UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE COMPUTAÇÃO

Student: André Luiz do Canto Portela RA: 220102 a220102@dac.unicamp.br
Student: Marcelo Aparecido Moreira RA: 183298 m183298@dac.unicamp.br
Professor: Marcelo da Silva Reis msreis7@unicamp.br

Tarefa #1:
Ética & Inteligência Artificial

Objetivo:
Entender e refletir sobre a ética (ou a falta dela) no processo do desenvolvimento de soluções baseadas em Inteligência Artificial, mais especificamente, **Aprendizado de Máquina**.

Observação: Como forma de complementar o conteúdo visto em sala de aula recomendamos que assistam a dois materiais extras:

- A aula 9 do curso FullStack Deep Learning da UC Berkeley que tem como foco Ética em IA.
- A palestra para o Instituto de Computação da UNICAMP do Prof. Dr. Moshe Vardi disponível no link.

Atividades:

1. Defina Ética em Inteligência Artificial.
   - A definição não precisa ser necessariamente com as suas palavras, mas precisa ser minimamente interpretada por vocês. Não basta copiar e colar a definição de alguém.
   - Coloque a referência da definição.

2. Apresente e descreva uma notícia recente (a partir de março de 2021) de um problema de Ética em IA.
   - Coloque a referência da notícia, ou seja, título da notícia, veículo de publicação, data da notícia, nome(s) da(s)/do(s) jornalista(s) e o link da notícia.

3. Apresente um artigo científico recente (a partir de março de 2021) de uma solução para um problema de Ética em IA. Descreva o problema abordado e a técnica utilizada no artigo.
   - Coloque a referência da artigo, ou seja, título do artigo, nome(s) da(s)/do(s) autora(e)(s), local de publicação (nome da conferência/workshop ou do periódico), data de publicação, número de citações do Google Acadêmico/Scholar (https://scholar.google.com/) e o link do artigo.
   - Pode ser artigo (ainda não publicado) do arXiv.
   - O número de citações do Google Acadêmico/Scholar pode ser zero. Tudo bem.
   - Sugestões (mas, não apenas) de termos para procura de workshop / conferência / periódico: fairness, accountability, transparency, ethics.

Observação: Para as três perguntas, as referências podem ser em português ou em inglês. Mas, possivelmente, encontraremos mais artigos científicos publicados na língua inglesa. As notícias contidas nos vídeos de motivação não podem ser utilizadas como resposta da atividade.

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE COMPUTAÇÃO

Prazo: 29 de março, terça-feira, 23:59.

Política de penalidade para submissões atrasadas: Você não está sendo encorajada(o) a submeter o trabalho depois da data de submissão. Entretanto, caso isso aconteça, a nota será penalizada da seguinte forma:

30 de março, 23:59: nota * 0,75

31 de março, 23:59: nota * 0,5

01 de abril, 23:59: nota * 0,25

Submissão:

Envie um arquivo PDF pelo Classroom. **Apenas UMA pessoa da dupla precisa enviar o PDF**. Não esqueça de colocar seus respectivos nomes & RAs.

O trabalho deve ser feito em 2-4 páginas. Formato: tamanho do papel A4, Fonte: Arial 12, Espaço Simples. Todas as perguntas devem ser respondidas em até 4 páginas.

Esta atividade NÃO é individual, deve ser realizada em dupla.

Não tem dupla? https://discord.gg/unBy3Yg3

1. Defina Ética em Inteligência Artificial.

Desde o seu surgimento em meados dos anos 50 com o Dartmouth Summer Research Project on AI, a Inteligência Artificial (IA) vem se tornando, com incrível rapidez, um dos ramos mais fecundos da Ciência. Os avanços nessa área mostram-se cada vez mais sofisticados e poderosos, com o surgimento de máquinas e algoritmos capazes de executar tarefas de altíssimo nível, em boa parte, com maestria superior à humana; os exemplos são diversos, de modo que invenções como programas de reconhecimento facial, algoritmos para diagnósticos médicos e até mesmo carros autônomos já se tornaram uma realidade.

Porém, em meio a tanto progresso, é notório o surgimento de algumas preocupações importantes, que levam em conta os impactos negativos (dos mais variados tipos, desde riscos à vida humana até problemas socioambientais) causados pela adoção de inteligências artificiais executando tarefas muito complexas – e nesse caso, os exemplos também não são poucos: algoritmos com vieses preconceituosos, principalmente racistas e sexistas, e acidentes causados por máquinas autônomas (e.g, o acidente de 2018 envolvendo uma motorista da Uber em um carro autônomo, que atropelou e matou uma mulher nos EUA) são apenas alguns na lista dos mais polêmicos. E, é claro, as questões envolvendo dados pessoais e a privacidade deles. Dessa forma, muito se discute sobre até que ponto automatizar funções primariamente humanas pode ser benéfico; quais as métricas para medir os impactos ocasionados e que tipos de soluções devem ser implantadas. É sobre esses e outros assuntos relacionados que se trata a Ética em IA.

