

Inteligência Artificial Generativa, dados pessoais e Literacia Digital em Saúde: possibilidades, desigualdades e limites

Arthur William Cardoso Santos

Mestre em Educação, Cultura e Comunicação (UERJ - Universidade do Estado do Rio de Janeiro), Rio de Janeiro/RJ, Brasil. ORCID: <https://orcid.org/0009-0000-6268-8734>

RESUMO

Este artigo apresenta um panorama geral do atual uso de ferramentas de Inteligência Artificial, em especial a Inteligência Artificial Generativa, na área da Saúde, apontando possibilidades, desigualdades e limites de sua adoção. Essa tecnologia tem o potencial de provocar uma grande revolução e sua rápida penetração exige um esforço para garantir que o impacto da IA na sociedade seja regulado com participação social e transparência. Como as aplicações de Inteligência artificial exigem o treinamento de grande quantidade de dados (Big Data), analisaremos também questões de interoperabilidade e uso de dados pessoais no contexto do *Open Health*. Outro ponto a ser discutido é como a Inteligência Artificial impacta a Literacia em Saúde, principalmente a Literacia Digital em Saúde. Adicionalmente, trataremos da importância da “explicabilidade” em sistema de IA no sentido de garantir a transparência de suas respostas. No mesmo caminho, vamos verificar como vieses e preconceitos podem interferir na análise destas aplicações. No campo da Inteligência Artificial Generativa, serão abordados aspectos como a relação entre mídia sintética e desinformação, além de possibilidades do uso de NLP (Processamento de Linguagem Natural) por profissionais de Saúde.

Palavras-chave: Inteligência Artificial, IA, IA Generativa, Literacia em Saúde, Literacia Digital em Saúde.

Generative Artificial Intelligence, personal data and eHealth Literacy: possibilities, inequalities and limits

ABSTRACT

This article presents an overview of the current use of Artificial Intelligence tools, in particular Generative Artificial Intelligence, in Healthcare, pointing out possibilities, inequalities and limits of its adoption. This technology has the potential to cause a major revolution and its rapid penetration requires an effort to ensure that the impact of AI on society is regulated with social participation and transparency. As Artificial Intelligence applications require the training of large amounts of data (Big Data), we will also analyze issues of interoperability and use of personal data in the context of

Open Health. Another point to be discussed is how Artificial Intelligence impacts the Health Literacy, especially the eHealth Literacy. Additionally, we will deal with the importance of “explainability” in an AI system in order to guarantee the transparency of its responses. In the same sense, we will verify how biases can interfere in the analysis of these applications. In the field of Generative Artificial Intelligence, aspects such as the relationship between synthetic media and disinformation will be addressed, as well as possibilities for the use of NLP (Natural Language Processing) by health professionals.

Keywords: Artificial Intelligence; AI; Generative AI; Large Language Model; eHealth Literacy.

Inteligencia Artificial Generativa, datos personales y Alfabetización Digital en Salud: posibilidades, desigualdades y límites

RESUMEN

Este artículo presenta una visión general del uso actual de las herramientas de Inteligencia Artificial, en particular de la Inteligencia Artificial Generativa, en Salud, señalando posibilidades, desigualdades y límites de su adopción. Esta tecnología tiene el potencial de causar una gran revolución y su rápida penetración requiere un esfuerzo para que el impacto de la IA en la sociedad se regule con participación social y transparencia. Como las aplicaciones de Inteligencia Artificial requieren el entrenamiento de grandes cantidades de datos (Big Data), también analizaremos cuestiones de interoperabilidad y uso de datos personales en el contexto de Open Health. Otro punto a discutir es cómo la Inteligencia Artificial impacta en la Alfabetización en Salud, especialmente en la Alfabetización en eSalud. Adicionalmente, trataremos la importancia de la “explicabilidad” en un sistema de IA para garantizar la transparencia de sus respuestas. En el mismo sentido, comprobaremos cómo los sesgos pueden interferir en el análisis de estas aplicaciones. En el campo de la Inteligencia Artificial Generativa se abordarán aspectos como la relación entre los medios sintéticos y la desinformación, así como las posibilidades de uso del NLP (Procesamiento del Lenguaje Natural) por parte de los profesionales sanitarios.

Palabras clave: Inteligencia artificial; Inteligencia Artificial Generativa; LLM; lenguaje natural; alfabetización en salud .

INTRODUÇÃO

Há alguns anos, a Inteligência Artificial (IA) já é aplicada na área da Saúde desde o diagnóstico antecipado de doenças até a criação de novos medicamentos. Contudo, é

relativamente recente a popularização da IA Generativa e seu uso na Saúde impõe uma série de questões, as quais discutiremos ao longo deste artigo.

Detecção de câncer com base na análise automatizada de imagens, desenvolvimento de biofármacos, melhor utilização dos recursos em um hospital (ocupação de salas de cirurgia e leitos), acompanhamento do paciente na atenção primária, além de embasamento para decisões de gestão e a criação de políticas públicas: todos estes são usos de IA já em operação na Saúde.

Com o avanço rápido da inteligência artificial a partir do início de 2023, uma série de empresários e estudiosos do tema lançaram uma carta pedindo a paralisação do desenvolvimento da tecnologia para que a sociedade consiga se adaptar ao impacto desta novidade (FUTURE OF LIFE, 2023). Tal proposta dialoga com o “Princípio da Precaução” (GOLDIM, 2002), muito utilizado na Bioética, em que uma atividade é restringida com o objetivo de prevenir possíveis danos ainda desconhecidos a partir de sua execução.

O processo de Saúde Digital vem caminhando ao longo dos anos, mas sofreu forte desenvolvimento no último período por conta da quarentena imposta pela pandemia de Covid-19. A área da Saúde já é conhecida por ser “*late adopter*” de novas tecnologias, pois precisa ter garantias de que sua adoção não vai trazer riscos à vida da população. Por isso, os algoritmos precisam ser validados pela Agência Nacional de Vigilância Sanitária (Anvisa) no Brasil. Inclusive, em 23 de março de 2022, a diretoria colegiada da Anvisa aprovou a resolução nº 657 (ANVISA, 2022) que regulamenta o uso de softwares como dispositivos médicos (*Software as a Medical Device* - SaMD).

Em janeiro de 2023, a renomada revista Nature proibiu que o ChatGPT seja creditado como autor de artigos (NATURE, 2023). Caso a ferramenta seja utilizada, o contexto deve ser esclarecido pelos autores humanos. No mesmo mês, a revista Science publicou editorial com o título “ChatGPT é interessante, mas não um autor” (THORP, 2023). Em março do mesmo ano, a Itália proibiu o ChatGPT no país até que a OpenAI, empresa desenvolvedora da ferramenta, explicasse questões de privacidade.

Diversos países, em especial na Europa, já desenvolvem legislações no sentido de regulamentar a adoção da inteligência artificial. No Brasil, há três projetos de lei em tramitação no Congresso e um anteprojeto produzido por uma comissão de juristas (BRASIL, 2023). O objetivo é tornar mais transparente o uso desta tecnologia, garantindo os direitos do cidadão, além de prevenir possível utilização de mídia sintética gerada por IA generativa em Fake News, cenário de desinformação preocupante na área da Saúde por conta da situação da Infodemia.

Outra questão jurídica diz respeito aos direitos autorais dos conteúdos (textos, fotos, vídeos e músicas) gerados por IA. Antes mesmo da regulação legal/infralegal, algumas plataformas já rascunham princípios de boas práticas na implementação de IA generativa ao dar mais transparência ao processo de criação de imagens e vídeos por IA, uma evolução em relação a ferramentas anteriores que usam material proprietário de terceiros sem autorização para gerar novos conteúdos. A alternativa seria apenas utilizar material expressamente autorizado, em

domínio público e de seu banco de imagens próprio (WIGGERS, 2023). Os produtores de conteúdo poderiam escolher ainda se autorizam que suas obras sejam usadas para o treinamento de robôs e a ferramenta recompensaria quem o assim desejar, criando uma economia da IA.

A privacidade também é uma questão central no contexto da Inteligência Artificial, na medida em que os sistemas são treinados a partir de grande base de dados. Essa *Big Data* muitas vezes conta com dados sensíveis de pessoas sem autorização das mesmas.

Sobre a acurácia dos textos gerados por IA, é preciso compreender que os sistemas são alimentados por “*machine learning*”, ou seja, não são ferramentas prontas ou totalmente programadas por humanos, mas sim em processo de aprendizado autônomo, então, por hora não poderiam ser consideradas como fonte totalmente confiável.

Outro ponto importante é a opacidade dos sistemas de aprendizado de máquina. Os chamados “*black boxes*” impedem que o usuário consiga entender quais foram os critérios utilizados para gerar a resposta recebida. Tal processo de “explicabilidade” (LIMA, 2022) é fundamental para verificar as possibilidades de uso das ferramentas de IA em serviços como a Saúde.

O desenvolvimento de ferramentas de Inteligência Artificial requer acesso a grandes bancos de dados (*dataset*) e à elevada capacidade de processamento de dados, recursos que demandam um custoso investimento em infraestrutura computacional. Esta situação projeta um cenário de potencial desigualdade entre os sistemas público e privado.

No momento, o SUS está discutindo as possibilidades e os desafios para o uso de inteligência artificial na Saúde Pública. O Conselho Nacional de Saúde (CNS) está pautando a IA dentro do contexto de Saúde Digital e da Transformação Digital em Saúde. A Resolução nº 659, de 26 de julho de 2021, elenca a IA entre as disposições da Política Nacional de Informação e Informática em Saúde (PNIIS).

IV - promoção da disseminação de dados e informações em saúde e do uso de inteligência artificial de forma a atender tanto às necessidades de usuários, de profissionais, de gestores, de prestadores de serviços e do controle social, quanto às necessidades de intercâmbio com instituições de formação, ensino e pesquisa, entre outras (PIGATTO, 2021).

Em conjunto com a Fiocruz, o CNS está discutindo a implementação da IA e outras possibilidades das ITIS (Informação e Tecnologia da Informação em Saúde) à luz dos princípios e diretrizes do SUS.

Outro passo neste sentido veio do Ministério da Saúde ao criar a Secretaria de Informação e Saúde Digital com os seguintes setores subordinados: Departamento de Informação para o Sistema Único de Saúde (Datusus); Departamento de Avaliação e Disseminação de Informações Estratégicas em Saúde; e Departamento de Saúde Digital e Inovação.

De acordo com o Artigo 53 do Decreto nº 11.358, de 1º de janeiro de 2023, compete à recém criada Secretaria de Informação e Saúde Digital:

[...] apoiar as Secretarias do Ministério da Saúde, os gestores, os trabalhadores e os usuários no planejamento, no uso e na incorporação de produtos e serviços de informação e tecnologia da informação e comunicação - TIC; incluídos telessaúde, infraestrutura de TIC, desenvolvimento de software, interoperabilidade, integração e proteção de dados e disseminação de informações (BRASIL, 2023).

A ABRASCO (Associação Brasileira de Saúde Coletiva) também trouxe elementos para este momento por meio do Plano Diretor para o Desenvolvimento da Informação e Tecnologia de Informação em Saúde para o quinquênio 2020-2024 (3º PlaDITIS). O documento lista cinco dimensões para: Governança e gestão da informação e tecnologia de informação e comunicação em saúde (ITICS); Pesquisa, desenvolvimento e inovação na área temática ITICS; Ensino e formação permanente de equipe de Informação e TIC em saúde; Ética, privacidade e confidencialidade; Informação e TIC em saúde: democracia, controle social e justiça cognitiva.

