa perspectiva, não tem sentido um determinado *Negócio* ser atacado com o *Dano* ‘Aumento dos desafios de saúde mental, como depressão, dismorfia corporal e transtornos alimentares’ e ser defendido com o *Recurso* ‘Tornar transparentes a tecnologia de IA e a forma de uso dos dados’. Isso fez com que os *Danos* de cada IA fossem tratados de forma superficial, sem trazer para os alunos discussões mais profundas sobre o porquê de cada abordagem. Alguns alunos relataram que, em alguns momentos, o jogo se tornou similar ao UNO (jogo de cartas), onde o pareamento de símbolos se sobrepôs à discussão crítica sobre os impactos dos negócios que estavam em jogo. Nesse caso, pela rapidez na jogada, não houve entre o grupo uma troca de reflexões éticas sobre o impacto de cada tipo de IA.

Ainda sobre a mecânica do jogo, a falta de um mecanismo que permitisse aos jogadores mitigar ou superar danos previamente defendidos em jogadas anteriores limitou as possibilidades estratégicas e tornou algumas jogadas repetitivas. Ou seja, uma mesma carta de *Negócio* poderia ser atacada diversas vezes pela mesma carta de *Dano* e ser defendida pela mesma carta de *Recurso*. Além disso, a ausência de uma mecânica de renovação de cartas na mão fez com que alguns jogadores ficassem travados, sem ações possíveis, comprometendo a participação dos alunos e a dinâmica do jogo.

Por último, o uso das cartas coringas foi ambíguo: enquanto que para alguns alunos elas impulsionaram a criatividade, para os demais, menos familiarizados com os temas abordados, as mesmas representaram uma barreira, pois eles não sabiam como argumentar para ou atacar o negócio de outro jogador ou defender o seu próprio negócio. Ao mesmo tempo, as cartas coringas limitaram alguns alunos que queriam criar danos mais elaborados, por exemplo, criar um dano sobre o impacto do descarte de baterias dos carros autônomos. Porém, por conta da regra que determina que o jogador, ao jogar uma carta de *Dano coringa*, deve escolher o símbolo que ele quer atacar, não faz sentido que esse dano seja jogado, visto que o mesmo não é correlacionado com nenhum dos símbolos existentes e, portanto, não sendo combatido por nenhum *Recurso* existente no jogo.

Sobre a interação dos alunos: A interação entre os alunos foi um dos pontos mais interessantes da experiência. O jogo conseguiu despertar debates importantes, especialmente nos grupos mais engajados. Durante as partidas, os alunos fizeram conexões com os conceitos discutidos em sala de aula, citando exemplos mencionados durante as aulas⁷, o que reforça o potencial do jogo em ser utilizado como ferramenta pedagógica. Adicionalmente, os alunos se interessaram bastante pelo jogo e não queriam parar de jogar, mesmo com o fim do tempo da aula. Isso foi um ponto muito interessante de observação, pois AI-AUDIT foi aplicado nas últimas aulas do semestre em todas as três disciplinas, o que poderia influenciar no interesse dos estudantes pela dinâmica devido ao cansaço do semestre e das atividades avaliativas finais das demais disciplinas.

Um efeito não esperado foi observado, no qual um jogo sobre ética em IA esti-

⁷Por exemplo, os problemas éticos nos sistemas de IA de classificação de imagens de lesões de pele que foram treinados sem considerar imagens de lesões de pele negra [Avila e Marini 2023].

mulou jogadas “antiéticas”. A competição levou jogadores a priorizarem a vitória sobre a reflexão crítica, reduzindo o potencial educativo do jogo. Entendemos que isso ocorreu porque o vencedor do jogo é aquele cujo negócio sobrevive aos ataques. Essa abordagem incentivou comportamentos negativos na defesa dos *Negócios* e do restante do grupo que se une para evitar a vitória de um determinado jogador. Dessa forma, as discussões éticas tiveram um papel menos importante do que a compatibilidade (e a sorte) das cartas de *Negócios*, *Danos* e *Recursos* recebidas. No entanto, como os alunos jogaram em duplas, o modo cooperativo favoreceu trocas de ideias e estratégias, enriquecendo a experiência.

5. Lições Aprendidas

A aplicação de AI-AUDIT nos mostrou que é possível fomentar discussões de problemas atrelados ao uso indiscriminado de IAs de forma divertida através de jogos. Alguns alunos relataram que a experiência os fez perceber danos relacionados à IA que antes não conheciam, como a correlação do problema de dismorfia corporal com IAs que colocam filtros faciais em rostos, o que indica seu potencial educativo. No entanto, foi perceptível que aqueles que já tinham base teórica prévia aproveitaram melhor a atividade, sugerindo que o jogo pode funcionar melhor quando precedido de uma aula introdutória. Com ajustes na mecânica e no *design* das cartas, o jogo tem potencial para se tornar uma ferramenta ainda mais eficaz para a educação crítica sobre o tópico de ética em IA.

Notamos que a mecânica do jogo pode reforçar uma visão otimista em relação às grandes empresas de tecnologias, sugerindo que os problemas éticos relacionados à IA podem ser resolvidos por meio de soluções técnicas que ignoram dimensões políticas e sociais. Embora essa abordagem não tenha comprometido a experiência dos alunos, ela pode ser um ponto de atenção para futuras reformulações do jogo que busquem considerar uma visão mais crítica sobre as relações de poder das empresas criadoras de IAs e suas implicações na sociedade, distanciando-se de uma visão reducionista sobre o assunto.

