

Utilização de Inteligência Artificial em Saúde: lições aprendidas durante o enfrentamento ao surto de COVID-19

Por Edson Amaro Jr¹, Michel Fornaciali², Andre Batista³, Murilo Gazzola⁴, Livia Paiva da Silva⁵, Diogo F. C. Patrão⁶ e Marcos Freitas Jr⁷

No contexto da pandemia de COVID-19, ficou mais evidente a necessidade de interação entre os profissionais de saúde e as tecnologias de análise de dados. As aplicações diretas envolvem cenários de atendimento (ajudar a atribuir pontuações de risco aos pacientes em relação ao prognóstico e melhorar as decisões sobre os cuidados médicos), de planejamento de curto prazo (organizar equipe e recursos para o atendimento), de avaliação de políticas públicas e ações macrorregionais de longo prazo (aplicações de modelos epidemiológicos e simulações).

Cada um desses cenários exige abordagens específicas. No atendimento, são primordiais soluções para os profissionais que estão na linha de frente, aplicadas quando um paciente está na

atenção primária, secundária ou terciária⁸. No planejamento de curto prazo, são necessárias ferramentas para o manejo de equipes ou de equipamentos que sejam acionáveis em até duas semanas, aplicadas na gestão de hospitais, municípios e estados. Por fim, para o planejamento de longo prazo é preciso pensar estratégias de micro e macrorregiões em cenários ajustáveis, aplicadas não só na gestão local, mas também em escala nacional. Em todas estas frentes, o objetivo é utilizar a tecnologia para prevenir mortes evitáveis ou causadas por complicações. Trata-se de um desafio, principalmente diante da falta de recursos, mas em um contexto como este as tecnologias de análise de dados podem contribuir para tornar mais eficientes o sistema de saúde e, em particular, a atuação das equipes.

¹ Médico neurorradiologista e Superintendente de Ciência de Dados e Analytics – Big Data Hospital Israelita Albert Einstein (HIAE) e professor livre-docente na Universidade de São Paulo (USP).

² Doutor e mestre em Engenharia da Computação pela Universidade Estadual de Campinas (Unicamp), bacharel em Ciência da Computação pela Unicamp e cientista de dados na equipe de Big Data Analytics do HIAE.

³ Doutor em Engenharia da Computação pela Escola Politécnica da USP e integrante da equipe de Big Data Analytics do HIAE.

⁴ Doutorando, mestre em Ciências de Computação e Matemática Computacional pela USP, membro do Núcleo Interinstitucional de Linguística Computacional da USP e cientista de dados na equipe de Big Data Analytics do HIAE.

⁵ Pós-graduanda em Gestão em Saúde no HIAE, pós-graduada em Gestão de Projetos e Engenharia Elétrica com ênfase em Telecomunicações pela Universidade Presbiteriana Mackenzie, tecnóloga em Telecomunicações pela Unicamp e técnica em Análise e Processamento de Sistemas e Dados pela Escola Técnica Federal de São Paulo.

⁶ Doutor em Oncologia pela Fundação Antônio Prudente (FAP/A.C. Camargo Cancer Center), bacharel em Física pela USP e cientista de dados na equipe de Big Data Analytics do HIAE.

⁷ Mestre em Sistemas de Informação pela USP, bacharel em Análise e Desenvolvimento de Sistemas pela Faculdade de Tecnologia Termomecânica (FTT) e cientista de dados na equipe de Big Data Analytics do HIAE.

⁸ O nível primário de atenção à saúde concentra as ações relacionadas à diminuição do risco de doenças e à proteção da saúde. Já o nível secundário é formado pelos hospitais e ambulatorios responsáveis por oferecer tratamento à população em pediatria, cardiologia e demais especialidades médicas, além de garantir a estruturação dos serviços de urgência e emergência. Por fim, no nível terciário estão reunidos os serviços de alta complexidade, representados pelos grandes hospitais e pelas clínicas especializadas.

Se modelos de IA dependem, em geral, de um grande volume de dados, como lidar com a escassez de dados na fase inicial da pandemia?

A experiência em Big Data Analytics do Hospital Israelita Albert Einstein

A organização de uma equipe dedicada ao estudo, à discussão e ao desenvolvimento de ferramentas de suporte baseadas em análise de dados e, em particular, algoritmos de Inteligência Artificial (IA), em si, não é tarefa fácil. O desafio se torna ainda mais considerável em um contexto de trabalho remoto, visando atender às medidas de distanciamento social impostas pelo enfrentamento ao novo coronavírus.

O grupo de Big Data Analytics do Hospital Israelita Albert Einstein (HIAE) vivenciou essa experiência inédita. Sediado na cidade de São Paulo, o Hospital atendeu o primeiro caso da doença no Brasil. Muitos foram os desafios de implementação de soluções de IA na rotina hospitalar, entre os quais se destaca o seguinte fato paradoxal: se modelos de IA dependem, em geral, de um grande volume de dados, como lidar com a escassez de dados na fase inicial da pandemia? Os aspectos práticos da formação de uma equipe para fazer frente às especificidades desse cenário, em que não há dados em quantidade ou qualidade desejada, podem trazer *insights* para momentos pós-pandêmicos.

Baseado no modelo ágil de gestão de projetos (Hidalgo, 2019), o time multifuncional de Big Data Analytics do HIAE foi subdividido em *squads* (equipes de profissionais dedicados a uma tarefa ou projeto) com perfis de expertise específicos. Designou-se uma pessoa como ponto focal para a interlocução com o grupo e como porta-voz em fóruns de definição de escopo, estratégia, governança e apresentação da evolução das soluções para as áreas que as utilizariam.

Iniciativa para envolver a comunidade aberta: AntennaCovid

A participação da comunidade que lida com processamento e análise de dados é crucial no contexto de uma pandemia. Por isso, é importante criar espaços que incluam diversas frentes, como pesquisadores, gestores e cientistas (Marston et al., 2020). É preciso que a gestão desses espaços participativos encontre pessoas com formações variadas, tendo a sensibilidade de reconhecer e fazer uso dos diferentes conhecimentos e experiências dos indivíduos, unidos em torno do objetivo de criar soluções de suporte aos profissionais de saúde.

Como é possível manter uma coprodução construtiva em um contexto de pandemia, com a pressão de gerar resultados práticos em um curto período de tempo? Foi nesse cenário que surgiu o AntennaCovid, um espaço para a interação de profissionais de saúde, cientistas de dados, engenheiros de *machine learning*, matemáticos, todos juntos no enfrentamento à COVID-19. Apoiado pela Rede Nacional de Ensino e Pesquisa e programado para existir enquanto houver necessidade, o AntennaCovid foi dividido em diversas plataformas para permitir aos participantes trabalhar e divulgar seus projetos – como o GitHub (github.com/antennaCovid), para o compartilhamento de código aberto; o Slack (antennaCovid.slack.com), para a comunicação facilitada e integrada à codificação; e o *site* (antennaCovid.org), para divulgação ao público.

Além disso, foram criadas reuniões virtuais semanais para que os integrantes dividam suas preocupações, ideias e trabalhos. Há também *roundups*, agendados de acordo com a necessidade para a discussão sobre o desenvolvimento dos projetos e a exposição de resultados práticos. Importante notar que tudo é divulgado de forma aberta no GitHub.

Assim, o AntennaCovid mantém uma gestão de espaço participativo para ganho de conhecimento em rede, aberto à comunidade, de forma que todos possam situar suas metas, controlar seus desenvolvimentos e interagir com times de diferentes instituições. Cabe notar que várias iniciativas semelhantes foram criadas no Brasil, algumas de interação direta com o AntennaCovid, reforçando o grande potencial de uso de ferramentas de colaboração em rede no país.

Soluções preditivas durante o atendimento: D+0⁹

As interações entre os profissionais de saúde e as tecnologias de análise de dados e de *machine learning* durante o atendimento são um solo fértil para o desenvolvimento de modelos preditivos. É possível criar soluções que ajudem a atribuir pontuações de risco aos pacientes sobre o prognóstico e melhorem as decisões sobre a necessidade e a intensidade dos cuidados médicos – contribuições fundamentais no enfrentamento à COVID-19, já que a identificação precoce de pacientes de alto risco pode, por exemplo, reduzir o uso de ventilação mecânica invasiva (Sun et al., 2020).

No caso de uma doença nova, tal como a COVID-19, os algoritmos são capazes de aprender os padrões referentes às interações das características do paciente que levam tanto a um risco maior de resultado positivo da enfermidade (Batista et al., 2020) quanto à evolução do quadro ou mesmo ao óbito, auxiliando a equipe em decisões clínicas e na alocação de recursos físicos para os casos mais graves (Alimadadi et al., 2020)¹⁰. Dessa forma, o desenvolvimento de modelos preditivos para prognósticos pode contribuir para a rotina médica de triagem de pacientes, antecipação de tratamentos, entre outros¹¹.

Em um trabalho de revisão sistemática de publicações na área da Saúde, Wynants et al. (2020) avaliaram os processos de construção, avaliação e divulgação científica de 66 modelos preditivos para COVID-19. Uma parcela considerável dos modelos foi elaborada para a predição da doença por meio de imagens. Além disso, os autores identificaram como preditores mais comumente utilizados para o diagnóstico de COVID-19 fatores como: idade, gênero, temperatura corpórea, sinais vitais, sintomas, pressão sanguínea e creatinina. Também se destacam entre os modelos de prognóstico os de desfechos para o óbito, deterioração clínica e tempo de internação.

⁹ Soluções para atuação no mesmo dia – “D+0”.

¹⁰ Mais especificamente, em Batista et al. (2020) apresentamos um modelo preditivo para COVID-19 a partir de informações como idade e sexo, bem como de exames básicos de hemograma. Foram construídos e avaliados cinco modelos preditivos supervisionados, dentre os quais o melhor (*support vector machine*) demonstrou que o uso de exames laboratoriais é promissor para o diagnóstico do novo coronavírus. Tal conclusão foi corroborada por trabalhos subsequentes em instituições de outros países (ver, em particular, Gao et al., 2020 e Brinati et al., 2020). A importância da construção de modelos preditivos para o prognóstico de COVID-19 é afirmada ainda pelo trabalho de Hirsch et al. (2020) sobre a incidência de insuficiência renal aguda em pacientes com a doença.

¹¹ Ainda nesse âmbito, temos concentrado esforços na predição de insuficiência renal aguda em pacientes hospitalizados. Os modelos são construídos a partir de arquiteturas computacionais de redes neurais recorrentes, com destaque para as *Long Short-Term Memory* (LSTM), discutidas em Swapnarekha et al. (2020).

As interações entre os profissionais de saúde e as tecnologias de análise de dados e de *machine learning* durante o atendimento são um solo fértil para o desenvolvimento de modelos preditivos.

Tanto a qualidade dos dados quanto suas condições de coleta e de utilização na rotina de atendimento são fundamentais para que as soluções analíticas tragam valor para a prática dos profissionais de saúde.

Sobressaem-se os trabalhos dedicados à construção de ferramentas *on-line* para predição de escore de risco para COVID-19 (DeCaprio et al., n.d.) e ao desenvolvimento de aplicativos *mobile* para predição de diagnóstico da doença a partir da inserção de resultados de exames laboratoriais e dos sintomas do paciente (Meng et al., n.d.). Tais estudos reforçam um conceito importante de comunhão de dados em ciência (*eScience* e *open data science*). Outros autores compartilharam seus códigos-fonte, possibilitando que a comunidade científica conheça as especificidades técnicas dos modelos construídos.

Apesar dos esforços para a divulgação de resultados e dos consequentes modelos preditivos, é importante destacar, conforme Wynants et al., que estes modelos precisam passar por um processo de validação externa antes de serem incorporados ao exercício médico (Cosgriff et al., 2019). Tanto a qualidade dos dados quanto suas condições de coleta e de utilização na rotina de atendimento são fundamentais para que as soluções analíticas tragam valor para a prática dos profissionais de saúde.

O avanço do uso de técnicas de *machine learning* e a abundância de publicações científicas que apresentam resultados otimistas sobre a maneira como os modelos preditivos podem contribuir com o exercício médico abrem espaço para reflexões e chamam atenção para a importância de boas práticas na divulgação de tais modelos. Por meio da adoção destas práticas tem-se a análise crítica pelos pares, o detalhamento das premissas, dos processos de seleção da amostra e das transformações aplicadas aos dados, além de considerações capazes de elucidar a compreensão da comunidade científica, otimizando o trabalho de validação externa¹².