Como referência mundial em Saúde Pública, o desenvolvimento de um sistema de Saúde Digital pelo SUS pode servir como referência para a adoção por outros países. Situação análoga ocorreu com o Sistema Brasileiro de TV Digital (SBTVD) que serviu de base para países como Angola, Argentina, Peru, Chile, Uruguai, Costa Rica, Paraguai, Bolívia e Venezuela. No caso da TV Digital, o Brasil fez melhorias no sistema japonês ISDB-T, incorporando tecnologia desenvolvida por pesquisadores brasileiros como a interatividade Ginga, cujas linguagens NCL e Lua foram criadas pelo laboratório TeleMídia da PUC-Rio.

Antes da popularização da Inteligência Artificial Generativa, a IA já estava presente no cotidiano de grande parte dos brasileiros de forma pouco perceptível. Nos conteúdos, a IA atua fortemente nos sistemas de recomendação de vídeos e músicas em aplicativos como Netflix, Spotify e Youtube. Nota-se que, para a IA atuar, é necessário ter acesso a informações pessoais dos usuários como dados de pesquisa e acesso na internet.

Em aplicativos de navegação por GPS, como Waze e Google Maps, por exemplo, a Inteligência Artificial é responsável por responder à pergunta indireta feita ao sistema: Qual o trajeto mais rápido de um ponto no mapa para outro?

Nem todos os motoristas utilizam e se guiam pelas respostas da IA, que algumas vezes cometem equívocos. Entretanto, os profissionais que se recusam a usar a tecnologia como auxílio a seu ofício (uma espécie de “Copiloto”) são considerados desatualizados e tendem a ficar fora do mercado. O mesmo fenômeno ocorreu com a adoção dos computadores pessoais nas empresas.

Contudo, da mesma forma que o bom taxista ou motorista de UBER seria aquele que utiliza as informações do GPS, mas utiliza sua experiência profissional para identificar possíveis erros nas rotas sugeridas, os médicos poderão utilizar a Inteligência Artificial como ferramenta auxiliar e não como um oráculo infalível, detentor de toda a verdade e sabedoria.

Neste sentido, a tendência é que os profissionais e gestores de Saúde atuem em conjunto com cientistas de dados para a resolução de problemas de seu cotidiano. Estamos apenas no início deste processo e as projeções acima ajudam a estabelecer marcos jurídicos que possibilitem a adoção da IA preservando as garantias legais do cidadão em uma democracia.

A Inteligência Artificial Generativa é um fenômeno recente, mas que já causa impactos na vida em sociedade, por isso a urgência em sua regulação. O ChatGPT foi o aplicativo de mais rápida adoção na história. Demorou apenas 5 dias para atingir a marca de um milhão de usuários. A título de comparação, o Facebook levou 10 meses para chegar a este patamar.

Além do ChatGPT, outras aplicações de IA Generativa já estão sendo utilizadas: Bard Ai, Bing AI, 365 Copilot, DeepComposer, Dall-E, Midjourney, Firefly, Stable Difusion, Stability AI, Bing Image Creator, Canva IA, Duolingo AI, Runway, Synthesia, entre outros. Entender as possibilidades, os riscos e limitações da tecnologia por trás destas ferramentas nos ajudará a projetar oportunidades e a regular seu uso na Saúde.

INTELIGÊNCIA ARTIFICIAL NA SAÚDE

Primeiramente, vamos elencar alguns dos principais conceitos de que trataremos durante o artigo. Inteligência Artificial (IA) é um conjunto de técnicas que pretendem simular o funcionamento da inteligência humana. O NLP (“Processamento de Linguagem Natural” na tradução livre) é um ramo da IA que trata da interação entre máquinas e seres humanos através de linguagem natural. Já a Inteligência Artificial Generativa é uma técnica de IA que possibilita a criação de textos, fotos, vídeos e música com base em modelos de dados treinados. Esses modelos, por sua vez, são compostos pelo processamento de uma grande quantidade de parâmetros, o que chamamos de LLM (“*Large Language Models*” ou “Grandes Modelos de Linguagem” na tradução livre).

A Inteligência Artificial na Saúde está inserida no rol de possibilidades das ITIS (Informação e Tecnologia da Informação em Saúde) e possibilita a criação de uma cultura de pensamento analítico em que profissionais e gestores de Saúde tomam decisões embasadas em evidências e não na intuição, processo conhecido como “*Data Driven*” (“orientado a dados” na tradução livre).

A IA atua basicamente em três eixos na Saúde: políticas públicas, otimização de recursos e melhoria do atendimento.

No primeiro eixo, a análise de dados de sistemas como o DataSUS e do IBGE auxilia na construção e na implementação de programas e ações de Saúde Coletiva pelo Estado. As informações mais utilizadas para este fim estão reunidas no Censo, na Pesquisa Nacional de Saúde (PNS), no Sistema de Informação sobre Mortalidade e Declaração de Óbitos Fetais (SIM-DOFET), no Sistema de Informações sobre Nascidos Vivos (SINASC), no Cadastro Nacional

de Estabelecimentos de Saúde (CNES), no Sistema de Informações Hospitalares do SUS (SIHSUS) e no Sistema de Informação sobre Mortalidade (SIM).

O segundo eixo permite uma melhor gestão de equipes médicas e da infraestrutura hospitalar. Já no terceiro, serão mais utilizadas as ferramentas de IA Generativa conforme veremos à frente.

A cultura de dados, difundida entre os profissionais da Saúde, possibilita que as decisões tomadas sejam baseadas em evidências e não apenas pautadas na experiência profissional e acadêmica. Ser “*Data Driven*” é, portanto, promover um processo organizacional baseado em dados a partir da coleta e análise de informações.

Com a grande quantidade de dados a serem analisados (*Big Data*), a IA lança mão de recursos de *machine learning* (aprendizado de máquina) para produzir suas próprias regras, numa escala muito maior do que um ser humano poderia programar. Na área da Saúde, técnicas de *machine learning* permitem a predição de doenças antes de os primeiros sintomas aparecerem, como o caso de câncer em estágio inicial. A máquina aprende a identificar padrões a partir do treinamento alimentado com casos anteriores.

A medicina preditiva usa complexos algoritmos para tentar prever condições futuras de saúde baseada em informações genéticas e no histórico familiar. O uso de dados pessoais também possibilita um atendimento preciso e personalizado com base em exames atuais e do passado. Contudo, coexistem aí questões sobre a privacidade do paciente e o uso dos dados pessoais para outros fins, conforme veremos à frente.

Ferramentas de análise automatizada de imagens radiológicas formam um dos principais serviços de Inteligência Artificial autorizadas pela FDA (*Food and Drug Administration*, agência federal de Saúde dos Estados Unidos), indo além do auxílio ao médico, mas também já indicando um possível diagnóstico de forma não supervisionada. Em 2018, a FDA autorizou o uso de um aplicativo que detecta retinopatia de forma autônoma, o IDx-DR. No contexto da pandemia do Coronavírus, a Anvisa liberou a venda em farmácias de testes rápidos para o autodiagnóstico da Covid-19.

A perspectiva é que a IA na Saúde não seja um serviço autônomo, mas que possa auxiliar os profissionais na automatização de rotinas, assim como fornecer subsídios para o raciocínio. No caso dos gestores, proporciona maior embasamento para a tomada de decisão, otimizando recursos. Exemplos são: melhor organização de escalas de profissionais e uso de equipamentos e insumos, além da construção de prontuários de forma automatizada e de previsões financeiras, promovendo conseqüentemente uma redução de custos operacionais.

- Dados de Saúde e Open Health

Um dos empecilhos para o uso de dados em Saúde é o fato de estarem em bases distintas, espalhados em sistemas de agências reguladoras, Ministério da Saúde, Ministério do Trabalho e sistemas privados como farmácias, laboratórios e hospitais.

Para o eixo das Políticas Públicas, a Fiocruz fornece acesso às informações do DataSUS (ligado ao Ministério da Saúde) através da Plataforma de Ciência de Dados aplicada à Saúde (PCDaS), projeto do Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICT/Fiocruz). Tais dados propiciam diversas soluções para a melhoria da saúde da população por meio de análises de informações de variadas fontes. Outra iniciativa que atua no mesmo sentido é Centro de Integração de Dados e Conhecimentos para Saúde (Cidacs) da Fiocruz Bahia, projeto de referência responsável por realizar pesquisas que utilizam a *Big Data* com o objetivo de aprimorar a compreensão das relações entre a Saúde populacional e as políticas ambientais e sociais.

Estas ações são importantes ainda pois atuam na limpeza e no enriquecimento dos dados brutos, garantindo a qualidade e a confiabilidade das informações coletadas, o que chamamos de “veracidade” em *Big Data*. Além disso, é possível contar com dados não estruturados de fontes como redes sociais.

No eixo de “Melhoria do atendimento”, a IA pode substituir a consulta a livros como as diretrizes clínicas de um Guideline Médico (ANS & AMB, 2012). No contexto de uma emergência, a consulta a tabelas de um Guideline por médicos não especialistas seria facilitada pela indicação de exames a partir dos indicadores fornecidos ao sistema de IA.

Junto com as informações baseadas em evidências de um Guideline, o processamento pela IA de dados pessoais do paciente admite a prescrição de tratamentos personalizados de acordo com cada pessoa, elevando a taxa de efetividade. Alguns hospitais brasileiros já utilizam a ferramenta Memed para a prescrição digital de medicamentos, contudo, conforme veremos posteriormente, a adoção de ferramentas do tipo pelo Sistema Único de Saúde (SUS) requer outras características como licenciamento de software livre. Exemplo é o sistema AGHUse, software livre sob licença GPL para gestão de hospitais universitários desenvolvido partir de parceria entre o Hospital das Clínicas de Porto Alegre (HCPA), o Ministério da Educação (MEC) e a Empresa Brasileira de Serviços Hospitalares (Ebserh). Os aplicativos de marcação de consultas online estão difundidos no sistema privado, mas podemos destacar o “Agenda Fácil” da Secretaria Municipal de Saúde de São Paulo além da experiência do AGHUse implementada também em hospitais das forças armadas (Aeronáutica, Marinha e Exército).

Mas atenção: as máquinas não substituirão os seres humanos. Como uma solução baseada em *machine learning* sempre estará sujeita a equívocos, é responsabilidade dos profissionais de Saúde entender o patamar de confiabilidade dos sistemas e dosar o nível de influência que os resultados automatizados terão nas decisões médicas. Além disso, como especialistas no ofício da medicina, esta categoria deve atuar em conjunto com os técnicos para

aprimorar o algoritmo a partir de técnicas como RLHF (*Reinforcement Learning from Human Feedback*, aprendizagem por reforço por resposta humana) e “*supervised learning*” (aprendizado supervisionado).

No momento, a Inteligência Artificial encontra-se em estágio inicial de desenvolvimento. Suas aplicações ainda povoam o ramo da IA Específica (ANI - *Artificial Narrow Intelligence*), sendo destinada a resolução de problemas limitados. A segunda etapa seria a Inteligência Artificial Geral (AGI - *Artificial General Intelligence*), cuja principal característica é contar com comportamentos humanos como sentimentos e emoções. A terceira etapa seria a superação da inteligência humana pelas máquinas, fenômeno chamado de superinteligência (ASI - *Artificial Super Intelligence*).

Paralelamente, a criação de humanóides, congregando IA Generativa e robótica, já está sendo experimentada na Ásia como o robô chinês Xiaoyi, além dos “*carebots*” (robôs cuidadores) no Japão, devido ao envelhecimento da população e à escassez de cuidadores humanos neste país. No Brasil, o Hospital Albert Einstein faz uso de robôs de telepresença dentro da dinâmica da telemedicina, permitindo que médicos possam acompanhar seus pacientes a distância.