Essa discussão nos leva, primeiramente, a definir o que vem a ser Ética. Não se trata de um tema simples e é tópico de inúmeras teorias. Como um dos objetos de estudo da Filosofia, a Ética é compreendida como um conjunto de princípios e valores definidos socialmente por meio dos quais as relações humanas têm, em tese, a plena capacidade de se estabelecerem em harmonia – do ser humano tanto para com os outros seres vivos, quanto para consigo próprio. Seria o pilar sobre o qual se equilibram as três grandes questões da vida (sob o olhar filosófico): “Quero?”; “Devo?”; “Posso?”. Ela é diferente, por exemplo, da Moral, que seria um conjunto de dogmas a serem seguidos por obrigação, anunciados por uma ordem superior, estando, às vezes, atrelada à Religião. A princípio, a Ética envolve valores que não precisariam ser “forçados”, mas aderidos por meio da compreensão genuína (que pode ser aprendida e praticada) de que são adequados. Em uma de suas aulas – em que ministrou um curso de Deep Learning na UC Berkeley – ao abordar o tema de Ética em IA, o Professor Sergey Karayev procura frisar, por exemplo, que a Ética não está necessariamente ligada às leis ou a convicções pessoais.

Mas o que é, então, a Ética em IA? De forma resumida, podemos dizer que se trata da Ética aplicada à IA, por meio de um conjunto de ferramentas desenvolvidas para fazer valer os princípios éticos nas aplicações de Inteligência Artificial. É uma forma de resguardar e garantir que esses princípios não sejam violados, e que o uso de IA seja feito de forma responsável. Do ponto de vista de Machine Learning, esses cuidados precisam ser tomados durante todo o processo de criação e pós-criação das máquinas e algoritmos, desde a concepção, desenvolvimento, testes e execução, até às fases posteriores de correção e implementação de melhorias. Porém, não bastam. Pois além disso, existem outras questões externas relacionadas. O problema nem sempre está nas máquinas, às vezes está em seus donos.

É o que defende o Professor Moshi Vardi, da Rice University. Em uma recente palestra concedida ao Instituto de Computação da Unicamp, o Dr. Vardi acredita sim que os

cuidados e boas práticas na criação de mecanismos de IA são fundamentais para que as mudanças aconteçam, mas é um pouco cético em relação à eficácia no macro. Para ele, há um conflito de interesses envolvido em torno deste tema, em que grandes corporações, aos exemplos de Google, Apple e Microsoft, não se importam em explorar de forma não-ética o poder atual dos dados e dos recursos em IA, em prol de lucros e de poder. Ele ilustra seus argumentos citando os escândalos envolvendo as políticas e práticas internas do Facebook, em um exposed feito no ano passado, e à demissão pela Google de Timnit Gebru em dezembro de 2020 – uma de suas mais proeminentes pesquisadoras e ex-líder do departamento de Ética em IA – após esta ter tido um de seus artigos censurados e ter se recusado a remover seu nome e o da Google desse artigo (o artigo denuncia uma série de problemas envolvendo Ética em IA, inclusive na própria Google). Conclui dizendo que para que seja perene, as questões envolvendo Ética em IA devem ser amparadas por políticas públicas, afirmando que o Estado ainda se mantém como uma forte instituição, a qual deve fazer valer seu poder e influência.

O que fica então é uma grande reflexão acerca de como a Ética em IA deve ser tratada. De forma crítica, é preciso concordar com o Prof. Vardi e se torna evidente que é quase impossível desvencilhar o assunto de políticas públicas. Além dos constantes esforços para educar e criar boas condutas em IA entre seus praticantes, é importante se ter uma legislação contumaz, que consiga monitorar e auditar as práticas de IA (inclusive nas grandes corporações), para que sejam sentidas na prática as mudanças. Trazendo a questão para o Brasil, já temos algumas melhorias na legislação quanto a isso, como é o caso da criação da Lei Geral de Proteção de Dados - LGPD (aliás, uma excelente palestra do Prof. Gustavo Xavier de Camargo pode ser vista aqui), mas o caminho a seguir ainda é longo.

Por fim, voltando à Filosofia, concluímos que todo o brainstorming em prol de descobrir formas de melhorar o problema é justo e necessário, mas que nesse meio tempo medidas táticas devem ser um imperativo. Buscar as respostas no mundo das ideias de Platão é válido e encorajado, mas levando em conta a velocidade com que a IA se desenvolve, é preciso encarar na prática e com urgência o problema, e levar a discussão para mais perto ainda das leis – numa pegada mais aristotélica.

