Uso responsável da IA para políticas públicas:

Manual de Formulação de Projetos

Gabriela Denis
María Paz Hermosilla
Claudio Aracena
Roberto Sánchez Ávalos
Natalia González Alarcón
Cristina Pombo

BID

Uso responsável da IA para políticas públicas:
Manual de Formulação de Projetos

Gabriela Denis
María Paz Hermosilla
Claudio Aracena
Roberto Sánchez Ávalos
Natalia González Alarcón
Cristina Pombo

Septiembre de 2021

https://www.iadb.org/

Copyright © 2021 Banco Interamericano de Desenvolvimento. Esta obra está licenciada sob uma licença Creative Commons IGO 3.0 Atribuição-NãoComercial-SemDerivações (CC BY-NC-ND 3.0 IGO) (http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode) e pode ser reproduzida com atribuição ao BID e para qualquer finalidade não comercial. Nenhum trabalho derivado é permitido.

Qualquer controvérsia relativa à utilização de obras do BID que não possa ser resolvida amigavelmente será submetida à arbitragem em conformidade com as regras da UNCITRAL. O uso do nome do BID para qualquer outra finalidade que não a atribuição, bem como a utilização do logotipo do BID serão objetos de um contrato por escrito de licença separado entre o BID e o usuário e não está autorizado como parte desta licença CC-IGO.

Note-se que o link fornecido acima inclui termos e condições adicionais da licença.

As opiniões expressas nesta publicação são de responsabilidade dos autores e não refletem necessariamente a posição do Banco Interamericano de Desenvolvimento, de sua Diretoria Executiva, ou dos países que eles representam.

Tradução de Mariana Fagundez, otraspalabras.com.

Agradecimientos

Expressamos a nossa gratidão a Erika Quiroz, Sebastián Elgueta, Rodrigo Moya, Claudio Collao, José Tomás Arenas e Romina Garrido, pelo material relevante que nos forneceram para o desenvolvimento do manual, bem como a Patricia Ardila, pelo seu trabalho de edição, e a Alejandro Scaff, pela diagramação.

ÍNDICE

SOBRE O MANUAL 5

INTRODUÇÃO 6

Sistemas de tomada de decisão e/ou suporte à decisão e aprendizado de máquina 7

Componentes de um sistema de IA para políticas públicas 8

PARTE 1. CONCEITUALIZAÇÃO E FORMULAÇÃO 12

1.1 Definição do problema 15

1.2 Análise de pré-viabilidade 21

1.3 Definição dos objetivos 24

1.4 Descrição das ações 28

1.5 Mapeamento de dados 32

1.6 Análise/ferramenta 36

1.7 Considerações éticas, jurídicas e de governança 40

1.8 Formação da equipe 50

PARTE 2. EXECUÇÃO 53

2.1 Coleta e processamento de dados 56

2.2 Desenvolvimento e validação do modelo 61

2.3 Uso e monitoramento 65

2.4 Prestação de contas 68

O que aconteceu com o DART na etapa de execução? 70

CONCLUSÕES 73

REFERÊNCIAS 75

ANEXOS 79

Anexo 1. Ficha de formulação e viabilidade do projeto 80

Anexo 2. Matriz de maturidade dos dados 89

Anexo 3. Lista de verificação para o diretor do projeto 91

Anexo 4. Perfil de dados 93

Anexo 5. Perfil do modelo 95

SOBRE O MANUAL

Iniciativa fAIr LAC

O Banco Interamericano de Desenvolvimento (BID), em colaboração com parceiros e aliados estratégicos, lidera a iniciativa fAIr LAC[1], por meio da qual busca promover a adoção responsável da inteligência artificial (IA) e dos sistemas de suporte à decisão. O objetivo é melhorar a prestação de serviços sociais e criar oportunidades de desenvolvimento, a fim de reduzir a desigualdade social. Este manual faz parte de uma série de ferramentas e documentos elaborados para orientar os[2] responsáveis pela formulação de políticas públicas e suas equipes técnicas na mitigação dos desafios inerentes aos sistemas de suporte à decisão baseados em IA e na promoção de sua adoção responsável (Cabrol et al., 2020).

Por que este manual?

Este manual tem como finalidade ajudar os responsáveis pela formulação de projetos com sistemas de suporte baseados em IA a realizarem sua conceitualização e formulação, bem como a conduzirem, posteriormente, sua execução e monitoramento. A IA é uma ferramenta muito poderosa que pode ajudar a resolver questões complexas, desde que seja contextualizada dentro do problema de política pública a ser solucionado, abordando-se as questões éticas e jurídicas envolvidas na aplicação de ferramentas automáticas de tomada de decisão.

A quem este manual se dirige?