Mesmo que apenas funcionando no ambiente virtual, os chatbots costumam usar nomes de pessoas para dar mais credibilidade à comunicação entre humanos e máquinas. Exemplos são: Bia (Bradesco), Lu (Magazine Luiza), Alexa (Amazon), Siri (Apple), Cris (Crefisa), CB (Casas Bahia), Nat (Natura) e Vivi (Vivo). Na área da Saúde, o Robô Laura atua na atenção primária. Criada inicialmente para a detecção antecipada da sepse via inteligência artificial, a Laura Care ganhou uma versão que funciona como uma assistente virtual que acompanha toda a jornada do paciente, desde a triagem inicial via WhatsApp até o monitoramento do paciente 14 dias depois do último atendimento.

Uma das barreiras à implantação da cultura de dados na área da Saúde é a digitalização das informações dos pacientes. Segundo dados de 2021, a adoção de Prontuário Eletrônico do Paciente (PEP) já ocorreu em 89% dos estabelecimentos de Saúde, contudo a maioria (84%) refere-se apenas a informações cadastrais do paciente, sendo que este percentual sofre redução quando relacionados a outras informações clínicas: Diagnóstico, problemas ou condições de saúde do paciente (73%), Alergias do paciente (66%), Resultados de exames laboratoriais do paciente (65%), Lista de medicamentos prescritos ao paciente (62%), Vacinas administradas ao paciente (47%) e Imagens de exames radiológicos do paciente (30%).

Nota-se no estudo acima (CETIC, 2021) que o percentual mais baixo está justamente nas imagens radiológicas, área em que a primeira onda de IA na Saúde teve grande presença. Na área pública, este percentual seria de apenas 26%.

Outro entrave é a padronização de dados estruturados e que possibilitem a interoperabilidade entre diversos sistemas, permitindo que o paciente tenha garantidas a transição e a continuidade no cuidado entre os diversos níveis da hierarquização do SUS, além do uso do

sistema de saúde complementar, formando uma EHR (sigla em inglês para *Electronic Health Record*).

Criada pela Portaria nº 1.434, de 28 de maio de 2020 e alinhada à política de Transformação Digital na Saúde, a Rede Nacional de Dados em Saúde (RNDS) tem o objetivo de sistematizar as informações dos pacientes dos sistemas público e privado. A RNDS utiliza o padrão FHIR, a terminologia LOINC e a arquitetura descentralizada Blockchain por apresentar melhor potencial de segurança, acesso, escalabilidade e desempenho.

A RNDS é uma das principais políticas estruturantes para a implementação do conceito de *Open Health* no Brasil, na medida em que viabilizaria a integração de dados de serviços públicos e privados, dentro de um sistema com controle social. Informações do SUS e de empresas comerciais (Einstein, Rede Dor, Dasa, Fleury, entre outros) seriam compartilhados aprimorando a experiência do paciente e evitando perda de tempo na realização de exames repetidos.

A experiência de intercâmbio de informações de usuários já acontece no Brasil com o chamado *Open Banking*, assegurando a portabilidade de dados entre diversos prestadores de serviços bancários. A modalidade é regulada pelo Estado através do Banco Central.

O mercado também possui iniciativas de integração de sistemas sem a mediação estatal, como é o caso do *Open Delivery*. Contudo, neste caso, na contramão do cenário de elevação na competitividade do setor, algumas empresas de entregas fecharam suas operações recentemente no Brasil, como Uber Eats e 99Foods.

Até 2023, a principal experiência exitosa da RNDS foi o projeto ConecteSUS no contexto da pandemia do Novo Coronavírus. Este sistema reúne registros de vacinação, exames de Covid-19, atendimentos, internações e medicamentos. O “ConecteSUS Cidadão” é uma evolução do “Meu digiSUS”, aplicativo que apresentava ao cidadão um conjunto de informações reunidas de sistemas como o Portal do Cidadão, o Cadastro Nacional de Usuário do Sistema Único de Saúde (CADSUS) e o Cadastro Nacional de Estabelecimento de Saúde (CNES). Com o *app*, era possível agendar procedimentos e exames inseridos no sistema da Atenção Básica e-SUS AB e no Sistema de Regulação (SISREG).

Lançado em agosto de 2020, a obrigatoriedade da emissão do comprovante de vacinação tornou o aplicativo ConecteSUS popular. Até dezembro de 2022, a ferramenta já tinha sido baixada mais de 36 milhões de vezes. Já em julho de 2021, foi criada uma versão focada nos profissionais de Saúde, o “ConecteSUS Profissional”. Este sistema permitia o acesso ao histórico médico do cidadão, condicionado à autorização do mesmo. Nos EUA, o *Health Information Exchange* (HIE) também exige consentimento do paciente para o intercâmbio de informações.

Essas ações estão alinhadas à Política Nacional de Informática e Informações em Saúde (PNIIS, 2021), à Estratégia da e-Saúde (CIT 2017), ao Plano de Ação, Monitoramento e Avaliação de Saúde Digital para o Brasil (PAM&A 2019) e à Estratégia de Saúde Digital para o Brasil 2020-2028 (ESD28) e à Estratégia e-Saúde para o Brasil.

A adoção de ferramentas de Inteligência Artificial na Saúde pode encontrar um empecilho na cultura do médico como protagonista exclusivo da Saúde. A presença da IA pode, em alguns episódios, contestar uma avaliação ou tomada de decisão por parte do médico. A perspectiva positiva é que a transformação digital propicie uma atuação do médico mais como tutor, aumentando e compartilhando de responsabilidades por meio da coparticipação do próprio paciente, na medida em que este último estará cada vez mais empoderado de sua situação clínica graças ao conhecimento de suas informações médicas conforme veremos no tópico sobre a “Literacia Digital em Saúde”.

Neste caminho, a jornada do paciente deve ser redesenhada com processos baseados nas necessidades do usuário (ou experiência do usuário - UX) e não nas necessidades do médico ou do hospital. Por fim, a IA não tem a pretensão de substituir os profissionais da saúde, mas, ao automatizar rotinas, o trabalho tende a ser cada vez mais dedicado a atividades de interação com o paciente e menos no preenchimento de relatórios.

- **Desigualdades**

Antes restrita a profissionais de Tecnologia da Informação (TI) e programadores, hoje a Inteligência Artificial passa por um processo de universalização, contudo ainda não há equidade no acesso a estas ferramentas.

Lançados inicialmente de forma gratuita, alguns aplicativos de IA já estão se tornando um *Software as a Service* (SaaS) cobrado dentro do modelo *Freemium* (algumas funcionalidades gratuitas e outras pagas), gerando desigualdades em seu uso.

O ChatGPT, por exemplo, permite o acesso gratuito à versão 3.5, enquanto os assinantes do plano *Plus* têm direito ao uso da versão 4.0. Além da frequente indisponibilidade da versão gratuita, seus usuários contam com um sistema que não consegue entender imagens, que não cita as fontes de pesquisa e que processa cerca de oito vezes menos palavras em comparação com a versão GPT-4.

Outro diferencial da versão Plus do ChatGPT é que ela tem acesso a informações atualizadas de seu *dataset*, diferentemente da opção gratuita cuja base de dados só vai até setembro de 2021.

Já a ferramenta de IA Generativa de imagens “Midjourney” encerrou os acessos gratuitos alegando sobrecarga do sistema após a viralização de imagens sintéticas do Papa Francisco e do ex-presidente dos EUA, Donald Trump, criadas a partir do aplicativo. Nesta esteira, o serviço de IA Generativa “Copilot” estará disponível no serviço oneroso “Microsoft 365”.

Mais uma desigualdade reside na criação de ferramentas de IA apenas para usuários de dispositivos da Apple, como iPhone e iPad. O aplicativo Kaiber, por exemplo, só foi disponibilizado inicialmente para usuários do sistema iOS, o que representa em 2023 uma ampla minoria de pessoas em todo o mundo.

Já o “Bard”, sistema de inteligência artificial do Google, foi lançado primeiramente apenas para residentes nos Estados Unidos e no Reino Unido. Numa segunda onda, abriu a ferramenta para 180 nações, mas excluiu da lista Brasil, Canadá e países da União Europeia, justamente as regiões que estão em processo de implementação de leis que regulamentam o uso da internet, impondo limites e responsabilidade às *Big Techs*.

Há ferramentas que possibilitam o uso de computadores pessoais para a geração de imagens, como o “Stable diffusion”, porém é necessário o uso uma GPU (unidade de processamento gráfico) potente, dispositivo de elevado custo. Já os sistemas de análises de dados com IA e as ferramentas de *machine learning* exigem grande capacidade computacional para serem treinados e terem as informações processadas, configurando mais uma desigualdade neste campo.

Outra barreira de implementação está no acesso à internet no Brasil. Apesar de estar difundida para a maioria da população, levantamento (IDEC, 2021) mostra que a adoção de planos com *Zero-Rating* pelas classes C, D e E cria um descompasso entre as possibilidades de acesso à internet de acordo com o poder aquisitivo do cidadão.

Mesmo proibido pelo Marco Civil da Internet (Lei nº 12.965/2014) e pelo Decreto 8771/2016, as operadoras continuam oferecendo planos com franquias pequenas de pacote de dados mas com acesso ilimitado a determinados aplicativos como Whatsapp e Facebook. Tal situação de limitação de acesso à informação é apontada como uma das causas para a desinformação durante a pandemia da Covid-19, fenômeno chamado de Infodemia (D’AGOSTINI, 2021).

Art. 10. As ofertas comerciais e os modelos de cobrança de acesso à internet devem preservar uma internet única, de natureza aberta, plural e diversa, compreendida como um meio para a promoção do desenvolvimento humano, econômico, social e cultural, contribuindo para a construção de uma sociedade inclusiva e não discriminatória (BRASIL, 2016).

Em janeiro de 2023, a partir de pedido do movimento social Coalizão de Direitos da Rede (CDR), o Governo Federal solicitou que o CADE analisasse os impactos do *zero-rating*.

Mais um ponto de destaque diz respeito às licenças de uso dos softwares e sistemas envolvidos na Saúde Digital. O Prontuário Eletrônico do Cidadão (PEC), por exemplo, é baseado no software público e-SUS AB. A mesma diretriz deve acompanhar o desenvolvimento de ferramentas de IA para o SUS, garantindo sua universalização.

Por outro lado, no que tange à infraestrutura, o DataSUS está hospedado em servidor privado (AWS - Amazon Web Services). Esta opção reduz a independência da gestão dos dados e pode colocar em risco a política pública se houver descontinuidade do serviço com impossibilidade de migração para outra infraestrutura. Tal percepção é aumentada pelo fato de a empresa contratada possuir um projeto de *Data Lake* para a Saúde (Amazon HealthLake), que possui o protocolo FHIR, escolhido pela Organização Mundial da Saúde (OMS) para sistemas de

Open Health. O Amazon HealthLake já está sendo utilizado pelo Hospital Sírio-Libanês em aplicações de ciência de dados na saúde privada.

A título de comparação, o Ministério da Educação (MEC) não conseguiu acesso ao acervo de conteúdos produzidos durante a execução de contrato de gestão com a Associação de Comunicação Educativa Roquette Pinto (ACERP) para funcionamento da TV Escola e da TV INES, situação que teve como consequência a descontinuidade dos projetos anteriores e a criação posterior de duas novas emissoras em parceria com a Empresa Brasil de Comunicação (EBC): Canal Educação e Canal Libras.

A própria lei de criação do SUS (Lei 8080 de 1990) em seu artigo 47 já prevê a criação de um “sistema nacional de informações em saúde, integrado em todo o território nacional, abrangendo questões epidemiológicas e de prestação de serviços”.

Quanto à inovação em Saúde, o setor privado tem apostado na incubação de *startups*, como é o caso da Eretz.bio do Hospital Albert Einstein e da Alma do Hospital Sírio-Libanês. Todavia, os empreendimentos do tipo são comumente adquiridos por grandes corporações ou morrem sem conseguir amadurecer o produto ou serviço. Por isso, para o SUS, o mais efetivo seria o investimento em políticas estruturantes com o desenvolvimento de laboratórios de inovação em centros de pesquisa públicos.