Soluções preditivas para a gestão de serviços: D+15

As predições de curto prazo – em até quinze dias, ou D+15 – assumiram papéis diferentes entre os contextos e estágios de avanço da COVID-19. No início da pandemia, observou-se, não apenas no Brasil, crescimento exponencial do número de pessoas infectadas (Remuzzi & Remuzzi, 2020). Naquele momento, era fundamental entender quando haveria uma mudança na inclinação da curva, sobretudo para o planejamento dos recursos físicos e humanos necessários para a melhor assistência.

Modelos de predição com técnicas tradicionais que apresentavam crescimento linear e/ou exponencial foram utilizados nesse estágio (Martinez et al., 2020). Países onde a disseminação do vírus estava mais avançada foram *proxys* importantes para compreender o comportamento das curvas (Barm-Paris & Tsironis, 2020), permitindo, inclusive, a projeção inicial da evolução de casos no Brasil por meio de métodos de agrupamento. Ainda que projeções de médio e longo prazos tenham sido mais demandadas pelos gestores, principalmente em razão do maior impacto das decisões necessárias, modelos de curto prazo com abordagem matemática simplificada não deixaram de ser relevantes. Isso porque – ainda que pareça um contrassenso – modelos matemáticos mais complexos não são necessariamente mais confiáveis e nem sempre apresentam os melhores resultados (Roda et al., 2020).

¹² O *guideline Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD)*, por exemplo, é um dos primeiros instrumentos de verificação de requisitos mínimos para a publicação e a divulgação de modelos preditivos (Collins et al., 2015), sendo ainda bastante útil.

Desse modo, lidar com a pluralidade e as limitações dos modelos preditivos apresentados (Roda et al., 2020) exige a compreensão dos diferentes contextos, bem como dos desafios e dos resultados de diversos grupos de pesquisadores no mundo. São exemplos de cenários típicos a escassez de dados (comum no início de uma pandemia), a inconsistência na coleta e a falta de parâmetros bem estabelecidos. Também podem ser citadas a adoção de critérios variados de testagem de infectados, decisões governamentais divergentes e, especificamente no Brasil, as diferenças socioeconômicas entre regiões, que limitam a possibilidade de generalização dos resultados e refletem na qualidade do registro dos dados¹³.

Das intensas trocas ocorridas entre os grupos em decorrência da pandemia, ficam as lições de que desenvolver soluções preditivas, ainda que de curto prazo e com um impacto positivo para os gestores, implica mais do que abordagens matemáticas, exigindo, por exemplo, a compreensão de contextos e de limites de utilização. Entender os desafios atuais é fundamental para uma interpretação precisa do que dizem as previsões realizadas e para a redução dos riscos quando estas são usadas para apoiar decisões, especialmente políticas públicas na área da Saúde.

Soluções de longo prazo: modelos D+30 e ciclos epidemiológicos

Ainda que com alguma variabilidade, os modelos preditivos de longo prazo são importantes não como instrumentos para “acertar” os números em meses futuros, mas em especial para entender como as intervenções não farmacológicas, as decisões governamentais e as dinâmicas naturais de cada macrorregião refletem no comportamento das curvas de predição no decorrer do tempo. A compreensão de que, heurísticamente (por construção), há um conjunto de projeções que indicam tendências com inclinação mais acentuada pode prover *insights* relevantes no que se refere à efetividade das ações.

A epidemiologia é, por definição, o estudo em termos quantitativos da distribuição dos fenômenos de saúde e de seus fatores determinantes na população. Trata-se, por isso, da ciência que mais fez uso dos modelos preditivos de longo prazo. No caso da COVID-19, nossa linha de utilização se voltou às simulações da propagação do vírus em 60 dias ou mais, bem como à identificação e caracterização dos grupos de risco – ações cruciais para a melhor gestão da pandemia em prol da saúde pública.

Simulações de longo prazo

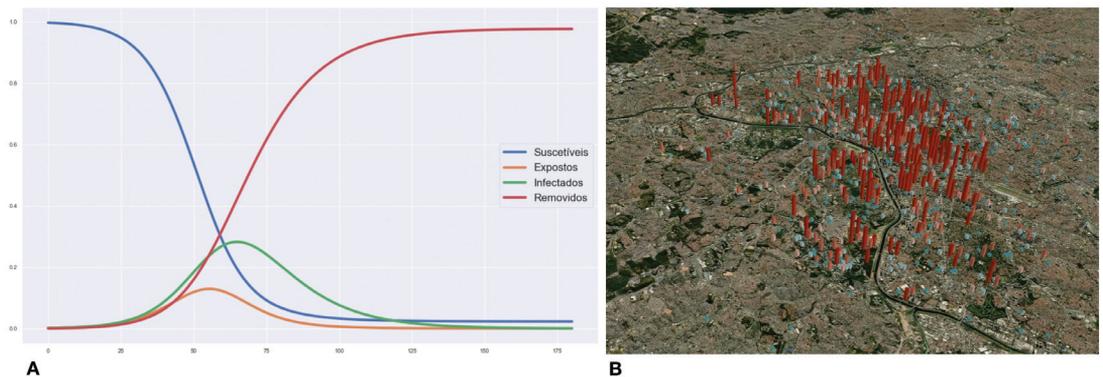
Toda epidemia é a manifestação de uma enfermidade caracterizada por sua rápida transmissão, atingindo simultaneamente diversas pessoas em um mesmo local. Compreender o ciclo epidemiológico específico, desde o início do surto até a sua extinção, é uma tarefa complexa, sobretudo para doenças novas. Assim, lança-se mão de modelos que descrevem, de forma simplificada, a transmissão entre indivíduos.

A epidemiologia é, por definição, o estudo em termos quantitativos da distribuição dos fenômenos de saúde e de seus fatores determinantes na população. Trata-se, por isso, da ciência que mais fez uso dos modelos preditivos de longo prazo.

¹³ Atrasos no registro de hospitalizações e de óbitos, por exemplo, são um problema enfrentado em diversos países, impactando na utilização de técnicas para a atualização de dados (*nowcasting*) com o intuito de aumentar a precisão das soluções preditivas desenvolvidas (Alfmejd et al., 2020; Schneble et al., 2020; Puca & Buonanno, n.d.).

Entre a ampla gama de modelos epidemiológicos existentes (Britton, 2009; Chowell et al., 2016), nós nos concentramos nos compartimentais determinísticos. Adotamos o modelo SEIR, que caracteriza uma população em epidemia em quatro estados: suscetível, exposta, infectada e recuperada (o que abrange óbitos). A Figura 1 mostra a evolução teórica da doença, descrevendo a porcentagem da população em cada estado ao longo do tempo.

Figura 1– Representação teórica de uma pandemia sob o modelo SEIR (A) e representação de testes para COVID-19 em São Paulo em 12/03/2020 (B)



Fonte: Elaboração própria.

Como toda simplificação matemática, o modelo SEIR possui limitações. Entre elas, destacam-se a superestimação de casos e a dependência de fortes premissas relacionadas à probabilidade uniforme de as pessoas se infectarem. Para a COVID-19, é preciso considerar ainda que se trata de uma doença desconhecida, com poucos estudos publicados e um cenário de compreensão altamente dinâmico. Tais complicações nos impedem de aplicar o modelo para afirmar com precisão quando ocorrerão fenômenos de interesse (o momento do pico da curva de infectados e o fim da epidemia, por exemplo) e qual será a sua magnitude (quantos serão os casos no pico da curva e quantos óbitos teremos em 90 dias).

Para amenizar as limitações técnicas e de contexto, introduzimos duas inovações em nossas modelagens: por um lado, trabalhamos com múltiplos cenários para evidenciar incertezas e, por outro, consideramos ações reais de mitigação da COVID-19 para aumentar a assertividade das curvas. Primeiramente, concentramos atenção na curva de infectados, uma vez que os números de novos casos, de hospitalizações e de óbitos derivam dela. Ao modelá-la, incorporamos a noção de incertezas por meio da ilustração visual de diversos cenários ao longo do tempo, representando flutuações entre contextos mais otimistas e mais pessimistas. Tal abordagem transmite ao interlocutor o fato de os modelos serem naturalmente imprecisos, impossibilitando a interpretação literal dos números apresentados nas simulações. Ao comparar diferenças entre cenários de uma mesma modelagem, podemos inferir os impactos no pior e melhor caso, o que vem sendo abordado com sucesso em outros estudos sobre o novo coronavírus (Lopez & Rodo, n.d.; Prem et al., 2020).

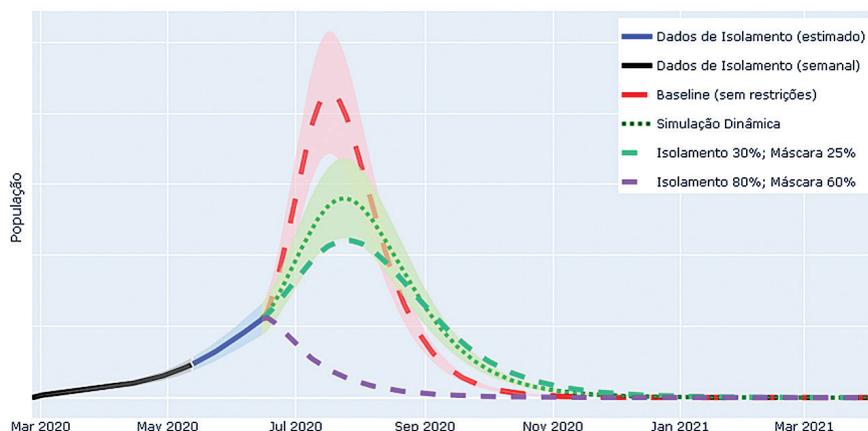
O modelo SEIR assume que toda a população pode se contaminar e que a transmissão da doença ocorre livremente, sem ações de mitigação, o que leva à superestimação de casos. Para amenizar este fenômeno, incorporamos às modelagens as iniciativas de controle da disseminação do vírus. Como ainda não há vacinas nem medicamentos específicos para a COVID-19, temos majoritariamente intervenções não farmacológicas (Lai et al., 2020), como a higienização pessoal e o distanciamento social.

Um dos principais parâmetros da modelagem é o número de reprodução (ou R_0), que determina o potencial de transmissão de uma doença, ou seja, quantos novos casos uma pessoa infectada produz. Em geral, o R_0 é tratado como uma taxa única ao longo do tempo. No entanto, durante uma pandemia aplica-se uma série de intervenções para mitigar a transmissão do vírus. Nossas modelagens visam estimar matematicamente quanto cada intervenção e sua adesão pela população afetam o R_0 , criando um R_0 comparável por meio de cenários.

Uma vez que as ações de maior impacto na mitigação da COVID-19 são o isolamento social (Hellewell et al., 2020) e o uso de máscaras (MacIntyre & Chughtai, 2020), implementamos essas medidas para corrigir o R_0 ¹⁴. Dessa forma, temos uma simulação mais realista, destacadas as incertezas da abordagem. O usuário pode navegar pelo simulador, criando cenários de aderência às intervenções não farmacológicas e comparando seus efeitos no longo prazo. Como exemplo, a Figura 2 confronta quatro cenários com diferentes taxas de adesão da população ao isolamento social e ao uso de máscaras, bem como uma simulação atual e um cenário sem nenhuma intervenção (*baseline*) – para estas duas últimas categorias, as faixas de sombreado ilustram as incertezas.

Nossas modelagens visam estimar matematicamente quanto cada intervenção e sua adesão pela população afetam o R_0 [número de reprodução], criando um R_0 comparável por meio de cenários.

Figura 2 – Ferramenta resultante de simulação de COVID-19 no longo prazo



Fonte: Elaboração própria.

¹⁴ As taxas de isolamento social são provenientes de estimativas de mobilidade a partir da triangulação de antenas de telefonia móvel. Já o uso de máscaras não é medido oficialmente, adotando-se uma estimativa empírica da porcentagem da população que teria aderido à medida. Também foi definida uma porcentagem de eficiência das máscaras, considerando a qualidade do produto e seu uso incorreto pelos indivíduos.

