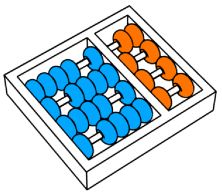


Universidade Estadual de Campinas
Instituto de Computação

Tarefa 1 - MO444

Ética e Inteligência Artificial

Discentes:
André Igor Nóbrega da Silva (RA:203758)
Taylla Milena Theodoro (RA:219596)

Docente:
Prof. Dr. Marcelo da Silva Reis

Campinas
Março, 2022

1 Ética: Definição e seu Contexto em Inteligência Artificial

O conceito de ética é algo mutável ao longo do tempo e das diferentes culturas e sociedades. Para o filósofo São Tomás de Aquino, por exemplo, a ética era pensada segundo uma Lei Natural em que "comportamentos morais são aqueles que promovem, e não prejudicam, um bem básico". Esse bem básico era algo inerentemente dado por uma entidade divina (BINZ, 2017). Na Grécia antiga, a ética era pensada a partir de uma Teoria de Virtudes (do inglês *Virtue Theory*), em que um comportamento ético era o caminho para atingir a eudaimonia. Para Aristóteles, uma pessoa moral era aquela que possuia virtudes básicas, tais como a coragem, a moderação, paciência, dentre outras (BINZ, 2017).

Diversos outros filósofos e pensadores contribuíram com diferentes pontos de vista sobre a ética ao longo dos séculos. Na modernidade, o autor e pesquisador sobre a ética Gustavo Dainezi afirma que a ética é um pensamento contínuo sobre decisões do cotidiano e não um conjunto de regras a ser seguido. Para ele, "ética é a deliberação racional sobre a melhor forma de agir, tanto no nível individual quanto no coletivo. Assim, ética é a atividade livre do pensamento que busca viver a melhor vida, entre todas as possíveis." (DAINEZI, 2021).

Diante desse contexto, antes de definir Ética em Inteligência Artificial (IA), primeiramente é preciso discutir sobre o próprio conceito da Inteligência Artificial. Embora assim como a ética esse também possua uma definição bastante ampla, o autor Stuart Russel afirma que a Inteligência Artificial se preocupa em entender a inteligência e em construir entidades inteligentes capazes de pensar e de agir de maneira racional e humana (RUSSEL; NORVIG, 1995).

Uma grande quantidade de aplicações envolvendo a IA já está presente e já impacta bastante a nossa sociedade. Dessa forma, discutir sobre práticas éticas na construção e no controle desses sistemas inteligentes é essencial. Pode-se definir Ética em Inteligência Artificial como um conjunto de práticas e de conhecimentos para o desenvolvimento de sistemas inteligentes a serviço da vida e da convivência. Isso vai além de elencar e respeitar normas estabelecidas, já que é uma disposição da inteligência coletiva e uma luta para alcançar uma convivência justa (BALDISSERA, 2021).

Podemos, portanto, afirmar que a discussão sobre Ética em IA deve envolver não somente as empresas e os desenvolvedores que implementam os sistemas de inteligência, mas também instituições governamentais e demais âmbitos da sociedade. Tais sistemas devem buscar um equilíbrio para respeito da convivência social.

2 Os Militares Americanos Querem que a IA Substitua a Tomada de Decisão Humana em Batalha

O jornal The Washington Post publicou em seu site, no dia 29 de Março de 2022, a notícia que os militares americanos querem que a IA substitua a tomada de decisão humana na batalha (https://www.washingtonpost.com/technology/2022/03/29/darpa-artificial-intelligence-battlefield-medical-decisions/), a qual foi escrita por Pranshu Verma, um jornalista

do time de tecnologia do jornal.

Movidos pelo ataque bomba suicida no aeroporto internacional de Cabul no ano passado (183 pessoas mortas, dentre elas 13 soldados americanos), a Agência de Projetos de Pesquisa Avançada de Defesa (DARPA) - organização voltada a inovação do exército americano - lançou um projeto de inteligência artificial chamado *In the Moment* que visa substituir humanos na tomada rápida de decisões em situações estressantes.

O programa visa limitar o erro humano em guerra e tentar atualizar o sistema usado em centros de triagem médica. A notícia cita a fala do gerente do programa, Matt Turek, que afirma que IA pode ajudar a identificar recursos de hospitais próximos que podem ajudar na tomada de decisão, tais como disponibilidade de remédios, estoque de sangue e disponibilidade da equipe médica. Ele ainda afirma que todas essas informações não caberiam no cérebro de um ser humano para tomada de decisão e os algoritmos de computador podem encontrar soluções que humanos não conseguem.