Diferentemente dos princípios e diretrizes do SUS, a inovação no setor privado tem como base um processo organizado em seis etapas por Peter Diamandis, sendo que somente o último seria o da “democratização” (DIAMANDIS, 2016). Deste modo, a inovação caminhará mais para uma medicina de precisão focada apenas no atendimento de um pequeno segmento da sociedade.

Desigualdade adicional diz respeito à quantidade de dados disponíveis sobre cada paciente. O fenômeno da Internet das Coisas Médicas (*internet of medical things* - IoMT) consiste na presença de diversos dispositivos que geram dados de saúde sobre as pessoas para além do ambiente hospitalar e do consultório médico. Estes dispositivos compartilham dados em tempo real que podem antecipar o diagnóstico de situações como um infarto, o que aumentaria a possibilidade de recuperação do enfermo. Dispositivos vestíveis (*wearable devices*) já estão presentes no cotidiano de uma parcela da sociedade, contudo ainda são inacessíveis à maioria da população relógios inteligentes (*smart watches*), sensores de pele (*skin sensors*), medidor de pressão conectado, além de sensores de movimento para a prática esportiva, batimentos cardíacos (HRM) e monitoramento do sono.

A extração dos dados da IoMT possibilita um atendimento personalizado, aprofundado e ágil aos pacientes, causando grande diferença se comparado ao atendimento generalista, superficial e demorado de quem não tiver acesso a tais dispositivos, principalmente na atenção primária. Por outro lado, a geração de grande volume de dados pessoais traz à tona questões de privacidade e controle social dos dados, as quais abordaremos mais à frente.

No caso de regiões remotas, como aldeias indígenas e comunidades rurais e ribeirinhas com reduzida assistência de Saúde, políticas públicas de IoMT poderiam complementar programas como o “Mais Médicos” (PMM), por exemplo, permitindo ao médico que as informações de Saúde sejam processadas de forma remota, auxiliando o trabalho do profissional de saúde localmente. Programas neste sentido seriam gestados por ações conjuntas dos ministérios da Saúde e das Comunicações.

O Programa de Apoio ao Desenvolvimento Institucional do Sistema Único de Saúde (Proadi-SUS) já possui um projeto-piloto neste sentido em cidades ribeirinhas do Amazonas e do Pará (projeto TeleAMEs), onde um médico generalista da atenção básica faz a intermediação para o teleatendimento por especialistas localizados em outras cidades brasileiras (LIRA, 2022). Projeto também de referência no SUS, o programa de TelessaúdeRS promove ações de teleeducação, telediagnóstico e teleconsultoria na atenção primária no Rio Grande do Sul.

Sobre a jornada do paciente, sem a efetiva implementação do *Open Health* no Brasil, apenas as empresas verticalizadas conseguiriam utilizar os dados clínicos coletados, formando uma espécie de oligopólio no setor de saúde privada. A Rede D’or, por exemplo, adquiriu recentemente o seguro Sulamérica, a rede de maternidades Perinatal e o laboratório Richet.

Uma defasagem no sentido da conectividade e de uso de dados em situações de emergência poderá ser encontrada nas ambulâncias com conexão 5G. Já em funcionamento a partir de parceria entre a TIM, a Deloitte e o Hospital Sírio-Libanês, este “veículo inteligente” promove uma melhor troca de informações entre a equipe do hospital e da ambulância, diminuindo os riscos de equívocos de comunicação na transição do paciente.

Dentro dos hospitais, a conectividade dos dispositivos viabiliza uma supervisão remota em tempo real reduzindo a probabilidade de erros médicos. O Hospital Israelita Albert Einstein (HIAE) possui um centro de controle operacional que supervisiona toda a jornada do paciente e aciona equipes locais caso algum dos indicadores esteja em desconformidade, como o caso de a equipe de enfermagem não administrar determinada medicação dentro do tempo previsto.

Em processo análogo à Divisão Internacional do Trabalho (DIT), outra perspectiva de análise sobre a desigualdade reside no papel de países desenvolvidos e subdesenvolvidos no processo de produção e processamento de dados em nível mundial. No “colonialismo de dados”, os países subdesenvolvidos forneceriam a matéria prima (dados) para o desenvolvimento de tecnologias pelos países desenvolvidos.

Em 2021, o Núcleo de Estudos sobre Bioética e Diplomacia em Saúde (Nethis) da Fiocruz Brasília criou o Observatório de Desenvolvimento e Desigualdades em Saúde e Inteligência Artificial (Odisseia) com o objetivo de fomentar a elevação do acesso do SUS à Transformação Digital em Saúde em conjunto com a IA.

- **Vieses e preconceito**

Além da desigualdade no acesso a recursos de Inteligência Artificial, a iniquidade de serviços de *machine learning* pode acarretar em respostas equivocadas por parte das ferramentas de Inteligência Artificial. Como as primeiras aplicações foram desenvolvidas com dados de populações de países desenvolvidos, a máquina pode reproduzir preconceitos com populações distintas, processo chamado de BIAS ou viés. A análise automática de imagens médicas agiliza a confecção de laudos, porém precisa levar em conta dados que muitas vezes não estão presentes.

Neste sentido, se não houver políticas públicas para a inclusão de populações vulnerabilizadas em ações de *machine learning*, especificidades étnicas podem ser desconsideradas pelos sistemas de IA, prejudicando a análise de dados de indígenas brasileiros, por exemplo. Isto porque, se estiver apenas pautada pelo potencial econômico da coleta de dados, informações de segmentos sociais historicamente excluídos podem não ser levadas em conta. A partir de 2023, o DataSUS anunciou que vai obrigar o preenchimento de informações como raça e cor (EVANGELISTA, 2022).

Outro ponto de atenção diz respeito à consideração da variável étnica nos indicadores médicos como a oxigenação do sangue, por exemplo. Como o oxímetro tem sua medição afetada pela cor de pele do paciente, um sistema de IA que desconsidere este fator pode ocorrer em grave equívoco com potencial para causar a morte de um paciente.

A Inteligência Artificial Generativa também esbarra no tema do preconceito. Isto porque as LLMs utilizam grandes modelos de linguagem que se alimentam de *datasets* compostos por conteúdos disponíveis na internet: textos e imagens que podem conter viés preconceituoso. A resposta de um chatbot como o ChatGPT é um processamento de variadas fontes opacas. A ausência de “explicabilidade” e das fontes das informações dificulta a identificação de BIAS.

Esta situação é de mais fácil percepção nas aplicações de imagens como “Midjourney” e o DALL-E. Ao imputar um *prompt* solicitando a criação de uma imagem de um médico ou de um empresário de sucesso, a probabilidade do material gerado conter uma pessoa negra é muito menor, pois os sistemas foram treinados com material que reproduz este estereótipo.

O GPT-3 foi treinado a partir de cinco conjuntos de dados: Common Crawl, WebText2, Wikipedia, Books1 e Books2. Os três primeiros conjuntos de fontes são compostos por conteúdos variados da internet.

Já o GPT-4 já possui uma preocupação maior com textos preconceituosos, porém esta versão encontra-se apenas acessível a usuários *PLUS* do ChatGPT ao custo de 20 dólares mensais.

- Privacidade e Controle Social dos Dados

A coleta de dados é premissa para sistemas de inteligência artificial, contudo é preciso garantir os direitos do cidadão à privacidade e protegê-lo de usos não autorizados de suas informações pessoais.

Nos Estados Unidos, a rede de farmácias CVS está atuando fortemente no uso de dados de Saúde (DARIN, 2021). No Brasil, o Grupo DPSP (dono das drogarias Pacheco e São Paulo) opera no sentido de construir um marketplace médico pautado nos dados de clientes dentro da modalidade de *OpenHealth*.

O Google também voltou a investir na coleta de dados de saúde dos usuários Android (SOARES, 2021). Já a startup Drumwave trabalha na monetização de dados pessoais de usuários de serviços financeiros e de saúde.

Entretanto, sem a devida regulamentação, o uso destes dados pessoais pode causar medidas prejudiciais ao cidadão em processos como o tempo de carência e o reajuste de planos de saúde com base na probabilidade de uso do serviço (sinistros). Por isso, a Rede Nacional de Dados de Saúde (RNDS) possibilitaria a implementação do *Open Health* com controle social, um dos pilares do SUS.

Alinhado a isto, a Lei Geral de Proteção de Dados Pessoais (LGPD) impõe limites claros para os tipos de uso permitidos para os de Saúde. A Lei nº 13.709, de 14 de agosto de 2018 estabelece que:

II - dado pessoal sensível: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural;

[...]

Art. 7º O tratamento de dados pessoais somente poderá ser realizado nas seguintes hipóteses:

[...]

VIII - para a tutela da saúde, exclusivamente, em procedimento realizado por profissionais de saúde, serviços de saúde ou autoridade sanitária;

[...]

Art. 11. O tratamento de dados pessoais sensíveis somente poderá ocorrer nas seguintes hipóteses:

[...]

f) tutela da saúde, exclusivamente, em procedimento realizado por profissionais de saúde, serviços de saúde ou autoridade sanitária;

[...]

§ 4º É vedada a comunicação ou o uso compartilhado entre controladores de dados pessoais sensíveis referentes à saúde com objetivo de obter vantagem econômica, exceto nas hipóteses relativas a prestação de serviços de saúde, de assistência farmacêutica e de assistência à saúde, desde que observado o § 5º deste artigo,

incluídos os serviços auxiliares de diagnose e terapia, em benefício dos interesses dos titulares de dados [...]

§ 5º É vedado às operadoras de planos privados de assistência à saúde o tratamento de dados de saúde para a prática de seleção de riscos na contratação de qualquer modalidade, assim como na contratação e exclusão de beneficiários (BRASIL, 2018).

Na área da Saúde, os dados têm origem em fontes diversas: pessoais, clínicos, de procedimentos e de exames. De acordo com a Lei nº 13.787, de 27 de dezembro de 2018, os prontuários eletrônicos devem ser guardados por um prazo mínimo de 20 anos. O Marco Civil da Internet (Lei nº 12.965, de 23 de abril de 2014) também trata do tema do tema em seu artigo 11º:

Art. 11. Em qualquer operação de coleta, armazenamento, guarda e tratamento de registros, de dados pessoais ou de comunicações por provedores de conexão e de aplicações de internet em que pelo menos um desses atos ocorra em território nacional, deverão ser obrigatoriamente respeitados a legislação brasileira e os direitos à privacidade, à proteção dos dados pessoais e ao sigilo das comunicações privadas e dos registros (BRASIL, 2014).

A LGPD prevê a possibilidade de eliminação dos dados pessoais quando houver solicitação por parte de seu titular, contudo, devido à importância de seu uso em pesquisas, os dados de Saúde podem vir a ser considerados como um “bem público global”.

O acesso a dados em Saúde podem prever uma possível epidemia. No caso da Saúde Coletiva, o monitoramento de informações sociais e clínicas podem auxiliar na previsão do maior impacto de uma epidemia, atuando no aprimoramento das políticas públicas. Entretanto, o direito à Saúde e à privacidade devem caminhar juntos, sem um canibalizar o outro.

Por tal motivo, a segurança dos dados é um ponto fundamental. Os sistemas público e privado devem ter regras claras de quais dados podem e devem ser coletados, onde serão armazenados, quem pode acessá-los e para que serão utilizados. A partir disso, o acesso aos dados deve ser controlado de acordo com perfis específicos. É preciso dar atenção a medidas de segurança como criptografia para evitar vazamento dos dados como o ocorrido com o Hospital do Barreiro (Portugal):

O Hospital do Barreiro tem 296 médicos colocados, mas os sistemas internos permitiam que mais de 900 médicos continuassem com as contas de acesso a repositórios clínicos ativos. Contas de assistentes sociais, falhas no sistema de autenticação e a inexistência de regras de acesso também contribuíram para a aplicação de coimas ao abrigo do novo Regulamento Geral de Proteção de Dados (SÉNECA, 2018).