REFERÊNCIAS

- “AI Ethics (AI Code of Ethics)”, artigo do jornalista George Lawton, disponível aqui;
- “Sobre ética e chocolates”, TEDx Talk pela Prof. Lúcia Helena Galvão, disponível aqui;
- Palestra sobre Ética em IA do Prof. Moshe Vardi, da Rice University, disponível aqui;
- Aula sobre Ética em IA do Prof. Sergey Karayev, da UC Berkeley, disponível aqui.

2. Apresente e descreva uma notícia recente de um problema de Ética em IA.

Em 11 de março de 2022 foi publicada uma notícia no Excalibur (jornal comunitário da Universidade de York no Canadá) entitulada *Addressing inherent racial and gender bias in artificial intelligence*.

A matéria discorre sobre sistemas que utilizam IA na área da saúde com o objetivo de melhorar a eficiência e eficácia em diagnósticos e tratamento ainda ignoram os riscos à saúde de pacientes negros e mulheres. A autora da matéria argumenta que uma porta de entrada para este tipo de viés racial e de gênero são datasets que não representam de maneira significativa a população ao qual os sistemas são destinados e já são construídos de maneira enviesada (ainda que inconscientemente) por pesquisadores e clínicos.

Também é citada uma pesquisa publicada no Journal npj Digital Medicine da Nature que demonstrou que a maioria dos algoritmos de IA com aplicações biomédicas não considerou diferenças de sexo e de gênero, mesmo que essas diferenças já sejam documentadas em condições de saúde baseadas em gênero. Ela cita também um relatório publicado pelo Instituto Alan Turing que demonstra que em geral esses algoritmos são criados por engenheiros e estatísticos (campos de conhecimento com amplos gaps de diversidade racial e de gênero) e que apenas 26% dos profissionais de data science se identificam como mulheres e estão concentradas em trabalhos de menor status (financeiro, responsabilidade, etc).

Situações como a crise sanitária da Covid-19 no Canadá motivaram estudos que demonstram que doenças podem afetar de maneira desproporcional algumas populações em termos de gênero e raça e que se os datasets não forem adequadamente estruturados, então as diferenças entre grupos marginalizados podem ficar escondidas e enviesar o dataset prejudicando as análises posteriores.

Por fim, argumenta-se que representantes destas populações e outros grupos relevantes deveriam estar diretamente e ativamente engajados em cada passo do processo estatístico para minimizar o risco de enviesamento dos treinamentos de algoritmos e análises em relação à grupos em vulnerabilidade social para que IA seja utilizada de uma maneira mais inclusiva e significativa.

A matéria é encerrada apontando que um relatório da Artificial Intelligence and Society Task Force foi publicado em novembro de 2021 recomendando o desenvolvimento de um espaço de pesquisa de IA centralizado em York capaz de atrair um time diverso e multidisciplinar como um passo importante nessa direção.

REFERÊNCIA

- Addressing inherent racial and gender bias in artificial intelligence, disponível aqui.

3. Apresente um artigo científico recente de uma solução para um problema de Ética em IA.

O artigo *“Putting AI ethics to work: are the tools fit for purpose?”* aborda uma problemática ampla que inclui bias, unfairness, lack of transparency e lack of accountability como ameaças em sistemas que aplicam IA e podem potencialmente podem utilizar modelos preditivos para tomada de decisão de maneira incorreta trazendo prejuízo à seres humanos, ter impactos éticos e consequências não intencionais na sociedade.

Os autores do artigo argumentam que existem diversos frameworks relacionados à ética para IA e que eles possuem níveis de escopo diferentes que vão desde a conceituação em alto nível, até ações práticas que podem ser implementadas por sistemas em produção e que apesar de que alguns conceitos sejam desejáveis, sua implementação pode apresentar gaps de aplicação em relação à pessoas, avaliação de impacto, auditoria, processos técnicos, de treinamento (de algoritmos) e de execução em produção.

Os autores propõem uma ampla análise em frameworks éticos práticos de IA para propor uma tipologia dividida em diversas dimensões de interesse para que as necessidades possam ser identificadas, os frameworks classificados e os gaps de necessidades identificados.

Diversos processos de auditoria são comparados em outras áreas que argumenta-se

terem aspectos de governança mais maduros, tais como Technology Assessment (TA), Scientific Technological Options Assessment (STOA), Participatory Technology Assessment (pTA), Environmental Impact and Risk Assessment (respectivamente EIA e ERA), Social Impact Assessment (SIA), Fair Information Practices (FIP), apenas para citar alguns.