Para a análise de dados complexos, como os da área da Saúde, é necessário unir informações que permitam compreender fenômenos de base quantitativa para a realização das análises de predição, prescrição ou classificação.

Em relação ao entendimento dos grupos de risco – tarefa relevante em termos de saúde pública –, o primeiro passo é reconhecer quem está mais suscetível a sofrer os efeitos graves da COVID-19. Em termos gerais, são os indivíduos com idade avançada e/ou comorbidades (hipertensão, obesidade, diabetes etc.) (Jordan et al., 2020). À medida que a pandemia se alastra, observa-se que a população de risco passa a ser aquela mais vulnerável a ser infectada e/ou com acesso restrito ao sistema de saúde – ou seja, as pessoas mais pobres e as que não podem se isolar por razões diversas, sobretudo falta de emprego formal e impossibilidade de aderir ao trabalho remoto.

Identificamos e geocalizamos essa população, atividade fundamental na priorização de regiões para a intervenção dos gestores de saúde. Além disso, é possível direcionar locais para a construção de polos de testagem em massa ou mesmo definir parâmetros para a reabertura dos municípios no pós-quarentena. Espera-se que essas ferramentas interativas permitam aos gestores estimar ou comparar medidas – ainda que não haja certeza da magnitude absoluta de seus efeitos, é bastante razoável pensar na sua comparação relativa.

Acreditamos, portanto, que as ações guiadas pela epidemiologia resultaram em uma solução dinâmica voltada aos gestores de saúde para o entendimento e a mensuração de como medidas adotadas no presente impactam o futuro da pandemia de COVID-19. Da mesma forma, a identificação e a localização da população vulnerável agilizam a tomada de decisões e promovem a melhor gestão da saúde pública.

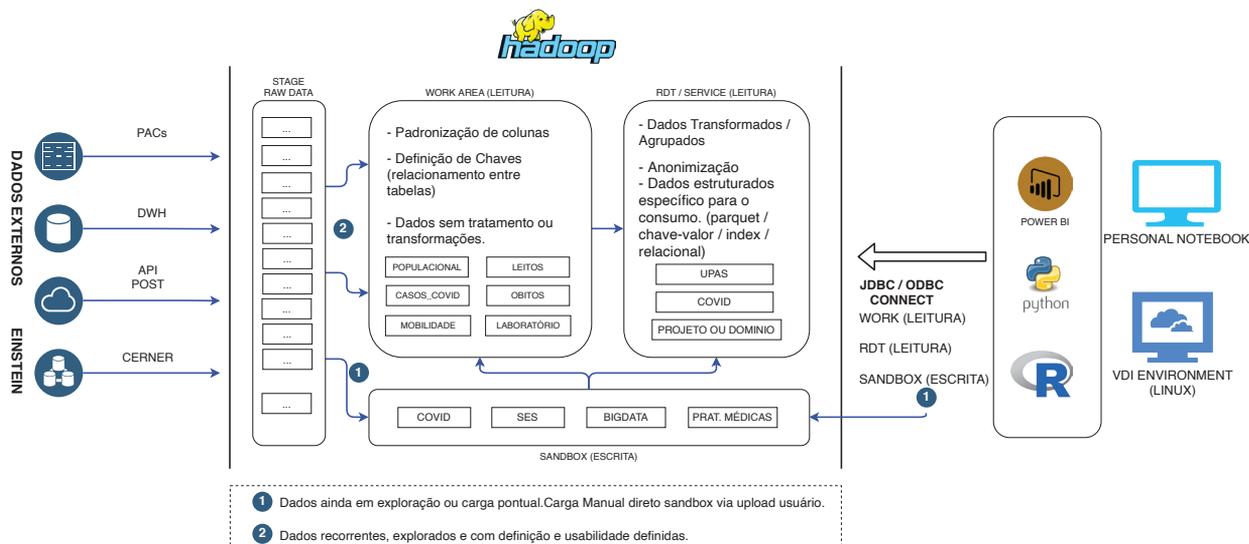
Soluções de data *lake*: organizando dados de múltiplas fontes

Para a análise de dados complexos, como os da área da Saúde, é necessário unir informações que permitam compreender fenômenos de base quantitativa para a realização das análises de predição, prescrição ou classificação. Diante da falta de dados disponíveis no início da pandemia de COVID-19, essa necessidade tornou-se ainda mais premente. Cada uma das três principais bases de dados contendo os números diários de casos da doença por país possuía uma metodologia diferente, gerando divergências em formatação, granularidade das localizações e endereço (URL) para recuperação.

O trabalho de identificação, extração, preparação e convergência dos dados pode ser árduo e moroso. Em uma equipe de cientistas, é possível que muitos precisem dos mesmos dados, gerando redundância ou mesmo incoerência nos resultados. Para reduzir esse problema, é comum o estabelecimento de um *data lake* (“lago de dados”, às vezes chamado de *data ocean* ou variantes). De forma geral, *data lake* significa um repositório de dados em formato bruto, contendo tanto dados estruturados quanto não estruturados. Trata-se de um ambiente capaz de realizar tarefas de extração, transformação e carga (ETL, *Extract, Transform, Load*), armazenamento de dados em grande escala e disponibilização desses dados de forma uniforme.

Para apoiar a criação de ferramentas de análise da pandemia de COVID-19 no Brasil, nossa equipe estabeleceu uma base de dados específica (*Covid Lake*), um *data lake* com dados de escopos diversificados, como mostra a Figura 3.

Figura 3 – Diagrama de arquitetura do *data lake*¹⁵



Fonte: Elaboração própria.

Essa estrutura serviu de suporte para as atividades dos diferentes *squads*. Foram consideradas 25 bases de dados públicas ou de projetos específicos, muitos deles conduzidos em parceria, cuja natureza inclui: comportamento (questionários aplicados à população brasileira, voltados a entender aglomerações e exposição à contaminação); deslocamento (dados referentes a mobilidade por meio de telefonia móvel); clima (estações meteorológicas e índices de poluição do ar); composição demográfica; e informações de outros países.

Conclusão

Apesar da complexa experiência trazida pelo momento de combate ao surto de COVID-19, esta poderá se concretizar como um acelerador em termos de contribuições para a adoção de soluções de ciência de dados – incluída a aplicação de Inteligência Artificial. Espera-se que os resultados sejam cada vez mais promissores para que no futuro, se condição semelhante surgir, tenhamos ferramentas analíticas mais maduras, capazes de impedir os danos ainda não evitáveis.

¹⁵ Estrutura baseada em *Hadoop*, HDFS e *HBase*, com autenticação *Kerberos* apontando para o *Active Directory* (AD) corporativo. O acesso é circunscrito à rede interna da instituição.

Referências

- Akselrod-Ballin, A. et al. (2019). Predicting breast cancer by applying deep learning to linked health records and mammograms. *Radiology*, 292(2), 331–342.
- Alimadadi, A. et al. (2020). Artificial intelligence and machine learning to fight COVID-19. *Physiological Genomics*, 52(4), 200–202.
- Altmejd, A., Rocklöv, J., & Wallin, J. (2020). Nowcasting COVID-19 statistics reported with delay: A case-study of Sweden. *arXiv.org*, q-bio.PE.
- Barmparis, G. D., & Tsironis, G. P. (2020). Estimating the infection horizon of COVID-19 in eight countries with a data-driven approach. *Chaos, Solitons & Fractals*, 135, 109842.
- Barroso, L. V. et al. (2020). GeoSES: A socioeconomic index for health and social research in Brazil. B. Lanza Queiroz, ed. *PloS one*, 15(4), e0232074.
- Batista, A. F. de M. et al. (2020). COVID-19 diagnosis prediction in emergency care patients: A machine learning approach. *medRxiv*, 1–8.
- Brinati, D. et al. (2020). Detection of COVID-19 infection from routine blood exams with machine learning: A feasibility study. *Journal of Medical Systems*, 44(8).
- Britton, T. (2009). Stochastic epidemic models: A survey. *arXiv.org*, math.PR.
- Chowell, G. et al. (2016). Mathematical models to characterize early epidemic growth: A review. *Physics of Life Reviews*, 18, 66–97.
- Collins, G. S. et al. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. 1–10.
- Cosgriff, C. V. et al. (2019). Developing well-calibrated illness severity scores for decision support in the critically ill. *npj Digital Medicine*, 1–8.
- DeCaprio, D. et al. (n.d.). Building a COVID-19 Vulnerability Index. *arXiv.org*.
- Gao, L. et al. (2020). Prognostic value of NT-proBNP in patients with severe COVID-19. *Respiratory Research*, 21(1), 83–7.
- Hellewell, J. et al. (2020). Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet. Global health*, 8(4), e488–e496.
- Hidalgo, E. S. (2019). Adapting the scrum framework for agile project management in science: Case study of a distributed research initiative. *Heliyon*, 5(3), e01447.
- Hirsch, J. S. et al. (2020). Acute kidney injury in patients hospitalized with COVID-19. *Kidney International*, 98(1), 209–218.
- Hornik, R. (2004). Some reflections on diffusion theory and the role of Everett Rogers. *Journal of Health Communication*, 9(sup1), 143–148.
- Jordan, R. E., Adab, P., & Cheng, K. K. (2020). COVID-19: Risk factors for severe disease and death. *British Medical Journal. Clinical Research Edition*, 368, m1198.
- Lai, S. et al. (2020). Effect of non-pharmaceutical interventions for containing the COVID-19 outbreak in China. *medRxiv*.
- Lopez, L. R., & Rodo, X. (n.d.). A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy: Simulating control scenarios and multi-scale epidemics. *medRxiv*.
- MacIntyre, C. R., & Chughtai, A. A. (2020). A rapid systematic review of the efficacy of face masks and respirators against coronaviruses and other respiratory transmissible viruses for the community, healthcare workers and sick patients. *International Journal of Nursing Studies*, 108, 103629.
- Marston, C., Renedo, A., & Miles, S. (2020). Community participation is crucial in a pandemic. *Lancet*, 395(10238), 1676–1678.
- Martinez, E. Z., Aragon, D. C., & Nunes, A. A. (2020). Short-term forecasting of daily COVID-19 cases in Brazil by using the Holt's model. *Revista da Sociedade Brasileira de Medicina Tropical*, 53(10223), 102433–35.
- Meng, Z. et al. (n.d.). Development and utilization of an intelligent application for aiding COVID-19 diagnosis. *medRxiv*.
- Nadim, S. S., Ghosh, I., & Chattopadhyay, J. (2020). Short-term predictions and prevention strategies for COVID-19: A model based study. *arXiv.org*, q-bio.PE.
- Prem, K. et al. (2020). The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study. *The Lancet. Public health*, 5(5), e261–e270.
- Puca, M., & Buonanno, P. (n.d.). Using newspapers obituaries to nowcast daily mortality: Evidence from the Italian COVID-19 hot-spots. *medRxiv*.
- Remuzzi, A., & Remuzzi, G. (2020). COVID-19 and Italy: What next? *Lancet*, 395(10231), 1225–1228.
- Roda, W. C. et al. (2020). Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling*, 5, 271–281.
- Santos, H. G. D. et al. (2020). Machine learning to predict 30-day quality-adjusted survival in critically ill patients with cancer. *Journal of Critical Care*, 55, 73–78.
- Schneble, M. et al. (2020). Nowcasting fatal COVID-19 infections on a regional level in Germany. *arXiv.org*, stat.AP.
- Sun, Q. et al. (2020). Lower mortality of COVID-19 by early recognition and intervention: Experience from Jiangsu Province. *Annals of Intensive Care*, 10(1), 33–34.
- Swapnarekha, H. et al. (2020). Role of intelligent computing in COVID-19 prognosis: A state-of-the-art review. *Chaos, Solitons & Fractals*, 138, 109947–48.
- Wynants, L. et al. (2020). Prediction models for diagnosis and prognosis of COVID-19 infection: Systematic review and critical appraisal. *British Medical Journal. Clinical Research Edition*, 369, m1328.