Caso similar ocorreu nos Países Baixos. O “Haga Hospital” foi multado em 460 mil euros por violar a GDPR (*General Data Protection Regulation*), lei europeia análoga à LGPD. O hospital possuía regras frágeis para o controle de acesso a dados de pacientes.

Além das informações clínicas, os dados genômicos também podem ser utilizados para prever a possibilidade de aparecimento de enfermidades. A “medicina preditiva” atua no diagnóstico de doenças hereditárias com base na análise de informações genéticas. A aposta da genômica clínica é promover tratamentos personalizados com o uso de medicamentos mais eficazes.

O acesso a dados do genoma humano ainda não é tão comum na medicina, contudo, com a crescente popularização de testes de ancestralidade, os próprios usuários geram dados genômicos pessoais que podem ser utilizados para outros fins até mesmo com autorização do titular a fim de supostamente tentar encontrar sua árvore genealógica.

Neste sentido, são os próprios usuários que autorizam o compartilhamento de suas informações pessoais de Saúde. O mesmo acontece com aplicativos ligados à qualidade de vida como Strava, Google Health, Garmin Connect, entre outros. Por meio de uma dinâmica, “gamificada”, os usuários destas ferramentas são engajados no compartilhamento cada vez maior de suas atividades cotidianas (esportes, sono e jornada de trabalho).

Por outro lado, mesmo com vedação legal, é preciso estar atento à tentativa de precificar planos de saúde suplementar a partir da probabilidade de uso futuro dos serviços médicos com base em informações pessoais. Conforme visto anteriormente, a Internet das Coisas Médicas (IoMT) possibilitará o monitoramento da Saúde em tempo real por meio de dispositivos conectados, tais quais os vestíveis (*wearable devices*). Pela lógica, uma pessoa desconfiaria se o plano de saúde tentasse acessar as informações dos *wearable devices*. Porém, situação similar ocorreu com a parceria entre farmácias e planos de saúde suplementar, que oferecem benefícios aos clientes em troca dos dados pessoais.

O acesso desregulado a dados como estes poderia ainda criar um cenário de interferência na contratação de funcionários, pois os recrutadores saberiam antecipadamente se uma pessoa está grávida ou se sofre de algum transtorno mental como depressão, bipolaridade ou pânico. Esta possibilidade é acentuada em populares sistemas de recrutamento online que utilizam inteligência artificial para a análise dos dados dos candidatos, como o Gupy que dispõe da IA Gaia.

Em 2023, a Sulamérica criou um sistema que chamou de “cashback” em parceria com o aplicativo de exercícios físicos Gympass. O usuário que comprovar ter feito atividades físicas (pelo menos duas vezes por semana durante seis meses) recebe, pela operadora, o reembolso do valor gasto com o aplicativo Gympass.

Em princípio, a “promoção” atuaria dentro da tentativa de mudança de cultura médica, migrando do “*sickcare*” para o “*healthcare*”, todavia possui potencial para acesso a dados pessoais de bem estar, deixando vulneráveis os usuários de planos de saúde suplementar na

medida em que suas informações pessoais podem ser levadas em conta no momento de reajuste, definição da carência ou da negativa de determinada cobertura.

- **Explicabilidade da IA**

Para assegurar a confiabilidade das informações fornecidas por aplicações de Inteligência Artificial, é fundamental conhecer o funcionamento dos algoritmos, contudo a maioria das ferramentas atua como caixas pretas (black boxes), tornando opacos e nada transparentes os critérios que levaram a máquina a chegar em determinado resultado. Na área financeira, por exemplo, um sistema de IA deve ser desenvolvido de forma a permitir que as avaliações de crédito sejam transparentes e permitam rastreabilidade.

O conceito da “explicabilidade” ou “interpretabilidade” diz respeito à capacidade de compreender os detalhes por trás do funcionamento de determinado algoritmo, explicando seu processo decisório.

Para o campo da saúde, apesar das enormes promessas de aplicações de Machine Learning, é fundamental entender o que motivou uma determinada decisão. Não parece razoável que médicos e pacientes recebam sem questionar diagnósticos que não podem ser explicados. Neste caso, a preocupação e busca por soluções tecnológicas “interpretáveis” são fundamentais não só para dar transparência para médicos e pacientes, mas para todas as partes interessadas, inclusive aos órgãos reguladores (LIMA, 2022).

A *Explainable Artificial Intelligence* (XAI) está, portanto, ligada à responsabilidade algorítmica e à transparência de processos de aprendizado de máquina (NUNES; MORATO, 2021).

Na versão 4 do GPT, a máquina adquiriu a capacidade de citar as fontes utilizadas para a geração da resposta, mas como está disponível apenas para os assinantes do plano *Plus*, a maioria dos usuários ainda enfrenta grande dificuldade em explicar as informações geradas pela IA.

Para o SUS, é fundamental que os sistemas de inteligência artificial funcionem dentro de uma mentalidade de *software livre*, com o compartilhamento não apenas de seus benefícios, mas também de abertura do “código”, medida que propiciaria o aprimoramento das ferramentas, o controle social da IA na Saúde, além do desenvolvimento de aplicações específicas derivadas.

O ideal é que pesquisa e ferramentas que utilizam dados públicos possam compartilhar não apenas gráficos e textos de suas conclusões, mas também o código utilizado para tal, criando um ecossistema de partilha de benefícios entre atores privados e públicos (*semicommons*), neste caso com *startups* e *healthtechs* de um lado e o SUS do outro.

INTELIGÊNCIA ARTIFICIAL GENERATIVA NA SAÚDE

Apesar de já abordado em tópicos anteriores, a partir daqui, lançaremos luz no uso de aplicações de Inteligência Artificial focadas em compreender, processar e gerar linguagem humana. Conforme explicado anteriormente, o NLP (“Processamento de Linguagem Natural” na tradução livre) é o ramo da Inteligência Artificial que abarca a interação entre seres humanos e máquinas por meio de linguagem natural. Já a IA Generativa é uma técnica que permite a geração de textos, fotos, vídeos e até música com base em grandes modelos de linguagem (LLM - *Large Language Models*).

Até recentemente, a NLP estava restrita a chatbots que apresentavam limitações nos processos de compreensão dos comandos humanos (NLU - *Natural-language understanding*) e na geração de respostas (NLG - *Natural-language generation*), tais como Alexa, Siri e Google Assistant.

O diferencial das novas ferramentas como o ChatGPT é a utilização de uma arquitetura de aprendizagem profunda (*Deep Learning*) baseada em “Transformers”, tecnologia desenvolvida pelo Google em 2017. O ChatGPT é nada mais que um chatbot baseado na tecnologia GPT (*Generative Pre-trained Transformer*). O GPT foi desenvolvido pela OpenAI, uma organização sem fins lucrativos cujo objetivo é popularizar o uso da Inteligência Artificial. O GPT também é o motor do DALL-E, ferramenta de IA que gera imagens sintéticas a partir de comandos de texto.

Para se ter uma ideia, o GPT-3 (lançado em 11 de julho de 2020) possui 175 bilhões de parâmetros de aprendizado de máquina.. Em comparação, o GPT-4, lançado em 14 de março de 2023, conta com 100 trilhões de parâmetros de aprendizado. Além disso, a versão mais recente agregou novas funcionalidades como a capacidade de compreender imagens, processar textos de até 25 mil palavras e possui maior preocupação com vieses e preconceitos. Atualmente, o GPT-4 está disponível apenas para assinantes do plano *PLUS* ao custo de 20 dólares mensais. O acesso gratuito dá direito apenas a versão 3.5 do GPT.

Na área da ciência de dados, a NLP pode atuar na estruturação de dados não-estruturados, enriquecendo as informações de bancos de dados a serem analisados pela Inteligência Artificial. Já a IA Generativa, ajudaria a romper a “síndrome da página branca”, constituindo-se como um *framework* para que a produção intelectual não parta do zero.

A IA Generativa impõe uma mudança nas habilidades fundamentais para o mundo profissional, sendo a metacognição uma delas. A capacidade de fazer boas perguntas (*prompt*) e de interpretar respostas pode vir a ser uma tendência num mundo com forte presença da IA Generativa, promovendo até mesmo o aparecimento de novas profissões com o “Engenheiro de Prompt”.

Embora ainda não se conheçam os limites do potencial do ChatGPT e de outras ferramentas de modelo de linguagem, os primeiros experimentos indicam que será necessário saber interagir com esses sistemas de inteligência artificial (IA) para

obter resultados satisfatórios. As novas aptidões moldariam uma profissão do futuro: o engenheiro de prompt. Nessa abordagem, prompt é a linha de comando na qual se digita, em texto corrente, a pergunta que o computador vai responder (STEIW, 2023).

- **Mídia sintética e Infodemia**

A principal característica da Inteligência Artificial Generativa é gerar textos, fotos e vídeos. Estas produções das ferramentas de IA são conhecidas como “mídia sintética”. Popularizada por aplicativos como o “Reface”, as primeiras experiências de *Deepfake* (troca de rostos e voz em vídeos) assustaram por demonstrar o potencial de sua utilização em ações de desinformação. Contudo o simples compartilhamento de textos não confiáveis gerados por LLM como o ChatGPT já pode promover tal situação.

Como o ChatGPT tem uma capacidade de processamento muito grande, pode produzir textos convincentes sobre diversos temas. O usuário, sem conseguir discernir a veracidade das informações, acaba compartilhando este material e os receptores o entendem como informação confiável.

O cenário acima já está em prática por diversos influenciadores da internet que utilizam o ChatGPT para produzir “dicas” para sua audiência sobre temas que não dominam. Tal situação é mais preocupante quando se trata da área da Saúde.

O canal no YouTube do influenciador Murilo Bevervanso possui quase 250 mil seguidores. No vídeo “Chatgpt cria videos do Youtube pra voce em alguns cliques Aprenda a como ganhar dinheiro No Youtube” (<https://youtu.be/zYsLmKtlyG8>), Murilo ensina a produção de dicas de emagrecimento com informações de nutrição, exercícios físicos e pressão arterial com base em consulta ao ChatGPT. Não há nas respostas do ChatGPT nenhuma indicação de fonte e nem confirmação de que as informações foram atestadas por profissionais de Nutrição e Educação Física.

No vídeo acima, o influenciador ainda mostra a como utilizar o texto gerado pelo ChatGPT para alimentar outra ferramenta de IA, a Synthesia, por onde é possível criar um avatar virtual e uma locução sintética para um vídeo que será postado no YouTube com o objetivo de influenciar pessoas.

Já no canal “Ganhando no Automático”, o influenciador Marcos Teixeira sugere o uso do ChatGPT em conjunto com a ferramenta de “Invideo”. No vídeo “Como criar vídeo em segundos com inteligência artificial grátis e zero de experiência” (<https://youtu.be/8wrnG4q0jgM>), Marcos mostra que o aplicativo “Invideo” processa o texto gerado pelo ChatGPT e monta automaticamente um vídeo-legenda com imagens relacionadas aos temas abordados.

Na mesma linha, o canal “Gus.A.I” perguntou ao ChatGPT quais os melhores assuntos para um canal do youtube e recebeu a indicação de “saúde” entre uma das possibilidades.

Também no vídeo “Como criar Videos Youtube com ChatGPT-OpenAI / 100% Automático com Edição de Inteligência Artificial” (https://youtu.be/Vqvc_xBoqkE), o youtuber ensina a usar a ferramenta “Pictory”, um aplicativo de edição de vídeo baseado em IA que monta automaticamente um vídeo-legenda a partir de textos do ChatGPT.