O programa da DARPA criará e avaliará algoritmos que auxiliam os tomadores de decisão militares em duas situações: ferimentos em pequenas unidades, como os enfrentados por unidades de Operações Especiais sob fogo; e eventos com vítimas em massa, como o bombardeio do aeroporto de Cabul. Mais tarde, eles podem desenvolver algoritmos para ajudar em situações de desastres, como terremotos, disseram funcionários da agência.

Porém, esse programa levanta o questionamento de colocar vidas humanas nas mãos de uma máquinas. Segundo a noticia, apesar da promessa, alguns especialistas em ética tinham dúvidas sobre como o programa da DARPA poderia funcionar: os conjuntos de dados que eles usam fariam com que alguns soldados fossem priorizados para atendimento em detrimento de outros? No calor do momento, os soldados simplesmente fariam o que o algoritmo mandasse, mesmo que o bom senso sugerisse algo diferente? E, se o algoritmo desempenha um papel na morte de alguém, quem é o culpado?

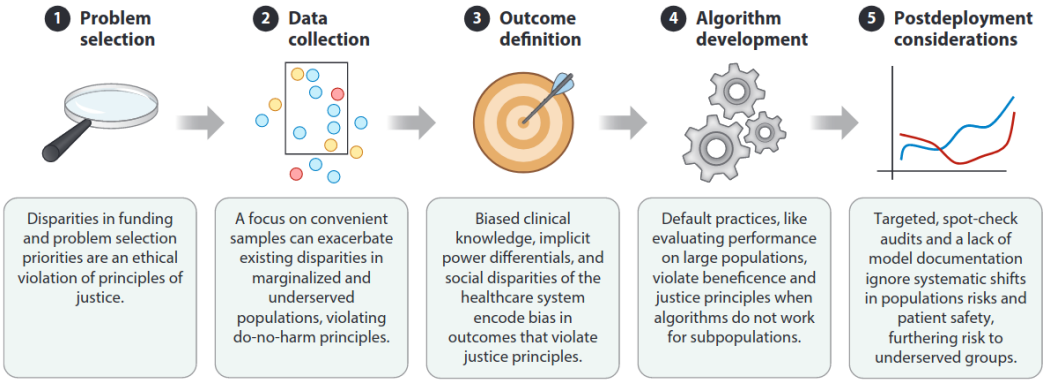
3 Aprendizado de Máquina Ético na Saúde

De acordo com (MURPHY et al., 2021), com o aumento da escala e difusão de tecnologias de IA na saúde em todo o mundo, é imperativo identificar e abordar as questões éticas sistematicamente, a fim de perceber os benefícios potenciais da IA e de atenuar seus danos potenciais, especialmente para os mais vulneráveis.

O artigo Ethical Machine Learning in Healthcare, ou em português, Aprendizado de Máquina Ético na Saúde (CHEN et al., 2021), possui 49 citações no google scholar e discute e propõem soluções interessantes (https://www.annualreviews.org/doi/abs/10.1146/annurev-biodatasci-092820-114757).

Foi proposto a pipeline ilustrada na Figura 1 para o desenvolvimento de modelos na saúde, onde cada estágio contém considerações para aprendizado de máquina em que ignorar desafios técnicos viola o princípio bioético de justiça, seja exacerbando injustiças sociais existentes ou criando o potencial para novas injustiças entre grupos. Tais considerações podem ser utilizadas no desenvolvimento da IA visada para a área médica na Seção 2.

Figure 1: Pipeline para o desenvolvimento do modelo de saúde.



Source: (CHEN et al., 2021)

3.1 Seleção de um problema

Ao selecionar um problema, o artigo propõem levar em consideração disparidades que influenciam este problema, tais como: **injustiça da saúde global**; **injustiça racial** (doenças afetam desigualmente diferentes etnias); **diversidade dos grupos dos próprios cientistas**.

3.2 Coleta de dados

O modelo reflete a distribuição de dados que é treinado, logo a coleta de dados é uma etapa muito importante para ter um modelo com equidade.

Enquanto alguns vieses de amostragem podem ser reconhecidos e possivelmente corrigidos, outros podem ser difíceis de corrigir. É apresentado vieses comuns na coleta de dados, considerando dois processos que resultam em perda de dados. Primeiro, os processos que afetam que tipo de informação é coletada (perda de dados heterogênea) em vários tipos de entrada. Segundo, examinamos os processos que afetam se as informações de um indivíduo são coletadas devido ao tipo de população do indivíduo (perdas de dados específicas da população), geralmente em categorias de entrada de dados.