Usos de Inteligência Artificial no contexto da COVID-19

As tecnologias de Inteligência Artificial (IA) estão sendo empregadas de maneiras diversas no combate à COVID-19. A Figura 1 apresenta exemplos de aplicações de IA em diferentes estágios da crise e para finalidades distintas, como a produção de informações customizadas, o monitoramento de contágios em tempo real ou mesmo o transporte e a entrega de materiais.

Figura 1 – Exemplos de aplicações de IA no enfrentamento à COVID-19

Aceleração de pesquisas Projetos de dados abertos e computação distribuída para encontrar soluções orientadas por IA para a pandemia. Ex.: desenvolvimento de medicamentos e vacinas	Deteção Alerta precoce Detecção de anomalias e “sinais de fumaça” digitais. Ex.: <i>BlueDot</i>	Diagnóstico Reconhecimento de padrões por meio de imagens médicas e dados de sintomas. Ex.: tomografias computadorizadas		
	Prevenção Predição Cálculo da probabilidade de contágio de uma pessoa. Ex.: <i>EpiRisk</i>	Vigilância Monitoramento e rastreo do contágio em tempo real. Ex.: rastreamento de contatos	Informação Notícias personalizadas e moderação de conteúdo para combater a desinformação. Ex.: via redes sociais	
	Resposta Entrega Drones para o transporte de materiais, robôs para tarefas com alto risco de exposição em hospitais. Ex.: <i>robô CRUZR</i>	Automação de serviços Implantação de <i>chatbots</i> e assistentes virtuais de triagem. Ex.: <i>chatbot</i> para a COVID-19 no Canadá		
	Recuperação Monitoramento Acompanhamento da recuperação econômica por satélite, GPS e dados de mídias sociais. Ex.: <i>WeBank</i>			

Fonte: OECD, *Using artificial intelligence to help combat COVID-19* (2020).

Saúde e vigilância pública: IA para o rastreamento de contatos

Uma das estratégias de vigilância pública que têm sido implementadas como mecanismo de prevenção no setor de Saúde é o rastreamento de contatos, que diz respeito ao processo usado para identificar, informar e monitorar indivíduos que tiveram contato próximo com alguém infectado por um vírus. Para tal, conta-se cada vez mais com o uso de tecnologias e, em particular, da IA.

No contexto da COVID-19, a IA tem sido empregada em estratégias que visam diminuir a transmissão do vírus, tais como: o escaneamento (*scanning*) de espaços públicos, que utiliza câmeras infravermelhas para identificar pessoas com febre; *drones* e robôs baseados em IA para detectar movimentações, reuniões sociais e indivíduos sem máscaras; e o uso de algoritmos de rastreamento de contatos orientados por IA para enviar aos cidadãos mensagens de texto personalizadas.

Na China, um exemplo de rastreamento de contatos auxiliado por IA se baseia em um *plug-in* instalado em aplicativos amplamente utilizados pela população. O *plug-in* coleta dados dos usuários e alimenta um banco de dados central, cujas informações são analisadas por meio de ferramentas de IA. Em seguida, o algoritmo emite códigos de cores aos usuários para indicar a imposição ou não de restrições ao seu deslocamento, além de alertar os contatos próximos recentes em caso de diagnóstico positivo para COVID-19.

Artigo II

Rastreamento digital de contatos na luta contra a COVID-19

Por Sacha Alanoca¹⁶, Nicolas Guetta Jeanrenaud¹⁷, Nyasha Weinberg¹⁸, R. Buse Çetin¹⁹ e Nicolas Mialhe²⁰

Uma crise de saúde pública sem precedentes

Desde o surto de COVID-19 em Wuhan, na China, no fim de 2019, e sua classificação como pandemia pela Organização Mundial da Saúde (OMS), em março de 2020, os números de infecções e de mortes continuam a crescer em todo o mundo. Até 10 de julho deste ano, a contagem era de mais de 12,3 milhões de casos diagnosticados e mais de 555 mil óbitos. Na América Latina, as estatísticas chegaram a mais de 3,3 milhões de pessoas contagiadas e 142 mil mortes, com maior concentração de casos no Brasil, Peru e Chile, respectivamente²¹. Para enfrentar essa crise inédita, a resposta imediata da maioria dos governos foi aplicar regras rígidas de distanciamento social, junto com *lockdowns* parciais ou totais.

Medidas como essas visam desacelerar a taxa de infecções, ou “achatar a curva” – em referência à representação gráfica da redução no número de casos diários –, permitindo que os sistemas de saúde consigam tratar o fluxo de pacientes recebidos. Contudo, sem a testagem em grande escala, por um lado, e sem uma perspectiva de tratamento e de vacina a ser implementados no futuro próximo, por outro, esses esforços apenas adiam, mas não impedem, a disseminação da pandemia.

Para enfrentar essa crise inédita, a resposta imediata da maioria dos governos foi aplicar regras rígidas de distanciamento social, junto com *lockdowns* parciais ou totais.

¹⁶ Pesquisadora sênior de Políticas de Inteligência Artificial (IA) e chefe de Desenvolvimento Comunitário na The Future Society. É mestre pela Sciences Po Paris, com especialização em economias emergentes e inovação digital.

¹⁷ Graduando em Tecnologia e Política no Instituto de Tecnologia de Massachusetts (MIT), onde se especializa em ciências sociais computacionais e sistemas de informação. Também é mestre em Engenharia Civil pela l'École des Mines de Paris.

¹⁸ Pesquisadora afiliada na The Future Society, com bolsa de estudos do programa *re:constitution*. É membro da Ordem dos Advogados do Reino Unido, especialista jurídica e em políticas públicas, com mestrado na Harvard Kennedy School.

¹⁹ Pesquisadora de Políticas de IA na The Future Society. É mestre em Gestão Pública Internacional pela Sciences Po Paris, com especialização em política econômica e governança global, e graduada em Ciências Sociais e Estudos do Médio Oriente pela Sciences Po Paris.

²⁰ Cofundador da The Future Society, membro do Comitê Diretor da parceria AI Commons, do grupo de especialistas em IA da Organização para a Cooperação e o Desenvolvimento Econômico (OCDE) e do Conselho Global de Inteligência Estendida (CXI), entre outras organizações. Leciona na Sciences Po Paris, na IE School of Global and Public Affairs, em Madri, e na Escola de Governo Mohammed bin Rashid, em Dubai. É pesquisador sênior associado na Universidade Harvard e *felllow* na IE Business School, em Madri.

²¹ A OMS atualiza regularmente esses dados. Disponível em: <https://COVID19.who.int>

Ademais, tais ações têm custos socioeconômicos muito significativos. Desde o início do surto, mais de 13 milhões de pessoas se declararam desempregadas na França²² e mais de 33 milhões nos Estados Unidos²³. Milhões de pessoas estão desempregadas ao redor da América Latina. No Brasil, um número recorde de 7,8 milhões de pessoas perdeu seu ofício entre março e maio de 2020, resultando em uma taxa abaixo de 50% de pessoas com trabalho²⁴. É preciso levar em conta também os impactos consideráveis do atual cenário na saúde mental da população, com níveis crescentes de estresse e ansiedade, bem como as consequências ainda desconhecidas da suspensão ou do adiamento de tratamentos médicos de rotina (Ornell et al., 2020).

Rastreamento de contatos como técnica de mitigação rápida e precisa

Confrontados com a necessidade urgente de mitigar essa crise, governos de todo o mundo têm explorado uma combinação de ferramentas médicas e tecnológicas. Entre elas, o rastreamento de contatos descreve uma variedade de técnicas utilizadas para identificar pessoas que tenham se aproximado de um indivíduo com diagnóstico positivo para COVID-19, tomando as ações apropriadas para informá-las, isolá-las e tratá-las.

No passado, o rastreamento manual de contatos foi implementado sistematicamente para combater epidemias como tuberculose, sarampo, HIV e, em anos recentes, para limitar a propagação do ebola e do SARS (Hart et al., 2020). Comparada a outras doenças, porém, a COVID-19 é considerada como altamente contagiosa²⁵. O intenso trabalho de rastreamento feito de modo manual, com base na memória humana e em uma equipe médica treinada, carece tanto de rapidez quanto de precisão para corresponder à propagação do vírus.

Por meio da mobilização de tecnologias de planejamento e de análise, o rastreamento digital de contatos contribui para melhorar esse quadro. Os *smartphones* e as pulseiras *smart* automatizam o processo, permitindo o rastreamento de proximidade, o monitoramento digital da saúde ou mesmo a identificação de pacientes. Quando aliado às demais medidas adequadas, como o distanciamento social e a testagem, o rastreamento digital pode ajudar a quebrar a cadeia de contaminação e prevenir novos casos ao enviar alertas a populações expostas ao risco. Diversos países que conseguiram “achatar a curva” na primeira onda de COVID-19, como Taiwan, Coreia do Sul, Singapura e China, adotaram a estratégia de maneira bem-sucedida. Segundo o *MIT Technology Review*²⁶, atualmente há 47 aplicações desse tipo implementadas em mais de 30 nações.

(...) o rastreamento de contatos descreve uma variedade de técnicas utilizadas para identificar pessoas que tenham se aproximado de um indivíduo com diagnóstico positivo para COVID-19, tomando as ações apropriadas para informá-las, isolá-las e tratá-las.

²² Saiba mais: <https://www.rtl.fr/actu/conso/coronavirus-pourquoi-y-a-il-autant-de-salaries-en-chomage-partiel-en-france-7800466953>

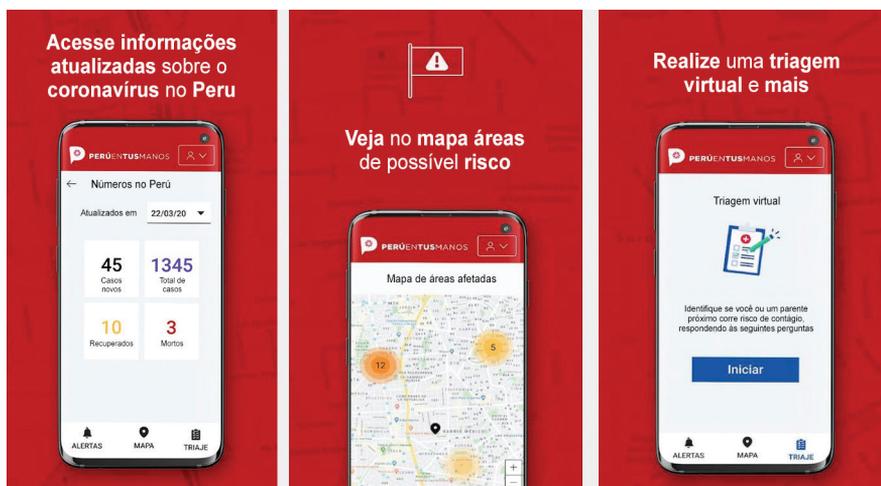
²³ Saiba mais: <https://www.bbc.com/news/business-52570600>

²⁴ Saiba mais: <https://edition.cnn.com/2020/07/01/americas/latin-america-coronavirus-unemployment-intl/index.html>

²⁵ Com incubação mediana em um período de cinco dias, 97,5% das pessoas que desenvolvem sintomas o fazem em 11,5 dias (Lauer et al., 2020; Li et al., 2020). Já as estimativas da parcela assintomática de indivíduos infectados variam de 15,5% (Mizumoto et al., 2020) a 56% (Arons et al., 2020).

²⁶ Saiba mais: <https://www.technologyreview.com/2020/05/07/1000961/launching-mittr-Covid-tracing-tracker>

Figura 1 – A aplicação PerúEnTusManos, implementada pelo governo peruano



Fonte: Presidencia Del Consejo de Ministros, Peru.

Foram colocados em xeque seu desempenho e impacto na privacidade de dados, nos direitos humanos, na estigmatização de indivíduos, na desconfiança em relação a autoridades públicas e, por último, no medo do estabelecimento de uma vigilância em massa.