Os três casos acima acendem o sinal amarelo para a criação de mídia sintética em massa a partir da integração de serviços de IA Generativa, principalmente na área da Saúde.

Há iniciativas no sentido de aprimorar a base de dados com a qual a IA Generativa é treinada. O GPT-3, por exemplo, é treinado lançando mão das bases Common Crawl, WebText2, Wikipedia, Books1 e Books2. A proposta alternativa seria limitar este universo apenas a textos acadêmicos, como testado pela ferramenta “Consensus”. Entretanto, parte dos artigos científicos encontra-se em repositórios fechados (situação contrária à *Open Science*), além disso existe o entendimento de que as publicações científicas apresentam visões sobre temas e não constituem uma verdade universal, sendo assim, mesmo em artigos revisados pelos pares, as conclusões precisam ser interpretadas pelo leitor como uma tese.

No mesmo sentido do Consensus, o investidor estadunidense Elon Musk anunciou o desenvolvimento da ferramenta TruthGPT, que teria como diferencial uma pretensa busca da verdade. Musk foi um dos signatários da carta do instituto Future of Life que pede uma pausa nos experimentos com Inteligência Artificial.

No campo da Saúde, esta preocupação é maior pois a desinformação pode resultar em mortes direta ou indiretamente, como constatado durante a Infodemia no contexto da Covid-19. Como vimos, há no Brasil uma desigualdade no acesso à internet reforçada pelos planos “*zero-rating*”, contribuindo para que a predominância das informações seja acessada via mídias sociais pelas classes C, D e E. Somada a isto, a descrença com meios de comunicação tradicionais e a popularização dos *influencers* (influenciadores digitais) cria um ambiente propício para a dinâmica da “apomediação” (SOUZA, 2020) (*apomediaction*), apontada pela pesquisadora Nicole Fajardo Maranhã Leão de Souza, ao analisar o comportamento de um grupo de estudantes que optava preferencialmente por informar-se pelas recomendações de outros usuários da internet.

Entre os alunos do colégio de Manguinhos, também parece ser comum a prática da busca de informações sobre saúde no Google ou no Youtube. No questionário, a maior parte indicou o Google como a fonte mais confiável para informações sobre saúde na internet. Outra atitude frequente entre esses jovens, conforme eles relataram no grupo focal, é consultar a avaliação e a opinião de outros usuários da Internet, seja para julgar a confiabilidade de uma fonte de informação ou para orientar-se sobre o uso de determinado recurso ou aplicativo. Na opinião deles, o que estiver bem avaliado por outros usuários, merece confiança (SOUZA, 2020).

Em paralelo à penetração da IA Generativa, é preciso apostar em políticas públicas que promovam a educomunicação, trabalhando o desenvolvimento de habilidades de leitura crítica da mídia. A Finlândia, por exemplo, é considerada como um país sem fake news, pois incluiu no currículo escolar a alfabetização midiática (BENKE; SPRING, 2022).

Mesmo tendo sua popularização apenas recentemente, a IA Generativa já causou alguns casos de desinformação em escala global. Imagens sintéticas do Papa Francisco vestindo um casaco fashion e do ex-presidente Donald Trump sendo detido viralizaram na internet. Muitas pessoas acreditaram nas imagens que já contavam com grande realismo após a atualização da ferramenta Midjourney. A polêmica trouxe novos usuários para o aplicativo que decidiu manter seu funcionamento apenas para assinantes a um preço mínimo de oito dólares mensais.

Mais um ponto de atenção reside na possibilidade de o ChatGPT exibir anúncios em suas respostas. Ainda não está claro como esta funcionalidade será implementada, mas a mesma deve ocorrer com muita precaução para que as respostas orgânicas não sejam confundidas com as publicidades, assim como ocorre na busca do Google, onde há grande destaque aos anúncios e a diferenciação visual é pouco perceptível.

A situação pode ser ainda mais grave se nas respostas houver privilégio à consulta de textos de determinado anunciante, dentro de uma estratégia de Marketing de Conteúdo. A título de comparação, o conflito de interesses seria análogo caso outras ferramentas de IA como Waze e Google Maps alterassem seus resultados para que a rota passasse por algum comércio anunciante. Com a Internet das Coisas (IoT), tal situação ficaria ainda mais crítica, pois o sistema de IA teria acesso a dados como o combustível restante e a agenda (consultando festas e indicando possibilidade de compra de presentes), fazendo com que o motorista pare em determinado shopping ou posto de gasolina a partir da influência do aplicativo

- Piloto automático x Copiloto

A euforia inicial com a IA Generativa levou à crença de que estas ferramentas substituiriam os seres humanos, numa espécie de “piloto automático”, contudo uma dinâmica que tem ganhado força leva a crer que a IA se portará mais como um “copiloto”. Este conceito ajuda a entender a relação entre humanos e máquinas, na medida em que os copilotos atuam no auxílio do comandante. Esta delegação de tarefas ao “copiloto” não retira o protagonista do piloto, mas permite que ele se engaje nas atividades mais importantes.

A existência de um copiloto em um avião aumenta a segurança em momentos críticos, a detecção de problemas e a indicação de possíveis caminhos para sua resolução. O copiloto é um auxiliar do comandante. Todas as suas sugestões são avaliadas e decididas pelo piloto.

Deste modo, as projeções de ficção científica em que os profissionais seriam substituídos por avatares sintéticos automatizados não é válida por hora. Na área da Saúde, a IA Generativa possibilitaria que o médico gaste mais tempo na comunicação com o paciente e menos tempo

preenchendo prontuários e relatórios graças à capacidade da máquina em analisar o conteúdo da conversa através de técnicas como *speech analytics* e *text mining*.

As ferramentas de IA Generativa, portanto, ajudariam no preenchimento automático de prontuários e outros documentos. Grosso modo, seria como se estivesse pedindo para um residente esboçar um relatório que seria revisado posteriormente pelo médico.

A capacidade de síntese da IA Generativa também pode reduzir as falhas de comunicação durante os "*handoffs*" de pacientes, momento em que há transferência de médico, de turno, de hospital ou de unidade (UTI, quarto, etc), garantindo um elevado nível de informação entre todas as equipes que acompanham a jornada do paciente. O paciente pode ainda ter seus exames explicados por meio de linguagem personalizada, proporcionando uma facilitação na literacia em Saúde.

A própria Educação já está se adaptando à chegada da IA Generativa. Ao invés de proibir o uso do ChatGPT em trabalhos escolares, alguns professores propõem que os alunos gerem textos pela plataforma e revisem as respostas dadas pelo robô a fim de encontrar erros nos textos produzidos por IA.

O livro "The AI Revolution in Medicine GPT-4 and Beyond", de Peter Lee, Carey Goldberg e Isaac Kohane, traz uma cena hipotética de um residente sendo orientado pelo GPT-4 durante uma situação de emergência. Ainda é muito cedo para prever todas as possibilidades de uso da IA Generativa na Saúde, mas numa área tão sensível, é preciso estar atento aos limites desta tecnologia, conforme veremos a seguir.

- **Limitações da IA Generativa**

O ChatGPT atingiu bons resultados no UMSLE, prova obrigatória para o exercício da profissão médica nos Estados Unidos (KUNG et al., 2023). Isto quer dizer que ele é suficientemente capaz de substituir um médico? A resposta é "não" e vamos explicar o porquê.

Primeiramente, por hora, as ferramentas de IA tendem a gerar bons resultados quando focadas em apenas uma tarefa e tendo como base uma grande quantidade de exemplos pré-treinados. No entanto, a Inteligência Artificial não é tão eficiente quando precisa executar uma tarefa complexa a partir de poucos dados.

A IA Generativa é calcada no *machine learning*, processo em que as máquinas aprendem sem serem programadas. Há métodos de aprendizagem supervisionados, não supervisionados e por reforço, porém ainda são ferramentas em processo de aprendizado, sendo que suas respostas devem sempre ser validadas posteriormente. Analogamente, seria como esperar respostas científicas definitivas de um estudante universitário na iniciação científica. Mesmo com acesso a livros da biblioteca e de repositórios online, sua capacidade cognitiva para a pesquisa científica ainda está em formação, podendo ser amadurecida em etapas posteriores como mestrado, doutorado e pós-doutorado.

O GPT utiliza o aprendizado profundo (*deep learning*) a partir da arquitetura Transformer, porém lança mão de algumas técnicas que explicam o aparecimento de respostas equivocadas ou a pura “alucinação” (*Hallucination*).

Por se tratar de um bate-papo, o ChatGPT precisa dar respostas rápidas. “A pressa é inimiga da perfeição” é um famoso ditado popular que ajuda a entender as consequências do uso das técnicas de *Zero-shot*, *One-shot* e *Few-shot* a partir do GPT-2. Ao se deparar com perguntas para as quais não tem nenhuma referência (*Zero-Shot*), apenas uma referência (*One-shot*) ou poucas referências (*Few-shot*), a ferramenta tenta prever uma resposta correta. É como se fôssemos perguntar para crianças sobre assuntos que desconhecem. Vão dar uma resposta, mas com elevado potencial de equívoco.

Este processo é muito natural em aplicações de inteligência artificial quando operadas por cientistas de dados em processos de treinamento por *reinforcement learning* e *supervised learning*. Por outro lado, estas questões não estão visíveis para o público leigo que muitas vezes entende a IA Generativa como uma espécie de oráculo.

Os sistemas de Inteligência Artificial possuem métricas para avaliar a tolerância de confiabilidade: acurácia, precisão, interpretabilidade e F1 score. A transparência destas métricas é fundamental para verificar a validade das informações.

Na mesma linha, a revista Nature proibiu que o ChatGPT seja creditado como autor de artigos científicos. Isto porque um dos princípios de autoria de artigos científicos é a responsabilidade pelo teor do texto e isto o ChatGPT não pode fazer.

O ChatGPT é uma ferramenta generalista e a tendência é a criação de aplicativos de LLM especializados em segmentos como o da Saúde. Neste caso, os profissionais devem entender a ferramenta não como uma solução pronta, mas sim como um projeto em constante construção e evolução (mentalidade do *software livre*). Requer, portanto, uma ação humana no sentido de detectar erros, propor melhorias, acompanhar a performance e avaliar os resultados.

LITERACIA EM SAÚDE E TELESSAÚDE

A Literacia em Saúde (*Health Literacy*) tem ação fundamental no papel ativo do cidadão no âmbito da Saúde, deixando de ser entendido como um mero coadjuvante dentro de uma postura passiva em relação aos profissionais, mas sim dando segurança a seu protagonismo e coparticipação nos processos envolvidos.

A literacia em saúde pode ser definida como um amplo e diverso conjunto de habilidades e competências que os indivíduos utilizam para buscar, compreender, avaliar e dar sentido a informações sobre saúde, visando ao cuidado de sua

própria saúde ou de terceiros. Essa ampla gama de habilidades (PERES et al., 2021).

- **Literacia Digital em Saúde**

Já a Literacia Digital em Saúde (*eHealth literacy* - eHL) consiste na articulação de seis domínios distintos: literacia tradicional e numérica; literacia Digital; Literacia midiática; literacia em informação; Literacia Científica; e Literacia em Saúde.

O conjunto completo de habilidades relacionadas à eHL, portanto, compreenderia competências relacionadas não apenas à saúde, mas também à alfabetização tradicional (ou básica), conceitos numéricos, computadores, mídia, ciência e informação. Possuir todas essas habilidades integradas da eHL permitiria aos usuários obter resultados positivos não apenas no comportamento relacionado à busca de informações, mas também uma maior probabilidade de se envolver em experiências proativas relacionadas à saúde (SOUZA, 2020).