3.3 Definição do resultado

A próxima etapa no pipeline do modelo do artigo em estudo é definir o resultado de interesse para uma tarefa de assistência médica. Mesmo tarefas aparentemente simples, como definir se um paciente tem uma doença, podem ser distorcidas pela prevalência das doenças ou como elas se manifestam em algumas populações de pacientes. Dentre as dificuldades, podemos citar: rótulos tendenciosos (incentivos econômicos podem alterar o registro do código de diagnóstico; alteração da frequência e observação de testes anormais devido ao protocolo clínico; desconfiança racial histórica pode atrasar o atendimento e afetar os resultados do paciente; coleta de dados ingênua pode gerar rótulos inconsistentes; e, por fim, modelos resultantes podem fazer com que os clínicos aloquem mal os recursos.

Logo, é imprescindível levar em consideração esses detalhes para uma escolha de saída correta.

3.4 Desenvolvimento do algoritmo

O artigo enfatiza que, assim como os dados não são neutros, os algoritmos também não são neutros. Uma quantidade desproporcional de poder está nas equipes de pesquisa que tomam decisões sobre componentes críticos de um algoritmo, como a função de perda. Logo, os autores destacam fatores cruciais no desenvolvimento do modelo que potencialmente afetam a capacidade de implantação ética, como:

- **Entendendo a confusão:** Características de confusão - ou seja, aquelas características que influenciam ambas as variáveis independentes e a variável dependente – requerem atenção cuidadosa. Métodos de aumento de dados e amostragem também podem ser usados para mitigar os efeitos da confusão do modelo;

- **Seleção de atributos:** A incorporação cega de fatores como raça e gênero em um modelo preditivo pode exacerbar as desigualdades para uma ampla gama de diagnósticos e tratamentos. O artigo cita como exemplo as pontuações de parto vaginal após cesariana (VBAC) são usadas para prever o sucesso da tentativa de parto de mulheres grávidas com cesariana prévia; no entanto, essas pontuações incluem explicitamente um componente de raça como entrada, o que reduz a chance de sucesso do VBAC para mulheres negras e hispânicas. Logo, eles concluem que a seleção de atributos automáticos deve ser combinada com o conhecimento do pesquisador.

- **Ajuste de parâmetros:** A falta de generalização é uma preocupação central para o ML ético, quando os dados carecem de diversidade e não são representativos da população-alvo onde o modelo seria implantado, algoritmos de superajuste a esses dados têm o potencial de prejudicar desproporcionalmente os grupos marginalizados. Logo os parâmetros devem ser bem ajustados buscando a generalização.

- **Métricas de performance:** As métricas apropriadas para otimizar dependem do caso de uso pretendido e do valor relativo de verdadeiros positivos, falsos positivos, verdadeiros negativos e falsos negativos.

- **Definição de imparcialidade do grupo:** A definição específica de imparcialidade para uma determinada aplicação geralmente afeta a escolha de uma função de perda e, portanto, o algoritmo subjacente.

3.5 Considerações após a implementação

A implementação é a parte principal do modelo, que é de fato sua utilização em um serviço clínico. A fim de evitar algum impacto ético duradouro, os autores destacaram algumas considerações para uma implantação robusta, como: quantificação do impacto, generalização do modelo, documentação do modelo e dos dados e regularização.

References

BALDISSERA, O. *Ética: aprenda com Clóvis de Barros Filho o que todo profissional do futuro precisa saber*. 2021. Disponível em: <https://posdigital.pucpr.br/blog/clovis-de-barros-filho-curso-etica>. page.11

BINZ, K. *An Introduction to Ethical Theories*. 2017. Disponível em: <https://kevinbinz.com/2017/04/13/ethical-theory-intro/>. page.11

CHEN, I. Y. et al. Ethical machine learning in healthcare. *Annual Review of Biomedical Data Science*, v. 4, n. 1, p. 123–144, 2021. Disponível em: <https://doi.org/10.1146/annurev-biodatasci-092820-114757>. page.22, page.33

DAINEZI, G. *O que é Ética?* 2021. Disponível em: <https://espacoetica.com.br/o-que-e-etica-3/>. page.11

MURPHY, K. et al. Artificial intelligence for good health: a scoping review of the ethics literature. *BMC Medical Ethics*, v. 22, n. 1, p. 14, fev. 2021. page.22

RUSSEL, S. J.; NORVIG, P. *Artificial Intelligence, A Modern Approach*. [S.l.]: Alan Apt, 1995. page.11