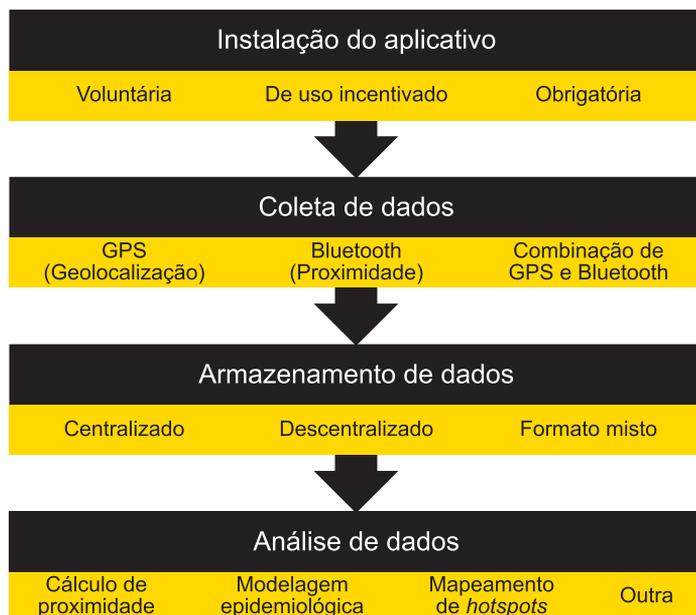
Logo surgiram, no entanto, preocupações sobre a adoção generalizada do rastreamento digital de contatos. Foram colocados em xeque seu desempenho e impacto na privacidade de dados, nos direitos humanos, na estigmatização de indivíduos, na desconfiança em relação a autoridades públicas e, por último, no medo do estabelecimento de uma vigilância em massa. Em diversos países, informações equivocadas e a falta de clareza quanto a salvaguardas éticas, legais e técnicas polarizaram o debate, em um momento em que, mais do que nunca, é urgente construir confiança e impedir que as liberdades civis sejam sacrificadas em nome da saúde pública.

Há uma necessidade premente de codensar mecanismos de governança que maximizem os benefícios à saúde trazidos pelos aplicativos de rastreamento de contatos, ao mesmo tempo que mitiguem seus possíveis efeitos adversos. A gravidade da situação sanitária não deve impactar nossa habilidade coletiva de tomar a decisão correta e de navegar com responsabilidade pelas tensões entre saúde, segurança e liberdades civis. As evidências sugerem que a adoção ampla desses aplicativos depende da confiança pública, o que reforça a demanda por um *framework* ético que governe seu desenvolvimento.

Alternativas técnicas e *tradeoffs*

Os aplicativos de rastreamento de contatos se baseiam em um leque de tecnologias e em protocolos para preservação da privacidade. Podem ser voluntários (disponíveis em lojas de aplicativos), obrigatórios ou de uso incentivado. O processo de coleta de dados se dá via proximidade (em geral, por meio de Bluetooth), dados de localização GPS ou um modelo misto. Os dados podem ser armazenados nos dispositivos móveis dos indivíduos, em uma abordagem descentralizada; em um servidor centralizado; ou em um formato misto. Por fim, a análise conduzida pela autoridade relevante varia, indo do simples cálculo de proximidade até a modelagem epidemiológica, o mapeamento de *hotspots* e muito mais.

Figura 2 – Características técnicas do rastreamento digital de contatos



Fonte: The Future Society.

Instalação dos aplicativos

Um imperativo técnico para o sucesso de um aplicativo de rastreamento de contatos é a sua instalação generalizada pelos indivíduos. Apesar de não haver consenso entre os epidemiologistas sobre a taxa de transmissão da COVID-19, um estudo conduzido na Universidade de Oxford estimou que, com uma taxa mínima de 60% de adoção, tais aplicativos poderiam de fato deter a epidemia²⁷. Este, porém, é um limiar mínimo – autores de relatórios subsequentes apontaram que níveis inferiores de adesão já seriam vitais para conter a propagação do vírus²⁸.

Ainda que a instalação desses aplicativos seja de grande importância para a luta dos governos contra a COVID-19, autoridades de diversos países têm adotado uma abordagem radical, tomando o uso obrigatório para todos os cidadãos. Na China, o “detector de contato” se conecta ao WeChat e ao Alipay, aplicativos amplamente utilizados. A ferramenta analisa dados de saúde e de viagens para atribuir escores de risco aos indivíduos, podendo negar-lhes acesso a lojas e serviços essenciais²⁹. No Kuwait e no Bahrein, o aplicativo é separado das funcionalidades telefônicas existentes, mas sua instalação é exigida por lei.

A maioria dos países tem instituído aplicativos voluntários, embora incentivos personalizados ou genéricos possam ser incorporados às suas funcionalidades. Alguns oferecem informações adicionais sobre onde realizar testes

Um imperativo técnico para o sucesso de um aplicativo de rastreamento de contatos é a sua instalação generalizada pelos indivíduos.

²⁷ Disponível em: <https://www.research.ox.ac.uk/Article/2020-04-16-digital-contact-tracing-can-slow-or-even-stop-coronavirus-transmission-and-ease-us-out-of-lockdown>

²⁸ Saiba mais: <https://www.technologyreview.com/2020/06/05/1002775/Covid-apps-effective-at-less-than-60-percent-download>

²⁹ Saiba mais: <https://www.nytimes.com/2020/05/26/technology/china-coronavirus-surveillance.html>

para a COVID-19 (Austrália) ou sobre a disponibilidade de leitos hospitalares (Turquia), enquanto outros divulgam atualizações a respeito da pandemia (Bulgária) ou cuidados com a saúde (Catar). Em que medida tais incentivos levarão à adoção compulsória do aplicativo dependerá da existência de rotas alternativas para acessar serviços idênticos.

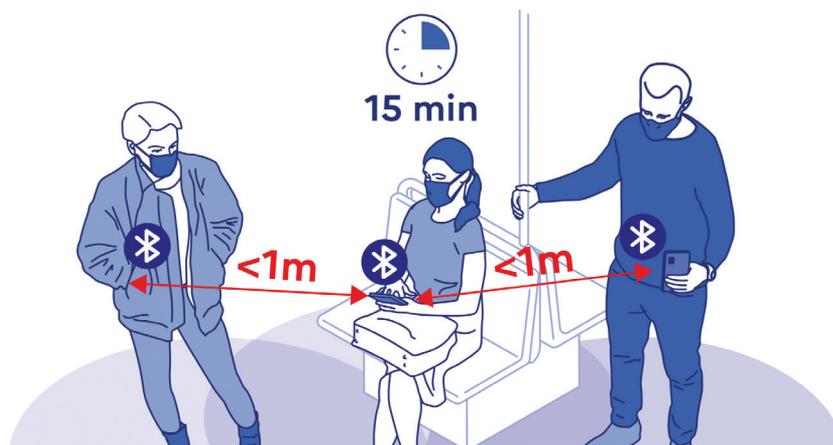
Coleta de dados

Segundo informações do *COVID-19 Digital Rights Tracker*, da empresa inglesa Top10VPN, no que se refere ao processo de coleta de dados, 34% dos aplicativos de rastreamento de contatos existentes utilizam GPS, 35% fazem uso de Bluetooth e 24% se valem de uma combinação de ambos. Essa distribuição destaca a falta de consenso em torno da tecnologia mais eficaz para a tarefa.

Os protocolos de Bluetooth dependem exclusivamente da detecção de proximidade. Um dispositivo móvel registra outros sinais de Bluetooth dentro de um determinado intervalo de tempo, no chamado “aperto de mão digital”. Durante a troca, cada dispositivo coleta *tokens*³⁰ com pseudônimos – tipicamente, cadeias de números que mudam várias vezes ao dia de forma aleatória – como identificadores dos demais usuários. Um mecanismo de notificação informa as pessoas quando elas estiverem próximas de um usuário sinalizado como infectado.

Países como México, Itália, Japão e Reino Unido adotaram protocolos de Bluetooth de baixa energia (BLE, *Bluetooth Low Energy*) ponto a ponto (P2P), enquanto Singapura usa o protocolo BlueTrace baseado em Bluetooth. Na maioria dos casos, a proximidade é medida dentro de um metro e os *tokens* pseudonimizados são trocados após 15 minutos, como ilustra a Figura 3 a partir do exemplo francês.

Figura 3 – Aplicativo de rastreamento de proximidade StopCovid, implementado pelo governo francês



Fonte: Governo da França (2020).

³⁰ Dispositivos físicos que geram uma senha temporária de proteção para as contas utilizadas pelo usuário, auxiliando em sua segurança pessoal.

Já o uso de GPS depende de dados móveis para reconstruir o histórico de localização dos usuários. Quando o indivíduo é marcado como infectado, a aplicação rastreia todas as pessoas com as quais ele teve contato. Os protocolos de GPS têm a característica adicional de destacar pontos de acesso geográficos do vírus, permitindo medidas direcionadas, o que não é possível com o Bluetooth. No entanto, essa técnica é menos precisa no registro de proximidade, especialmente em ambientes internos ou durante viagens subterrâneas. Em termos de privacidade, esses protocolos têm sido associados a algumas das aplicações mais invasivas, como BeAware, do Bahrein, Shlonik, do Kuwait, e Smittestopp, da Noruega³¹.

Há, por fim, protocolos que combinam dados de Bluetooth e de GPS para o rastreamento de contatos. Exemplo disso é a plataforma desenvolvida pelo Instituto de Tecnologia de Massachusetts (MIT), Safe Paths, que faz uso de ambos para registrar a localização das pessoas. Gratuita e de código aberto, a ferramenta foi adotada em vários países, como o Chipre³².

Armazenamento de dados

Além das aplicações de rastreamento de contatos, diversas tecnologias solicitam informações pessoais adicionais do usuário. Ketju, na Finlândia, e Hayat Eve Siğar, na Turquia, exigem números de telefone válidos para autenticação, enquanto na Islândia (somente para indivíduos diagnosticados com COVID-19), no Catar e no Kuwait a autenticação é feita via números de identificação nacional. Alguns aplicativos, como ViruSafe, na Bulgária, pedem a idade e o histórico médico dos usuários no momento do cadastro. Todas as aplicações de rastreamento de contatos também coletam dados de diagnóstico de COVID-19 para notificar os indivíduos em situação de risco.

Diante da natureza sensível dos dados em questão, a arquitetura de armazenamento de dados tem sido tema de grande debate entre desenvolvedores de aplicativos e especialistas em privacidade. Nas soluções centralizadas, os dados coletados pelos aplicativos de rastreamento de contatos são gravados diretamente em um servidor principal, que costuma ser bastante seguro e de fácil acesso para agências governamentais. Soluções descentralizadas, por outro lado, mantêm registros de proximidade e de localização nos aparelhos dos usuários, em um armazenamento local³³.

Independentemente do protocolo de armazenamento escolhido, a maioria dos aplicativos de rastreamento de contatos recorre à desidentificação dos dados para mitigar as preocupações de privacidade. Embora, na prática, a anonimização completa dos dados não seja possível (Rocher, Hendrickx & Montjoye, 2019), diferentes métodos de pseudonimização costumam ser aspectos integrais das aplicações. Stopp Corona, na Áustria, StopCovid, na França, e TraceTogether, em Singapura, atribuem identificações únicas aos usuários que podem mudar ao longo do tempo para segurança adicional.

Diante da natureza sensível dos dados em questão, a arquitetura de armazenamento de dados tem sido tema de grande debate entre desenvolvedores de aplicativos e especialistas em privacidade.

³¹ Saiba mais: <https://www.amnesty.org/en/latest/news/2020/06/bahrain-kuwait-norway-contact-tracing-apps-danger-for-privacy>

³² Saiba mais: <https://www.financialmirror.com/2020/04/06/technology-recruited-in-fight-against-coronavirus>

³³ Como exemplos das diferentes situações, a solução descentralizada conhecida como DP3T, desenvolvida em parceria pela Google e pela Apple, é utilizada em aplicativos da Finlândia, Malásia e Suíça. Já a França e a Alemanha adotaram o que chamam de "protocolo de rastreamento de contatos centralizado e descentralizado". O protocolo ROBERT (ROBust and privacy-presERving proximity Tracing), por sua vez, armazena dados no dispositivo, mas os envia para um servidor centralizado quando o usuário testa positivo para COVID-19.

Tanto a plataforma Safe Paths, do MIT, quanto os protocolos Google-Apple usam privacidade diferencial para garantir que os dados agregados disponibilizados não permitam a reidentificação dos usuários individuais.