Antes passivos, aguardando a ação dos profissionais da Saúde, os pacientes informados ou pacientes digitais estão engajados na busca ativa por informações em Saúde disponíveis na internet. Estas informações podem vir da busca do Google (fenômeno conhecido popularmente como Dr. Google), das redes sociais e de mensagens do WhatsApp.

A construção da RNDS e a implementação do *Open Health* no Brasil facilita o acesso do cidadão a seus dados de Saúde, antes restritos a laboratórios, consultórios e hospitais. A partir do empoderamento do paciente com o domínio de suas informações, o SUS caminharia para a construção de uma “tecnodemocracia”.

A proposta de uma tecnodemocracia, que tenha por objeto de seu exercício a definição de uma PMIS, representa uma tentativa de dar materialidade à idéia da apropriação pela sociedade de um aparato até então quase que de uso restrito pelos dispositivos de vigilância, de controle de gestão da vida, cuja chave de acesso é de domínio reservado de uma elite científica e técnica, política e econômica (MORAES, 1998).

Este empoderamento do paciente cria uma relação mais horizontal com o médico, tendo o profissional de Saúde vindo a desempenhar mais um papel de tutor na atenção primária com a coparticipação do paciente.

A prática de Telessaúde sofreu grande impulso devido à pandemia de Covid-19 e foi autorizada por meio da Lei nº 13.989, de 15 de abril de 2020. Posteriormente, a Lei nº 14.510, de 27 de dezembro de 2022, veio a regulamentar definitivamente a prática da Telessaúde no Brasil alterando a Lei do SUS (Lei nº 8.080 de 1990).

As ferramentas de IA Generativa também ajudam a intensificar a relação entre médico e paciente durante as consultas. Ao ter automatizado o preenchimento do prontuário, o profissional da Saúde pode privilegiar a conversa com o paciente, na medida em que os dados serão estruturados no banco de dados pela IA a partir do processamento em tempo real do áudio da conversa no consultório.

Como a perspectiva de que a IA não seja autônoma, ao final da conversa, o médico revisaria as informações sistematizadas pela ferramenta, corrigindo, se necessário, algum equívoco no entendimento antes da finalização da consulta.

A partir do carregamento de dados pessoais históricos de Saúde, a IA já poderia indicar em tempo real perguntas para o médico fazer ao paciente. No caso de esquecer de alguma informação protocolar, o sistema sinalizaria o fato ao profissional.

Por conta da confidencialidade das informações do paciente, é fundamental que os sistemas de NLP que vão acompanhar a conversa durante a consulta forneçam garantias seguras de que o áudio não será gravado, armazenado ou utilizado para outros fins senão aqueles autorizados.

A adoção da IA Generativa nos consultórios não será mandatória, porém com a melhoria na comunicação daqueles que optarem por seu uso, esta será uma reivindicação natural dos pacientes, assim como os recursos de agendamento online, receita digital e teleconsulta. O nível de literacia tende a ser um dos fatores de escolha do profissional de Saúde, na medida em que é baixo quando o paciente nem ao menos consegue entender o nome do medicamento e a posologia indicada na prescrição médica por conta da letra escrita no papel em dinâmicas de atendimento não digitalizadas.

CONSIDERAÇÕES FINAIS

Este artigo fez um panorama geral do atual uso de ferramentas de Inteligência Artificial, em especial as generativas, na área da Saúde. Essa tecnologia tem o potencial de provocar uma nova revolução industrial e sua rápida adoção exige um esforço para garantir que o impacto da IA na sociedade seja regulado com participação social e transparência. A preocupação reside no fato de ainda serem desconhecidos os impactos da IA na sociedade, por isso iniciativas como a carta do instituto Future of Life dialogam com o “Princípio da precaução”, muito comum na área ambiental, medida que assegura a sustentabilidade do planeta.

Além da desinformação, como visto anteriormente, o desemprego estrutural é uma ameaça a trabalhadores cujos ofícios têm potencial para serem automatizados pela Inteligência Artificial. A carta do instituto Future of Life elenca algumas destas demandas para o enfrentamento desse cenário:

Sistemas poderosos de IA devem ser desenvolvidos apenas quando estivermos confiantes de que seus efeitos serão positivos e seus riscos serão administráveis.

Essa confiança deve ser bem justificada e aumentar com a magnitude dos efeitos potenciais de um sistema.

[...] os desenvolvedores de IA devem trabalhar com os formuladores de políticas para acelerar drasticamente o desenvolvimento de sistemas robustos de governança de IA. Estes devem incluir, no mínimo: autoridades reguladoras novas e capazes dedicadas à IA; supervisão e rastreamento de sistemas de IA altamente capazes e grandes conjuntos de capacidade computacional; sistemas de proveniência e marca d'água para ajudar a distinguir o real do sintético e rastrear vazamentos de modelos; um ecossistema robusto de auditoria e certificação; responsabilidade por danos causados pela IA; financiamento público robusto para pesquisa técnica de segurança de IA; e instituições com bons recursos para lidar com as dramáticas perturbações econômicas e políticas (especialmente para a democracia) que a IA causará (FUTURE OF LIVE, 2023, tradução nossa).

Algumas das sugestões já estão em processo de incorporação pelas ferramentas dentro do estabelecimento de boas práticas para a IA Generativa, tais quais a citação das fontes utilizadas e o treinamento de IA apenas utilizando como base material previamente autorizado.

Estas ações por parte das próprias empresas dialogam com o “Dever de Cuidado”, situação em que as ferramentas se responsabilizam pelas consequências do mau uso das mesmas. Este pensamento fez com que gigantes como Google e Facebook não lançassem como produto, até então, as ferramentas de inteligência artificial que desenvolviam em seus laboratórios.

A popularização do ChatGPT, entretanto, causou uma corrida desenfreada pelo desenvolvimento de novas aplicações de IA, fato que coloca um ponto de atenção sobre as questões éticas envolvidas, principalmente a partir da demissão em massa das equipes responsáveis por esta análise em alguns das principais *Big Techs* (CRIDDLE, 2023).

A autorização de uso de aplicações de IA na área médica já foi normatizada no Brasil pela Anvisa, mas para o caso das IAs Generativas focadas para o público em geral, a certificação de ferramentas que utilizam repositórios confiáveis seria fundamental a fim de evitar desinformação. Já existem selos como o HONcode e o “Selo Sergio Arouca de Qualidade da Informação em Saúde na Internet”. Quanto aos repositórios, podemos citar o ARCA para artigos científicos e o PCDaS para dados estruturados.

Assim como os curadores da Wikipedia ajudam a prevenir que a plataforma reproduza preconceitos, informações equivocadas de colaboradores mal intencionados e propaganda disfarçada de conteúdo informativo, é preciso estabelecer mecanismo de aprendizagem por reforço a partir de feedbacks por parte de estudiosos no tema.

Para além da marca d'água sugerida pelo instituto Future of Life, outra ação importante seria no sentido de obrigar as ferramentas a incluir um identificador nos metadados de fotos e vídeos gerados por IA, assim como já acontece nas câmeras fotográficas que registram no

arquivo informações como modelo da câmera, abertura de íris , lente utilizada, entre outros dados. Em sentido similar, o rastreamento digital dos conteúdos gerados via IA poderia ser verificado por meio da Blockchain, em funcionamento similar à geração de NFTs. Se tal medida infringir os ditames da LGPD, a verificação pode ser apenas booleana, retornando somente dois valores: 1 - Produzido por IA; 2 - Não produzido por IA.. Este sistema de verificação funcionaria como um plataforma de *Fact checking*, evitando assim a desinformação a partir da disseminação de mídia sintética ou *deepfake*.

Também fundamental é a transparência sobre o funcionamento da IA, condição chamada de “explicabilidade” ou “interpretabilidade”. Adicionalmente, as respostas de IA deveriam exibir as métricas para avaliação de sua confiabilidade, como acurácia e precisão.

Ainda na mesma linha, a adoção de padrões abertos permitiria a interoperabilidade das informações que devem ser entendidas como um bem público, sejam elas de *Open Health* ou *Open Science*.

Por último, o papel da regulação legal/infralegal, garantindo que os direitos fundamentais do cidadão sejam assegurados como o Direito à Saúde e o Direito à Comunicação, tão caros numa sociedade desigual como a brasileira.

O Projeto de Lei 21/2020 objetiva criar um Marco Legal da Inteligência Artificial. O texto inicial previa a criação de um “relatório de impacto”, além da responsabilização legal dos agentes de IA, sejam eles desenvolvedores como também as empresas que implementam as aplicações. Mesmo sendo aprovado pela Câmara dos Deputados em 2021, não seguiu adiante e já é considerado obsoleto.

Por este motivo, um grupo de juristas formou uma comissão para construir um anteprojeto em 2022 que fosse atemporal e não se tornasse obsoleto com a evolução tecnológica. Além do PL 21/2020, o anteprojeto visa substituir outros projetos de lei (PL 5.051/2019 e PL 872/2021) que tratam do mesmo tema. A proposta é que, com a regulamentação, as ferramentas possam “obedecer a princípios como liberdade de escolha, transparência, rastreabilidade de decisões, responsabilização, reparação de danos, inclusão e não-discriminação” (FRAGOSO, 2022), assegurando a explicabilidade e a privacidade de que tratamos ao longo deste artigo.

Esta última proposta foi apresentada oficialmente pelo presidente do Senado Federal, Rodrigo Pacheco, na forma do Projeto de Lei nº 2338/2023.

Ainda estamos no início da adoção da IA pela sociedade em geral e esperamos que os princípios e diretrizes do SUS pautem o uso da Inteligência Artificial na Saúde, promovendo universalização, equidade, integralidade, regionalização, hierarquização, descentralização e participação social, além de associar sua implementação no SUS ao desenvolvimento tecnológico pelo Complexo Econômico-Industrial da Saúde (CEIS) como um dos principais pilares da Quarta Revolução Tecnológica.

REFERÊNCIAS

ALMEIDA, C. V.; MORAES, K. L.; BRASIL, V. V. **50 Técnicas Literacia em Saúde na Prática: Um Guia para a Saúde**. Novas Edições Acadêmicas, 2020.

ANS & AMB. **Diretrizes Clínicas na Saúde Suplementar 2012**: Associação Médica Brasileira e Agência Nacional de Saúde Suplementar. Disponível em: https://bvsmms.saude.gov.br/bvs/publicacoes/diretrizes_clinicas_2012.pdf. Acesso em: 25 abr. 2023.

ANVISA. RESOLUÇÃO DE DIRETORIA COLEGIADA - RDC Nº 657, DE 24 DE MARÇO DE 2022 - RESOLUÇÃO DE DIRETORIA COLEGIADA - RDC Nº 657, DE 24 DE MARÇO DE 2022 - **DOU - Imprensa Nacional**. Disponível em: <https://www.in.gov.br/en/web/dou/-/resolucao-de-diretoria-colegiada-rdc-n-657-de-24-de-marco-de-2022-389603457>. Acesso em: 25 abr. 2023.

BENKE, E.; SPRING, M. Como a Finlândia tem conseguido derrotar as “fake news”. **BBC News Brasil**, 25 out. 2022. Disponível em: <https://www.bbc.com/portuguese/internacional-63390825>. Acesso em: 25 abr. 2023.

BRASIL. Decreto nº 11.358, de 1º de janeiro de 2023. Altera o Decreto nº 11.358, de 1º de janeiro de 2023, que aprova a Estrutura Regimental e o Quadro Demonstrativo dos Cargos em Comissão e das Funções de Confiança do Ministério da Saúde. **Diário Oficial da União**, Brasília, DF, 2023. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2023-2026/2023/decreto/D11391.htm. Acesso em: 25 abr. 2023.