Também percebemos em alguns casos que, por se tratar de um jogo, o desejo dos alunos era ganhar independentemente de qualquer circunstância e não, necessariamente, refletir sobre o que estava de fato acontecendo na partida. Assim, entendemos que o jogo poderia se beneficiar de regras que não forçasse a discussão, mas que, sim, trouxesse naturalmente (mesmo que somente por reflexão individual e não necessariamente através de discussão entre os jogadores) a reflexão sobre os sistemas de IA do cotidiano.

Por fim, é importante ressaltar que este relato parte da experiência de aplicação do jogo AI-AUDIT em um grupo específico de alunos (ensino superior da UNICAMP) e o jogo passou por adaptações específicas conforme relatado na Seção 3.1. Dessa forma, os relatos deste artigo podem ser diferentes se aplicado em outros contextos.

Agradecimentos. Agradecemos à Marina Bitencourt dos Santos pela revisão do texto. Este projeto foi apoiado pelo Ministério da Ciência, Tecnologia e Inovações, com recursos da Lei nº 8.248, de 23 de outubro de 1991, no âmbito do PPI-SOFTEX, coordenado pela Softex e publicado Arquitetura Cognitiva (Fase 3), DOU 01245.003479/2024-10. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. J.S. e G.S. são financiados pela FAPESP 2024/23118-1 e 2024/07969-1, respectivamente. H.P. é parcialmente financiado pelo CNPq 304836/2022-2. S.A. é parcialmente financiada por CNPq 316489/2023-9, FAPESP 2013/08293-7, 2020/09838-0 e 2023/12086-9.

Referências

- [Ali et al. 2023] Ali, S., Kumar, V., e Breazeal, C. (2023). AI Audit: A Card Game to Reflect on Everyday AI Systems. In *AAAI Conference on Artificial Intelligence*.
- [Avila e Marini 2023] Avila, S. e Marini, M. (2023). “O Dia em que Descobri que Colaborava para um Projeto Eugênico”: Sobre Imponderáveis na Pesquisa Colaborativa. <https://tinyurl.com/castac-avila>. Platypus: The CASTAC Blog.
- [Bell e Vahrenhold 2018] Bell, T. e Vahrenhold, J. (2018). CS Unplugged — How is It Used, and Does It Work? *Adventures Between Lower Bounds and Higher Altitudes*.
- [Brown et al. 2024] Brown, N., Xie, B., Sarder, E., Fiesler, C., e Wiese, E. S. (2024). Teaching Ethics in Computing: A Systematic Literature Review of ACM Computer Science Education Publications. *ACM Transactions on Computing Education*.
- [Fort e Couillault 2016] Fort, K. e Couillault, A. (2016). Yes, We Care! Results of the Ethics and Natural Language Processing Surveys. In *International Conference on Language Resources and Evaluation*.
- [Google-Ipsos 2025] Google-Ipsos (2025). Global AI Optimism Increases as Usage Grows. <https://tinyurl.com/global-ai-optimism>.
- [McGonigal 2017] McGonigal, J. (2017). *A Realidade em Jogo*. Editora Best Seller.
- [Mehrabani et al. 2021] Mehrabani, N., Morstatter, F., Saxena, N., Lerman, K., e Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*.
- [Miranda 2001] Miranda, S. (2001). No Fascínio do Jogo, a Alegria de Aprender. *Ciência Hoje*.
- [Nunes et al. 2024] Nunes, M., Soares, T., e Ferro, M. (2024). Questões Éticas em IA e PLN. In *Processamento de Linguagem Natural: Conceitos, Técnicas e Aplicações em Português*, capítulo 29. BPLN.
- [O’Neil 2017] O’Neil, C. (2017). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown.
- [Rao et al. 2024] Rao, D. V. S., Mr.A.Satish, e Prasad, M. B. (2024). *Artificial Intelligence: Principles and Applications - Chapter 1*. Leilani Katie Publication.
- [Silva 2024] Silva, J. (2024). Avaliação de Ferramentas de Ética em Modelos de Linguagem Desenvolvidos em Português. Master’s thesis, Instituto de Computação, Universidade Estadual de Campinas.
- [Silva et al. 2024a] Silva, J., Ferreira, A., Moreira, D., Santos, G., Bonil, G., Gondim, J., Pereira, L., Maia, H., Silva, N., Hashiguti, S., Avila, S., e Pedrini, H. (2024a). Avaliação de Ferramentas de Ética no Levantamento de Considerações Éticas de Modelos de Linguagem em Português. In *Conferência Latino-Americana de Ética em Inteligência Artificial*.
- [Silva et al. 2024b] Silva, M., Seixas, E., Ferro, M., Viterbo, J., Seixas, F., e Salgado, L. (2024b). Ética e Responsabilidade na Era da Inteligência Artificial: Um Survey com Estudantes de Computação. In *Workshop sobre Educação em Computação*.
- [Silva 2022] Silva, T. (2022). *Racismo Algorítmico: Inteligência Artificial e Discriminação nas Redes Digitais*. Edições SESC SP, São Paulo.
- [Stanford University 2025] Stanford University (2025). The Stanford Emerging Technology Review 2025. A Report on Ten Key Technologies and Their Policy Implications.
- [Touretzky et al. 2019] Touretzky, D., Gardner-McCune, C., Martin, F., e Seehorn, D. (2019). Envisioning AI for K-12: What Should Every Child Know About AI? In *AAAI Conference on Artificial Intelligence*.