Análise de dados

Os dados coletados pelas aplicações de rastreamento de contatos podem servir a diferentes objetivos dos órgãos governamentais. O uso mínimo consiste em sinalizar a proximidade entre indivíduos infectados e usuários da aplicação, conforme medido por Bluetooth ou GPS. Entretanto, diversos países utilizam os dados para outros propósitos. O plano inicial do Reino Unido para o rastreamento digital de contatos permitia, por exemplo, que os dados coletados pelo aplicativo do Sistema Nacional de Saúde referente à COVID-19 alimentassem também modelos epidemiológicos para estudar a propagação do vírus³⁴.

Riscos éticos levantados pelo rastreamento digital de contatos

A adoção de aplicativos de rastreamento digital de contatos suscita uma série de questões éticas relevantes. Considerando a extensão e a natureza sensível dos dados coletados, essas tecnologias têm o potencial de comprometer liberdades civis e valores humanos fundamentais, como a privacidade, a proteção de dados, a autonomia e a equidade.

Privacidade e proteção de dados

Todo rastreamento digital exige certo grau de acesso a informações que têm o potencial de violar a privacidade, como condições de saúde, localização e dados de cartão de crédito. Dessa forma, riscos à privacidade são inerentes ao uso dessa tecnologia. É possível, por exemplo, que outros aplicativos instalados em um *smartphone* captem a aplicação e enviem dados para terceiros, e dados anonimizados são suscetíveis à reidentificação. De acordo com o *COVID-19 Digital Rights Tracker*, 20% dos aplicativos de rastreamento digital não têm políticas de privacidade – portanto, a falta de informações e de salvaguardas legais, técnicas e políticas levanta grandes preocupações.

Nesse contexto, há dois aspectos importantes: a existência de características *reais* de preservação da privacidade nos próprios aplicativos, de um lado, e o grau de *controle* de privacidade oferecido ao usuário, de outro. A privacidade real implica limitar ao máximo a exposição de informações, enquanto o controle dessa privacidade significa permitir que os indivíduos façam escolhas transparentes sobre o uso de seus dados (Loi et al., 2020). O combate à COVID-19 por meio do rastreamento digital de contatos pode exigir ajustes temporários nos termos de privacidade; contudo, as pessoas devem estar empoderadas para tomar suas próprias decisões.

O uso mínimo consiste em sinalizar a proximidade entre indivíduos infectados e usuários da aplicação, conforme medido por Bluetooth ou GPS. Entretanto, diversos países utilizam os dados para outros propósitos.

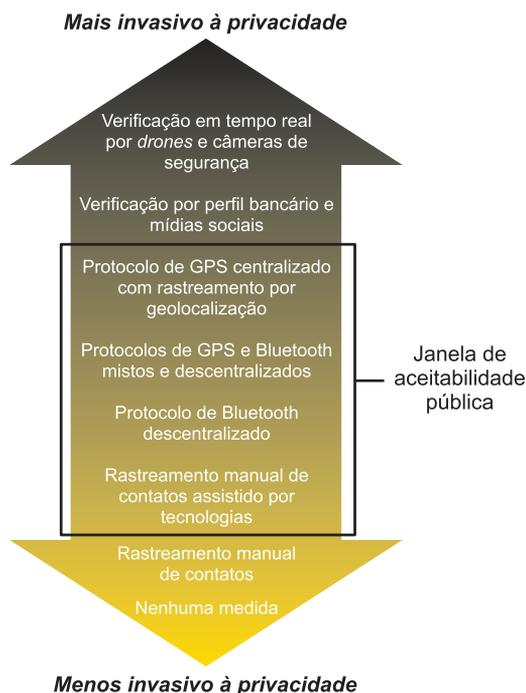
³⁴ Saibamais: <https://social.techcrunch.com/2020/05/05/nhs-COVID-19-the-uks-coronavirus-contacts-tracing-app-explained>

Um aspecto importante é que não necessariamente a pseudonimização dos dados e as medidas técnicas tomadas por aplicativos de rastreamento de contatos nesse sentido são capazes de eliminar o risco. Alegações foram feitas de que várias aplicações, incluindo StopCovid, da França, não preservavam o anonimato de maneira suficiente, apesar da criptografia do usuário. Em 2013, pesquisadores estudaram dados de localização de 1,5 milhão de pessoas e descobriram que estes eram tão específicos em relação aos hábitos individuais que 95% dos usuários podiam ser identificados (Rocher, Hendrickx & Montjoye, 2019). Quando os aplicativos de rastreamento combinam dados de localização com outras informações sensíveis, como acontece em Taiwan e na Coreia do Sul, há uma chance maior de violação da privacidade e até mesmo de vigilância em massa.

Nesse sentido, há preocupações sobre o uso do rastreamento digital de contatos para o estabelecimento de infraestruturas de vigilância por parte de governos e de grandes empresas de tecnologia. No dilema entre o colapso econômico e o desastre social, autoridades públicas implementaram ferramentas de rastreamento de forma veloz, sem passar por procedimentos democráticos e de consentimento informado. Existe o temor de que medidas excepcionais que infringem as liberdades civis permaneçam após a pandemia e se tornem o “novo normal”. Garantir que o rastreamento digital de contatos seja voluntário e temporário é crucial para impedir a expansão do alcance da vigilância para além de seus propósitos iniciais.

(...) não necessariamente a pseudonimização dos dados e as medidas técnicas tomadas por aplicativos de rastreamento de contatos nesse sentido são capazes de eliminar o risco.

Figura 4 – Escala de privacidade dos aplicativos para o combate à COVID-19



Fonte: The Future Society.

Se não forem tomadas medidas específicas destinadas a superar o hiato digital, a base de usuários não refletirá toda a população, e subgrupos serão privados dos serviços da aplicação.

Estigma e discriminação

Preocupações têm sido levantadas sobre o impacto dos aplicativos de rastreamento de contatos em grupos sociais historicamente submetidos a discriminação, estigma e abuso. Há três riscos significativos, em particular nos Estados que impõem sanções criminais à transmissão da COVID-19 ou que utilizam os dados coletados pelas aplicações para decidir sobre um “passaporte” de acesso a certos serviços. O primeiro risco é o de que os dados sejam usados para estigmatizar segmentos específicos da população, como mulheres (Davis, 2020). O segundo, de que os dados de localização sejam empregados para culpar determinadas minorias³⁵. Por fim, o terceiro risco é o de que grupos evitem usar os aplicativos de rastreamento de contatos, tendo vedado, assim, o acesso a serviços ou espaços.

O risco de estigmatização é intrínseco ao rastreamento digital de contatos porque este envolve a coleta de dados sobre populações inteiras. Aplicados para além do propósito de monitorar a exposição à COVID-19, os dados podem levar à identificação de “zonas de incidência” do vírus. Bairros habitados por determinados grupos étnicos se tornam alvo fácil para a discriminação, assim como trabalhadores na linha de frente e de suporte do combate à pandemia, como entregadores de comida, operadores de caixa e motoristas. Quando essas profissões são exercidas por grupos marginalizados, o impacto do rastreamento digital é ainda mais severo. A estigmatização cria alienação e prejudica a coesão social.

Acessibilidade

Os aplicativos de rastreamento de contatos também dependem do acesso à tecnologia e da alfabetização digital para sua adoção em massa. Se não forem tomadas medidas específicas destinadas a superar o hiato digital, a base de usuários não refletirá toda a população, e subgrupos serão privados dos serviços da aplicação. Embora cada vez mais difundido, há lacunas no acesso à tecnologia móvel: segundo o Pew Research Center, em 2019 dois terços da população mundial não possuíam um *smartphone* (no Brasil, a penetração do dispositivo era de 60%)³⁶.

Com frequência, o letramento digital tem alta correlação com outras características socioeconômicas. Uma análise dos usuários do Twitter na Itália, por exemplo, descobriu que eles são, em média, mais jovens e mais instruídos do que a população geral (Vaccari et al., 2013). Cabe apontar ainda que pessoas com deficiência têm sido historicamente marginalizadas pela Internet e pelas plataformas digitais³⁷.

De acordo com Floridi (2020), o aplicativo funcionará melhor quando estiver mais amplamente disseminado. No entanto, isso ocorrerá onde a posse de telefones celulares e o letramento digital forem maiores. Assim, há um risco muito concreto de se privilegiar aqueles que já são privilegiados, bem como suas áreas residenciais. Nesse contexto, o hiato digital pode se tornar um hiato biológico.

³⁵ Saiba mais: <https://news.un.org/en/story/2020/03/1060602>

³⁶ Saiba mais: <https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally>

³⁷ Saiba mais: <http://www.theguardian.com/technology/2015/jun/29/disabled-people-internet-extra-costs-commission-scope>

Em defesa de um *framework* ético para o rastreamento digital de contatos

A tarefa de desenvolver e adotar um *framework* ético para navegar pelos conflitos entre saúde pública, segurança, atividade econômica e liberdades civis é uma responsabilidade coletiva. A urgência da situação provocada pela COVID-19 não deve afetar a capacidade de construção de um marco de referência abrangente para guiar as decisões dos cidadãos, dos desenvolvedores de aplicativos e dos formuladores de políticas públicas. Quando as escolhas são voluntárias e eticamente justificadas, indivíduos informados se tornam mais dispostos a aceitar sacrifícios potencialmente significativos mas necessários. Assim, o estabelecimento de um *framework* ético e confiável é fundamental para garantir a coesão social e a efetividade dos aplicativos de rastreamento de contatos.

Quadro de princípios éticos

Objetivo e desempenho	O objetivo dos aplicativos precisa ser claro, compreensível dentro do contexto mais amplo, mensurável e passível de auditoria independente.
Voluntariedade e reversibilidade	Os indivíduos devem ter o direito de escolher se desejam instalar ou não os aplicativos conforme seu livre-arbítrio, sem consequências negativas em caso de recusa. Os usuários devem ser capazes de desativar os aplicativos temporária ou permanentemente a qualquer momento, sem que os dados pessoais ou as informações de proximidade remanescentes sejam armazenados pelos desenvolvedores das aplicações ou por terceiros.
Privacidade by Design	Os aplicativos de rastreamento de contatos devem alcançar os níveis mais altos de proteção de privacidade. O armazenamento dos dados deve ser seguro e pseudonimizado.
Uso mínimo de dados e tecnologias	A coleta de dados deve ser proporcional, justificada e ter uma data de validade definida. Apenas os dados minimamente necessários para cumprir com o objetivo dos aplicativos devem ser usados e armazenados.
Transparência e verificabilidade	O código-fonte completo dos aplicativos e os protocolos de rastreamento centrais devem ser de livre acesso e reprodução, sem restrição para auditorias.
Não discriminação e não estigmatização	Desenvolvedores de aplicativos e formuladores de políticas públicas devem garantir que os aplicativos de rastreamento de contatos não estigmatizem nem discriminem pessoas que testaram positivo para COVID-19 ou seus familiares, categorias de trabalhadores, bairros ou aqueles que não desejam usar as aplicações.
Acessibilidade	Deve-se reconhecer que aplicativos de <i>smartphone</i> e conexão à Internet não são acessíveis a toda a população. Alguns cidadãos podem não ter <i>smartphones</i> , e pessoas com deficiência, idosas ou sem amplo conhecimento sobre tecnologia talvez não consigam usar as aplicações. É preciso desenvolver soluções complementares e alternativas para assegurar a acessibilidade.
Aviso e consentimento informado	Informações sobre os objetivos, os recursos dos aplicativos e os dados coletados devem ser claramente apresentadas aos usuários. O consentimento informado e explícito deve ser um pré-requisito para as aplicações. Devem ser evitados padrões de <i>dark design</i> , ou desenho oculto (por exemplo, incômodos incentivos via notificações <i>push</i> , aplicativos pré-instalados em <i>smartphones</i> e ocultação de recursos para desativá-los ou removê-los).
Prestação de contas	Aplicativos de rastreamento de contatos devem ser avaliados continuamente, fiscalizados por entidades independentes e legítimas nas quais o público confie. Todas as partes interessadas envolvidas no desenho e na implementação dos aplicativos devem prestar contas de acordo com um claro arcabouço legal de responsabilidades e penalidades.