BRASIL. Portaria nº 1.434, de 28 de maio de 2020. Altera o Anexo XLII da Portaria de Consolidação GM/MS nº 2, de 28 de setembro de 2017, para dispor sobre a Política Nacional de Informação e Informática em Saúde (PNIIS). **Diário Oficial da União**, Brasília, DF, 2021. Disponível em: <https://www.in.gov.br/en/web/dou/-/portaria-gm/ms-n-1.768-de-30-de-julho-de-2021-335472332>. Acesso em: 25 abr. 2023.

BRASIL. Lei nº 12.965, de 23 de abril de 2014. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil. **Diário Oficial da União**, Brasília, DF, 2014. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm. Acesso em: 25 abr. 2023.

BRASIL. Lei nº 14.510, de 27 de dezembro de 2022. Altera a Lei nº 8.080, de 19 de setembro de 1990, para autorizar e disciplinar a prática da tele-saúde em todo o território nacional, e a Lei nº 13.146, de 6 de julho de 2015; e revoga a Lei nº 13.989, de 15 de abril de 2020. **Diário Oficial da União**, Brasília, DF, 2022. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2022/lei/L14510.htm. Acesso em: 25 abr. 2023.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). **Diário Oficial da União**, Brasília, DF, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 25 abr. 2023.

BRASIL. Lei nº 8.080, de 23 de abril de 2014. Dispõe sobre as condições para a promoção, proteção e recuperação da saúde, a organização e o funcionamento dos serviços correspondentes e dá outras providências. **Diário Oficial da União**, Brasília, DF, 1990. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l8080.htm. Acesso em: 25 abr. 2023.

BRASIL. Lei nº 13.787, de 27 de dezembro de 2018. Dispõe sobre a digitalização e a utilização de sistemas informatizados para a guarda, o armazenamento e o manuseio de prontuário de paciente. **Diário Oficial da União**, Brasília, DF, 2018. Disponível em:

https://www.planalto.gov.br/ccivil_03/ato2015-2018/2018/lei/l13787.htm. Acesso em: 25 abr. 2023.

BRASIL. Decreto n° 8.771, de 11 de maio de 2016. Regulamenta a Lei n° 12.965, de 23 de abril de 2014, para tratar das hipóteses admitidas de discriminação de pacotes de dados na internet e de degradação de tráfego, indicar procedimentos para guarda e proteção de dados por provedores de conexão e de aplicações, apontar medidas de transparência na requisição de dados cadastrais pela administração pública e estabelecer parâmetros para fiscalização e apuração de infrações. **Diário Oficial da União**, Brasília, DF, 2016. Disponível em: https://www.planalto.gov.br/ccivil_03/ato2015-2018/2016/decreto/d8771.htm. Acesso em: 25 abr. 2023.

BRASIL. Senado Federal. Projeto de Lei n° 2.338, de 3 de maio de 2023. Dispõe sobre o uso da Inteligência Artificial. Brasília: Senado Federal, 2013. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>. Acesso em: 9 mai. 2023.

CETIC. **TIC Saúde 2021**. Disponível em: <https://cetic.br/pt/tics/saude/2021/estabelecimentos/B2/>. Acesso em: 25 abr. 2023.

CRIDDLE, C.R. Big techs geram preocupação ao demitir equipes que cuidavam de ética em IA. **Folha de S. Paulo**, 29 mar. 2023. Disponível em: <https://www1.folha.uol.com.br/tec/2023/03/big-techs-geram-preocupacao-ao-demitir-equipes-que-cuidavam-de-etica-em-ia.shtml>. Acesso em: 25 abr. 2023.

D'AGOSTINI, J. **Zero-rating, desinformação e precariedade do acesso à Internet no Brasil**. **LAPIN**, 2 ago. 2021. Disponível em: <https://lapin.org.br/2021/08/02/zero-rating-e-desinformacao-a-relacao-entre-a-precariedade-do-ace-sso-a-internet-no-brasil-e-a-disseminacao-de-conteudos-enganosos/>. Acesso em: 25 abr. 2023

DARIN, B. **CVS Health Paves Path to Better Care With Data, AI - WSJ**. Disponível em: <https://deloitte.wsj.com/articles/cvs-health-paves-path-to-better-care-with-data-ai-01615406532>. Acesso em: 25 abr. 2023.

Ep.129 - O que é Chat GPT e qual seu uso na saúde?. [Locução de:] Fernanda Velloni, Felipe Kitamura, Giancarlo Domingues. [S. l.]: DASA EDUCA, 23 fev. 2023. *Podcast*. Disponível em: <https://www.youtube.com/watch?v=CfWTbzK3V-s>. Acesso em: 25 abr. 2020.

DIAMANDIS, P. H. **The 6 D's**. Disponível em: <https://www.diamandis.com/blog/the-6ds>. Acesso em: 25 abr. 2023.

EVANGELISTA, A. P. **Sistemas do SUS não têm dados suficientes de raça/cor da pele de pacientes durante a pandemia**. Disponível em: <https://www.epsjv.fiocruz.br/podcast/sistemas-do-sus-nao-tem-dados-suficientes-de-racacor-da-pele-e-de-pacientes-durante-a-pandemia>. Acesso em: 25 abr. 2023.

FRAGOSO, R. Comissão de juristas aprova anteprojeto do marco legal da inteligência artificial. **Rádio Senado**, 1 dez. 2022. Disponível em: <https://www12.senado.leg.br/noticias/audios/2022/12/comissao-de-juristas-aprova-anteprojeto-do-marco-legal-da-inteligencia-artificial>. Acesso em: 25 abr. 2023.

FUTURE OF LIFE. **Pause Giant AI Experiments: An Open Letter - Future of Life Institute**. Disponível em: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>. Acesso em: 23 abr. 2023.

GADELHA, C. A. G.; TEMPORÃO, J. G. **Desenvolvimento, inovação e saúde: a perspectiva teórica e política do Complexo Econômico-Industrial da Saúde**. *Ciência & Saúde Coletiva*, Rio de Janeiro, v. 23, n. 6, p.1891-1902, 2018. DOI: <https://doi.org/10.1590/1413-81232018236.06482018>.

GADELHA, C. A. G. **O Complexo Econômico-Industrial da Saúde 4.0**: por uma visão integrada do desenvolvimento econômico, social e ambiental. Cadernos do Desenvolvimento, Rio de Janeiro, v. 16, n. 28, p. 25-49, 2021. Disponível em: <http://www.cadernosdodesenvolvimento.org.br/ojs-2.4.8/index.php/cdes/article/view/550>. Acesso em: 25 abr. 2023.

GADELHA, C. A. G. (org.). **Saúde é Desenvolvimento**: o Complexo Econômico-Industrial da Saúde como opção estratégica nacional. Rio de Janeiro: Fiocruz - CEE, 2022.

GOLDIM, J. R. **Princípio da Precaução**. Disponível em: <https://www.ufrgs.br/bioetica/precau.htm>. Acesso em: 25 abr. 2023.

IDEC. **Modelo de internet restrito prejudica acesso a direitos básicos, diz pesquisa**. Disponível em: <https://idec.org.br/noticia/modelo-de-internet-restrito-prejudica-acesso-direitos-basicos-diz-pesquisa>. Acesso em: 25 abr. 2023.

KUNG, T. H. et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. **PLOS Digital Health**, v. 2, n. 2, p. e0000198, 9 fev. 2023.

LEE, P; GOLDBERG, C; KOHANE, I. **The AI Revolution in Medicine**: GPT-4 and Beyond. Pearson Education; 2023.

LIMA, J. DA C. **Desafios para a adoção de Inteligência Artificial pelo Sistema Único de Saúde (SUS): ética, transparência e interpretabilidade**. Tese (Doutorado em Informação e Comunicação em Saúde) - Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, 2022.

LIRA, M. **Sespa avança com o Projeto Telemedicina Pará em 40 municípios – SESPA**. , 3 jun. 2022. Disponível em: <http://www.saude.pa.gov.br/sespa-avanca-com-o-projeto-telemedicina-para-em-40-municipios/>. Acesso em: 25 abr. 2023

MORAES, I. H. S. DE. **Informações em saúde: para andarilhos e argonautas de uma tecnodemocracia emancipadora**. 1998. 285 f. Tese (Doutorado em Saúde Pública) - Escola Nacional de Saúde Pública Sergio Arouca, Fundação Oswaldo Cruz, Rio de Janeiro, 1998.

NATURE. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. **Nature**, v. 613, n. 7945, p. 612–612, 24 jan. 2023.

NUNES, D.; MORATO, O. **A explicabilidade da inteligência artificial e o devido processo tecnológico**. Disponível em: <https://www.conjur.com.br/2021-jul-07/opinioao-explicabilidade-ia-devido-processo-tecnologico>. Acesso em: 25 abr. 2023.

PERES, F. **A LITERACIA EM SAÚDE NO CHATGPT: EXPLORANDO O POTENCIAL DE USO DE INTELIGÊNCIA ARTIFICIAL PARA A ELABORAÇÃO DE TEXTOS ACADÊMICOS**. SciELO Preprints, 2023. DOI: 10.1590/SciELOPreprints.5658. Disponível em: <https://preprints.scielo.org/index.php/scielo/preprint/view/5658>. Acesso em: 18 maio. 2023.

PERES, F; Rodrigues, K.M.; Silva, T.S.. **Literacia em saúde**. Rio de Janeiro: Ed. FIOCRUZ; 2021.

PIGATTO, F. **RESOLUÇÃO Nº 659, DE 26 DE JULHO DE 2021**. Disponível em: <http://conselho.saude.gov.br/resolucoes-cns/1922-resolucao-n-659-de-26-de-julho-de-2021>. Acesso em: 25 abr. 2023.

SÉNECA, H. **Visão | CNPD: Hospital do Barreiro multado em 400 mil euros por permitir acessos indevidos a processos clínicos.** Disponível em: <https://visao.sapo.pt/exameinformatica/noticias-ei/mercados/2018-10-19-cnpd-hospital-do-barreiro-multado-em-400-mil-euros-por-permitir-acessos-indevidos-a-processos-clinicos/>. Acesso em: 23 abr. 2023.

SILVEIRA, S. A. DA; SOUZA, J.; CASSINO, J. F. (orgs.). **Colonialismo de dados: como opera a trincheira algorítmica na guerra neoliberal.** São Paulo: Autonomia Literária, 2021.

SOARES, L. Google volta a tentar coletar dados de saúde dos usuários. **Olhar Digital**, 12 abr. 2021. Disponível em: <https://olhardigital.com.br/2021/04/12/medicina-e-saude/google-volta-a-tentar-coletar-dados-de-saude-dos-usuarios/>. Acesso em: 25 abr. 2023.

SOUZA, N. F. M. L. DE. **eHealth literacy entre jovens: estudo exploratório sobre o papel das condições socioeconômicas no uso da informação sobre saúde na Internet.** 2020. 195 f. Dissertação (Mestrado em Informação e Comunicação em Saúde) - Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, 2020.

STEIW, L. **Com o ChatGPT, a engenharia de prompt desponta como uma nova profissão. Insper: Ensino Superior em Negócios, Direito, Engenharias e Ciência da Computação**, 7 mar. 2023. Disponível em: <https://www.insper.edu.br/noticias/com-o-chatgpt-a-engenharia-de-prompt-desponta-como-uma-nova-profissao/>. Acesso em: 25 abr. 2023.

THORP, H. H. **ChatGPT is fun, but not an author | Science.** Disponível em: <https://www.science.org/doi/10.1126/science.adg7879>. Acesso em: 23 abr. 2023.

WIGGERS, K. With Firefly, Adobe gets into the generative AI game. **TechCrunch**, 21 mar. 2023. Disponível em: <https://techcrunch.com/2023/03/21/adobe-firefly-generative-ai/>. Acesso em: 25 abr. 2023.