Classificações como Auditoria, Avaliação de Impacto, ferramentas técnicas e de projeto, estágio de aplicação, stakeholders, avaliação de risco e outras são propostas melhor estratificar os frameworks éticos. Para cada uma dessas dimensões é proposta uma tipologia que identifica e caracteriza os frameworks para facilitar a análise do escopo individual de cada um, bem como seus gaps e portanto suas vantagens e desvantagens na aplicação.

Por fim, essa tipologia é apresentada de maneira a sugerir uma possível classificação que auxilie na utilização informada de cada framework ético em IA

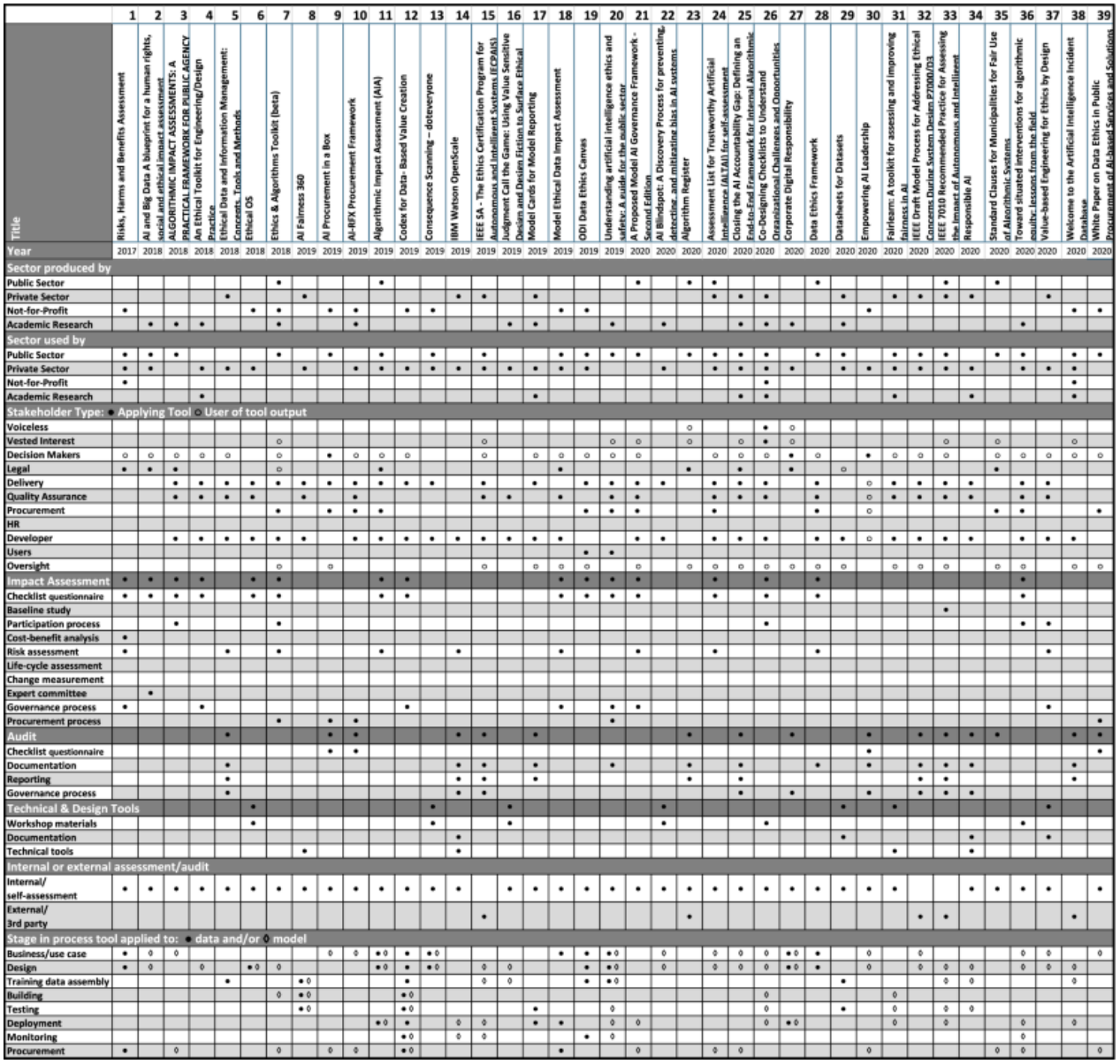

Tabela 10: Overall results for coded document set (n = 39)

REFERÊNCIA

Ayling, J., Chapman, A. Putting AI ethics to work: are the tools fit for purpose?. AI Ethics (2021). https://doi.org/10.1007/s43681-021-00084-x. 6 Citações no Google Schoolar.