Conclusão

Diante do crescimento no número de casos de COVID-19 ao redor do mundo, é uma tarefa coletiva explorar todas as soluções médicas e tecnológicas, mas também gerenciá-las com responsabilidade. Nos últimos anos tem-se erodido a confiança da população nas instituições públicas e nas empresas de tecnologia, tornando ainda mais relevante a busca proativa pela conscientização e pelo consentimento da sociedade antes de implementar aplicativos de rastreamento de contatos. Um *framework* ético pode ajudar ao construir um entendimento comum e nortear as escolhas das diferentes partes interessadas. Para que esse marco possa ser usado na prática, seus princípios precisam ser traduzidos em critérios específicos, bem como estar associados a parâmetros de auditoria desenvolvidos e implementados por terceiros, de forma a aumentar a confiança entre os usuários e os demais atores envolvidos.

Referências

- Arons, M. M. et al. (2020). Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. *New England Journal of Medicine*, 382, 2081.
- Davis, S. (2020, 29 Abril). *Contact tracing apps: Extra risks for women and marginalized groups*. Health and Human Rights Journal. Recuperado de <https://www.hhrjournal.org/2020/04/contact-tracing-apps-extra-risks-for-women-and-marginalized-groups>.
- Floridi, L. (2020, 18 Abril). *Mind the app: Considerations on the ethical risks of COVID-19 apps*. Onlife. Recuperado de <https://thephilosophyofinformation.blogspot.com/2020/04/mind-app-considerations-on-ethical.html>.
- Governo da França. (2020). *ProjetStopCovid: Dossier de presse (21 mai 2020)*. Recuperado de <https://decryptageo.fr/wp-content/uploads/2020/05/Dossier-de-presse-Pojet-StopCovid.pdf>.
- Hart, V. et al. (2020). Outpacing the virus: Digital response to containing the spread of COVID-19 while mitigating privacy risks. *Edmond J. Safra Center for Ethics*. Recuperado de <https://ethics.harvard.edu/outpacing-virus>.
- Lauer, S. A. et al. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine*, 172, 577.
- Li, Q. et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, 382, 1199.
- Loi, M. et al. (2020). *Ethical framework for human-centric public health digital surveillance*. Recuperado de https://docs.google.com/document/d/19F_hXIIpVdKck&JTFxNXOuex0r_GpmPEUuoSUMEj-zY/edit.
- Mizumoto, K. et al. (2020). Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance*, 25, 2000180.
- Ornell, F., Schuch, J. B., Sordi, A. O., & Kessler, F. H. P. (2020). “Pandemic fear” and COVID-19: mental health burden and strategies. *Brazilian Journal of Psychiatry*, 42(3), 232-235.
- Rocher, L., Hendrickx, J. M., & Montjoye, Y. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*, 10, 3069.
- Vaccari, C. et al. (2013). Social media and political communication: A survey of Twitter users during the 2013 Italian general election. *Rivista Italiana di Scienza Politica*.

Entrevista I

P.S._ *O debate sobre as possíveis contribuições da Inteligência Artificial (IA) para a área da Saúde tem ganhado força. Quais são os desafios enfrentados na implementação de IA no setor?*

E.M._ A utilização de algoritmos e de ferramentas de IA para auxiliar médicos, enfermeiros e demais profissionais em sua prática tem crescido de maneira constante. Essas soluções têm o potencial de melhorar a eficiência, a eficácia e a equidade de acesso na área da Saúde. No entanto, embora a visão geral seja otimista, é importante considerar os desafios envolvidos na tarefa. A discrepância entre os esforços no desenvolvimento de modelos e de soluções de IA, por um lado, e o sucesso no seu uso, por outro, reflete as dificuldades na implementação de ferramentas de apoio à decisão de maneira geral. Ainda que existam obstáculos técnicos, grande parte do desafio está: na complexidade de ajustar as aplicações para que se integrem a sistemas como prontuários eletrônicos; no desalinhamento em relação às reais necessidades dos usuários; na falta de gestão de expectativas; e nos aspectos éticos e legais. Todos esses fatores precisam ser equacionados para que haja um verdadeiro benefício.

A IA oferece muitas oportunidades em soluções que auxiliem no diagnóstico de problemas clínicos, na terapia de doenças, na redução de custos, gerando um impacto bastante positivo na área da Saúde como um todo. Entretanto, devemos estar cientes da capacidade e das limitações atuais dessas tecnologias para que se evite a frustração e, conseqüentemente, um movimento antagonista que prejudique o progresso da IA no setor. O ideal é um equilíbrio realista entre aquilo que pode ser feito e aquilo que é adequado. Informar e educar todos os públicos envolvidos – informatas, engenheiros, gestores, profissionais de saúde, população geral – em relação aos princípios e ao uso de tais ferramentas é fundamental.

P.S._ *No contexto do ecossistema de dados no setor da Saúde, em que há grande heterogeneidade de fontes e alto volume de dados disponíveis, como deve se dar a governança de IA?*

E.M._ O engajamento da comunidade científica, dos profissionais de saúde e dos cidadãos é essencial para que a integração e o uso dos dados sejam equitativos e transparentes. A governança de IA para a prática clínica deve incluir não só os desenvolvedores de tecnologia, mas também representantes dos provedores de serviços, dos gestores, dos usuários e dos pacientes. Ela deve considerar uma análise tanto do desempenho da aplicação quanto de seu potencial custo-benefício para a melhoria da saúde e a redução dos custos de atendimento.

Para que tenham sucesso na área da Saúde, os algoritmos de IA precisam ser treinados com dados representativos das características da população real – só assim os modelos gerados conseguem atingir resultados com a precisão adequada. Diante do aumento da quantidade e da diversidade de



Dra. Eneida A. Mendonça

Professora de Pediatria e Biostatística na Universidade de Indiana, vice-presidente de Desenvolvimento de Pesquisa e diretora interina do Centro Clem McDonald para Informática Biomédica no Instituto Regenstrief.

"Métodos de aprendizado de máquina costumam mostrar bom desempenho em extrair características, padrões e regras a partir de dados complexos e em grande volume (*Big Data*), mas não produzem significado, propósito, senso de justiça ou equidade."

dados relativos ao setor (prontuários eletrônicos, dados genéticos e determinantes sociais de saúde, por exemplo), faz-se necessária a criação de um ecossistema que possibilite o armazenamento, a integração e a análise destes dados.

Diversas opções de estruturas de dados e padrões de representação da informação biomédica já estão disponíveis, como códigos CID, LOINC e SNOMED. Entretanto, seu emprego não é universal, reduzindo substancialmente a capacidade de uso em IA. A adoção de *common data models* – por exemplo, *Observational Medical Outcomes Partnership* (OMOP) e *Informatics for Integrating Biology & the Bedside* (i2b2) –, bem como de padrões como o HL7 FHIR, que permitem a transferência de dados de forma padronizada, precisa ser priorizada, junto com a continuidade das discussões sobre políticas de troca de dados, governança, mecanismos de avaliação de sistemas e de algoritmos.

P.S._ Uma vez que as soluções de IA são baseadas em dados, quais são os riscos envolvidos na sua implementação em Saúde? Como contorná-los?

E.M._ Além da complexidade dos dados usados no treinamento de modelos, o ciclo de desenvolvimento, validação e implementação de ferramentas de IA exige que sejam adotados modelos muito controlados e rigorosos do ponto de vista técnico-científico. Na prática, em termos de riscos aos pacientes, esses algoritmos se assemelham a medicamentos e equipamentos médicos. Os modelos e as ferramentas também devem ser avaliados em relação ao desempenho, à utilidade, às vulnerabilidades e aos vieses. Variações nos padrões locais de atendimento e de terapias, populações com características diferentes e vieses na seleção dos dados de aprendizado e de validação são exemplos de fatores que podem afetar significativamente o uso dos algoritmos e causar consequências indesejáveis.

Métodos de aprendizado de máquina costumam mostrar bom desempenho em extrair características, padrões e regras a partir de dados complexos e em grande volume (*Big Data*), mas não produzem significado, propósito, senso de justiça ou equidade. É essencial reconhecer que os algoritmos “aprendem” baseados somente nos dados que os treinam – se o conjunto de dados contém vieses, o algoritmo terá o potencial de amplificá-los. Também é importante que haja transparência sobre como os dados são usados e os modelos, validados. Além da reprodutibilidade em diferentes populações, o balanço entre inovação, segurança e privacidade dos dados dos indivíduos precisa ser respeitado.

Por fim, sugiro a leitura de uma publicação recente da *National Academy of Medicine*³⁸ sobre a implementação adequada de aplicações da IA na área da Saúde. Os autores do documento apresentam recomendações detalhadas e oferecem uma reflexão interessante para a comunidade.

³⁸ Disponível em: <https://nam.edu/artificial-intelligence-special-publication>

Entrevista II

P.S._ *Existe um marco de referência proposto pela OMS para regulamentar, comparar e certificar métodos baseados em Inteligência Artificial (IA) no campo da saúde digital? Se sim, como esses processos são implementados na prática?*

B.M.J._ Na luta contra a COVID-19, tecnologias como a IA são usadas para ajudar na triagem da população, no rastreamento de casos de infecção, no monitoramento de recursos e principalmente para definir os determinantes sociais da saúde, elemento fundamental de equidade e direitos humanos nessa luta. Contudo, embora a IA e tecnologias similares tenham um potencial enorme, é preciso abrir espaço para discutir questões éticas em torno de seu uso na vida real. Temos de garantir que a IA seja empregada corretamente nos sistemas de saúde, em especial no que diz respeito ao acesso equitativo e à privacidade dos pacientes.

A OMS está desenvolvendo a Estratégia Global de Saúde Digital, que propõe um marco de referência para regulamentar, realizar avaliações comparativas (*benchmarking*) e certificar a IA e os dispositivos médicos de saúde digital para implantação pelos Estados-Membros. Em parceria com a União Internacional de Telecomunicações (UIT), a OMS está trabalhando em um guia de *benchmarking* para IA em programas de saúde. Foram criados grupos de especialistas sobre ética da IA e regulamentos na área para desenvolver o marco de referência.

O marco terá como foco os seguintes aspectos:

- Garantir acesso equitativo à IA;
- Estabelecer de que forma a IA interage com o problema do hiato digital;
- Preservar os direitos individuais de autonomia, privacidade, consentimento informado, proteção contra o preconceito e a discriminação;
- Garantir educação sobre como a IA funciona e toma decisões;
- Garantir igualdade de acesso às bases de dados, independentemente dos recursos financeiros disponíveis;
- Manter o controle humano sobre a IA; e
- Reforçar a vigilância pública e a regulação do setor privado.

Além disso, em julho de 2018 a OMS e a UIT criaram o Grupo Focal em Inteligência Artificial para a Saúde (FG-AI4H), que desenvolve um processo de *benchmarking* para modelos de IA em saúde. Tal processo pode ser usado como um marco de avaliação internacional, independente e padronizado.



**Sr. Bernardo
Mariano Junior**

Chief Information Officer (CIO) e diretor do Departamento de Saúde Digital e Inovação da Organização Mundial da Saúde (OMS).

"O cerne da EIOS é criar uma comunidade de práticas de inteligência em saúde pública que inclua Estados-Membros, organizações internacionais, institutos de pesquisa e outros parceiros, com o objetivo principal de salvar vidas".

P.S._ Há iniciativas específicas de IA implementadas ou monitoradas pela OMS no contexto da COVID-19?

B.M.J._ A OMS lidera desde setembro de 2017 a iniciativa Inteligência Epidemiológica de Fontes Abertas (EIOS, *Epidemic Intelligence from Open Sources*), uma colaboração única entre vários atores de saúde pública no mundo. A iniciativa reúne redes e sistemas novos ou já existentes para criar uma abordagem integrada *all-hazards* (focada em todo o espectro de emergências ou desastres) e *One Health* (ações colaborativas, multissetoriais e transdisciplinares) para a detecção precoce, a verificação, a avaliação e a comunicação de ameaças à saúde pública, utilizando informações disponíveis a todos. O cerne da EIOS é criar uma comunidade de práticas de inteligência em saúde pública que inclua Estados-Membros, organizações internacionais, institutos de pesquisa e outros parceiros, com o objetivo principal de salvar vidas.

Essa comunidade de práticas tem o apoio de um sistema EIOS em evolução para a inteligência em saúde pública que não apenas conecta sistemas e atores – incluindo os programas de monitoramento ProMED, HealthMap e a Rede Global de Inteligência sobre Saúde Pública (GPHIN, *Global Public Health Intelligence Network*) –, mas também promove e catalisa desenvolvimentos colaborativos inovadores por meio do uso da IA e da Inteligência Aumentada, especialmente no domínio da linguística computacional e do processamento de linguagem natural. O sistema EIOS é construído com base em uma colaboração antiga com o Centro Comum de Investigação da Comissão Europeia para responder à necessidade de uma iniciativa global que reúna esforços de inteligência em saúde pública.

Desde o início da pandemia de COVID-19, membros da comunidade EIOS trabalham para desenvolver recursos avançados de IA e aprendizado de máquina que ajudem a gerenciar o volume sem precedentes de informações de fontes oficiais e não oficiais. Estão sendo investigadas formas aprimoradas de filtrar, contextualizar e visualizar todo o fluxo de conteúdo, como a introdução de dados adicionais das redes sociais (Twitter) e o reconhecimento da confiabilidade de artigos de notícia a partir da análise do tom e do estilo da escrita.

Relatório de Domínios

A dinâmica dos registros de domínios no Brasil e no mundo

O Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br) monitora mensalmente o número de nomes de domínios de topo de código de país (do inglês, *country code Top-Level Domain* – ccTLD) registrados entre os países do G20³⁹. Somados, eles ultrapassam 79,50 milhões de registros. Em agosto de 2020, os domínios registrados sob .de (Alemanha) chegaram a 16,53 milhões. Em seguida, aparecem China (.cn), Reino Unido (.uk) e Rússia (.ru), com, respectivamente, 16,05 milhões, 9,50 milhões e 4,96 milhões de registros. O Brasil teve 4,41 milhões de registros sob .br, ocupando a quinta posição na lista, como mostra a Tabela 1⁴⁰.

Tabela 1 – Registro de nomes de domínios entre os países do G20 – Agosto/2020

Posição	Países G20	Domínios	Fonte
1	Alemanha (.de)	16.531.825	www.denic.de
2	China (.cn)	16.048.918	research.domaintools.com/statistics/tld-counts/
3	Reino Unido (.uk)	9.501.094	www.nominet.uk/news/reports-statistics/uk-register-statistics-2020/
4	Rússia (.ru)	4.960.220	ccTld.ru
5	Brasil (.br)	4.408.632	registro.br/dominio/estatisticas/
6	França (.fr)	3.566.512	www.afnic.fr/en/resources/statistics/detailed-data-on-domain-names/
7	União Européia (.eu)	3.531.400	research.domaintools.com/statistics/tld-counts/
8	Itália (.it)	3.315.268	nic.it
9	Austrália (.au)	3.205.126	www.auda.org.au/
10	Canadá (.ca)	2.935.244	www.cira.ca
11	Índia (.in)	2.200.000	www.registry.in/
12	Estados Unidos da América (.us)	1.685.195	research.domaintools.com/statistics/tld-counts/
13	Japão (.jp)	1.596.328	jprs.co.jp/en/stat/
14	África do Sul (.za)	1.264.123	www.zadna.org.za
15	Coréia do Sul (.kr)	1.107.229	krnic.or.kr/jsp/eng/domain/kr/statistics.jsp
16	México (.mx)	910.012	research.domaintools.com/statistics/tld-counts/
17	Argentina (.ar)	612.656	nic.ar/es/dominios/estadisticas
18	Indonésia (.id)	434.525	pandi.id/?lang=en
19	Turquia (.tr)	430.555	www.nic.tr/index.php?USRAC TN=STATISTICS
20	Arábia Saudita (.sa)	69.293	www.nic.sa/en/view/statistics

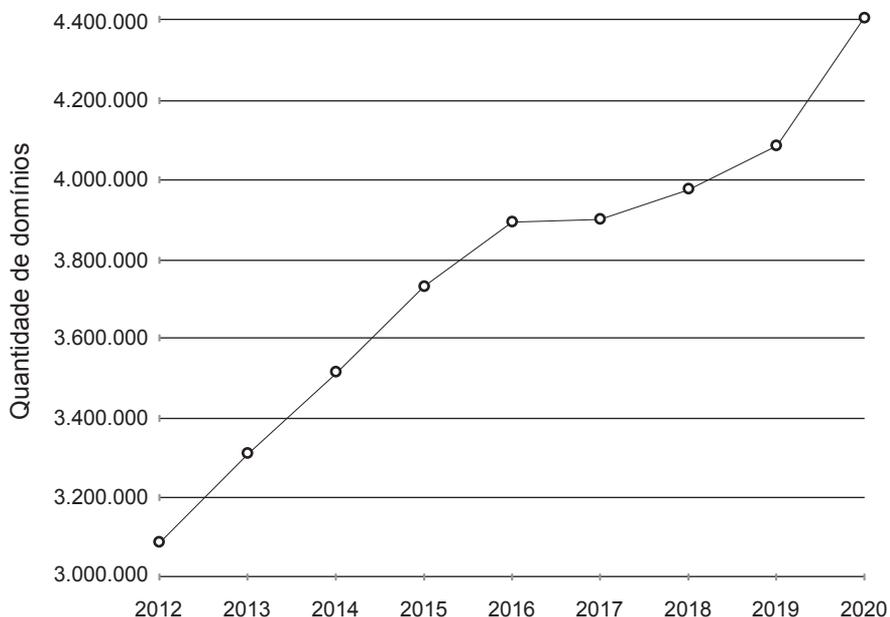
Data da coleta: 31/08/2020

³⁹ Grupo composto pelas 19 maiores economias mundiais e a União Europeia. Saiba mais: <https://g20.org/en/Pages/home.aspx>

⁴⁰ A tabela apresenta a contagem de domínios ccTLDs segundo as fontes indicadas. Os valores correspondem ao registro publicado por cada país do G20. Para países que não apresentam ou publicam uma estatística oficial fornecida pela autoridade de registro de nomes de domínios, a contagem foi obtida em: <https://research.domaintools.com/statistics/tld-counts>. É importante destacar que há variação no período de referência, embora seja sempre o mais atualizado para cada localidade. A análise comparativa de desempenho de nomes de domínios deve considerar ainda os diferentes modelos de gestão de registros ccTLDs. Assim, ao observar o ranking é preciso atentar para a diversidade de modelos de negócio existentes.

O Gráfico 1 apresenta o desempenho do .br desde o ano de 2012.

Gráfico 1 – Total de registros de domínios ao ano do .Br – 2012 a 2020*



* Dado referente ao mês de agosto de 2020.
Fonte: Registro.br

Em agosto de 2020, os cinco principais domínios genéricos (do inglês, generic *Top-Level Domain* – gTLD) totalizaram mais de 182 milhões de registros. Com 149,28 milhões de registros, destaca-se o .com, conforme apontado na Tabela 2.

Tabela 2 - Principais gTLDs - Agosto/2020

Posição	gTLD	Domínios
1	.com	149.279.349
2	.net	13.246.836
3	.org	10.240.183
4	.icu	5.618.105
5	.info	4.298.038

Fonte: DomainTools.com
Recuperado de: research.domaintools.com/statistics/tld-counts

VOCÊ SABE DE QUE MANEIRA OS ESTABELECIMENTOS DE SAÚDE BRASILEIROS REGISTRAM AS INFORMAÇÕES CLÍNICAS DOS PACIENTES? QUAIS DADOS SÃO COLETADOS EM FORMATO ELETRÔNICO?

Conheça estes e outros indicadores sobre a produção de dados no setor de Saúde⁴¹, matéria-prima para as aplicações de Inteligência Artificial (IA).

92%

DO TOTAL DE ESTABELECIMENTOS DE SAÚDE UTILIZARAM INTERNET NOS ÚLTIMOS 12 MESES⁴².



DESTES ESTABELECIMENTOS:

82%

POSSUEM SISTEMA ELETRÔNICO PARA O REGISTRO DAS INFORMAÇÕES DOS PACIENTES

64%

EM A MANUTENÇÃO DAS INFORMAÇÕES CLÍNICAS E CADASTRAIS NO PRONTUÁRIO DOS PACIENTES É FEITA PARTE EM PAPEL, PARTE EM FORMATO ELETRÔNICO

18%

EM APENAS EM PAPEL

18%

EM APENAS EM FORMATO ELETRÔNICO

⁴¹ Com base nos dados da pesquisa TIC Saúde 2019, do Cetic.br/NIC.br. Os indicadores se referem aos 12 meses que antecederam a pesquisa. Saiba mais: <https://cetic.br/pt/pesquisa/saude/indicadores/>.

⁴² Do total de estabelecimentos de saúde.

/Tire suas dúvidas

QUAIS DADOS SOBRE O PACIENTE ESTÃO DISPONÍVEIS ELETRONICAMENTE NOS ESTABELECIMENTOS DE SAÚDE?

[89%] DADOS CADASTRAIS

[65%] HISTÓRICO/ANOTAÇÕES CLÍNICAS SOBRE O ATENDIMENTO

[64%] DIAGNÓSTICO E CONDIÇÕES DE SAÚDE

[54%] ALERGIAS

[46%] VACINAS TOMADAS

[61%] RESULTADOS DE EXAMES LABORATORIAIS

[53%] LISTA DE MEDICAMENTOS PRESCRITOS

[56%] ADMISSÃO, TRANSFERÊNCIA E ALTA

[64%] PRINCIPAIS MOTIVOS QUE O LEVARAM AO ATENDIMENTO/CONSULTA

[51%] ANOTAÇÕES DE ENFERMAGEM

[38%] LAUDOS DE EXAMES RADIOLÓGICOS

[24%] IMAGENS DE EXAMES RADIOLÓGICOS



CREATIVE COMMONS
Atribuição 2.0
Genérica (cc-by-2.0)



/Créditos

REDAÇÃO

ARTIGO I

Edson Amaro Jr, Michel Fornaciali,
Andre Batista, Murilo Gazzola,
Lívia Paiva da Silva, Diogo F. C. Patrão,
Marcos Freitas Jr
(Hospital Israelita Albert Einstein - HIAE)

ARTIGO II

Sacha Alanoca, Nicolas Guetta
Jeanrenaud, Nyasha Weinberg,
R. Buse Çetin, Nicolas Miallhe
(The Future Society)

RELATÓRIO DE DOMÍNIOS

José Márcio Martins Júnior
(Cetic.br/NIC.br)

INFOGRAFIA

Giuliano Galvez, Klezer Uehara,
Maricy Rabelo
(Comunicação/NIC.br)

DIAGRAMAÇÃO

Grappa Marketing Editorial

TRADUÇÃO

Isabela Ayub, Maya Bellomo-Johnson,
Luana Guedes, Luísa Caliri
(Prioridade Ltda.)

EDIÇÃO DE TEXTO EM PORTUGUÊS

Mariana Tavares

COORDENAÇÃO EDITORIAL

Alexandre Barbosa, Tatiana Jereissati,
Stefania L. Cantoni
(Cetic.br/NIC.br)

AGRADECIMENTOS

Edson Amaro Jr, Michel Fornaciali, Andre
Batista, Murilo Gazzola, Lívia Paiva da
Silva, Diogo F. C. Patrão, Marcos Freitas Jr
(Hospital Israelita Albert Einstein - HIAE)
Sacha Alanoca, Nicolas Guetta
Jeanrenaud, Nyasha Weinberg, R. Buse
Çetin, Nicolas Miallhe (The Future Society),
Eneida A. Mendonça (Universidade de
Indiana/Instituto Regenstrief),
Bernardo Mariano Junior (Organização
Mundial da Saúde – OMS)



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

cetic.br

Centro Regional de Estudos
para o Desenvolvimento da
Sociedade da Informação
sob os auspícios da UNESCO

nic.br

Núcleo de Informação
e Coordenação do
Ponto BR

egi.br

Comitê Gestor da
Internet no Brasil

CREATIVE COMMONS

Atribuição

Uso Não Comercial
Não a Obras Derivadas
(by-nc-nd)





POR UMA INTERNET CADA VEZ MELHOR NO BRASIL

CGI.BR, MODELO DE GOVERNANÇA MULTISSETORIAL

www.cgi.br

nic.br cgi.br