Este manual dirige-se aos tomadores de decisão (gerentes, diretores ou profissionais não necessariamente especialistas em ciência de dados) de instituições públicas que conduzam projetos de IA, da fase de formulação até sua implementação. O documento foi dividido em duas partes principais: **conceitualização e formulação**, cujos trabalhos ficarão a cargo do diretor do projeto (responsável pela tomada de decisão), e **execução**, em que o diretor desempenha um papel crucial nas tarefas que serão executadas em conjunto com uma equipe multidisciplinar, composta pela equipe técnica de desenvolvedores de modelos de IA, especialistas do setor e especialistas jurídicos, entre outros.

Este manual complementa o manual técnico de IA responsável dirigido principalmente às equipes técnicas, disponível em https://publications.iadb.org/es/ia-responsable-manual-tecnico-ciclo-de-vida-de-la-inteligencia-artificial.

1 Para mais informações, consulte https://fairlac.iadb.org/.

2 Por motivos puramente estilísticos, neste documento, utiliza-se o masculino genérico inclusivo, independentemente do gênero gramatical. Os cargos e funções corresponderão indistintamente a pessoas de ambos os sexos.

INTRODUÇÃO

Atualmente, os sistemas de suporte à decisão baseados em inteligência artificial (IA) permitem processar enormes quantidades de dados para gerar recomendações, previsões ou classificações que podem ser usadas para aprimorar diferentes processos. Estamos cada vez mais familiarizados com suas aplicações no nosso cotidiano: de sugestões de possíveis amizades nas nossas redes sociais ou de séries ou filmes que gostaríamos de ver nas plataformas de streaming a anúncios que respondem especificamente aos nossos interesses e gostos, passando pela utilização de técnicas de reconhecimento facial para desbloquear os nossos dispositivos eletrônicos.

No entanto, a IA também pode ser usada para solucionar problemas de política pública, e cada vez mais instituições estão implementando soluções baseadas em IA, com as quais procuram gerar um impacto positivo no bem-estar da sociedade.

As ferramentas de IA, junto com a capacidade dos computadores modernos de processar informações, têm permitido que sejam aplicadas nos diferentes afazeres das instituições, solucionando problemas internos e externos. A seguir, serão mencionados alguns usos da IA em entidades públicas que atuam em diversas áreas de atividade da sociedade:

- Distribuição de professores ou alunos pelas escolas.[3]
- Prevenção ou tratamento em tempo hábil de doenças por meio da detecção precoce.[4]
- Correspondência entre vagas de emprego e candidatos com base em informações mais completas.[5]
- Resposta automática às solicitações dos usuários e beneficiários de uma instituição.
- Direcionamento da concessão de subsídios e benefícios à população destinatária.
- Supervisão do uso de autorizações de acesso a fontes de água por meio da análise de imagens.
- Fiscalização da poluição causada pelas empresas com base nos dados das estações de monitoramento.
- Direcionamento do patrulhamento policial.
- Previsão de tempos de deslocamento e espera no transporte público.

No documento "Inteligência artificial a serviço do bem social na América Latina e no Caribe: panorama da região e retrato de doze países" da iniciativa fAIr LAC[6] (Gómez, del Pozo, Martínez e Martín del Campo, 2020), pode ser encontrada uma descrição detalhada de algumas iniciativas de aplicação da IA para o bem social na região.

3 Para mais informações sobre a aplicação da IA na educação, clique aqui.

4 Para mais informações sobre a aplicação da IA na saúde, clique aqui.

5 Para obter mais informações sobre o uso da IA em sistemas de pesquisa de emprego, clique aqui.

6 Para mais informações sobre a iniciativa, consulte https://fairlac.iadb.org/

Sistemas de tomada de decisão e/ou suporte à decisão e aprendizado de máquina

A OCDE descreve os sistemas de suporte à decisão como "sistemas computacionais que podem fazer previsões e recomendações ou tomar decisões que influenciam ambientes reais ou virtuais para um determinado conjunto de objetivos definidos pelos seres humanos". Esses sistemas são projetados para operar com diferentes níveis de autonomia (OECD, 2019).

Este manual busca abordar os aspectos mais comuns referentes ao uso de sistemas de suporte à decisão e tomada de decisão, através da lente das pessoas responsáveis, inclusive a determinação da viabilidade do projeto, detecção de vieses e avaliação da possibilidade de estar gerando resultados indesejáveis para a sociedade ou para uma determinada instituição.

Embora os métodos de aprendizado de máquina (AM) não sejam o único tipo de algoritmo que pode ser usado por sistemas de IA, eles tiveram o maior crescimento nos últimos anos. Trata-se de um conjunto de técnicas que permitem que um sistema aprenda comportamentos automaticamente por meio de padrões e inferências, não de instruções explícitas ou simbólicas inseridas por seres humanos (OCDE, 2019). Existem dois arquétipos de uso do aprendizado de máquina em processos decisórios (González, Ortiz e Sánchez Ávalos, 2020): os sistemas de suporte à decisão e os sistemas de tomada de decisão.

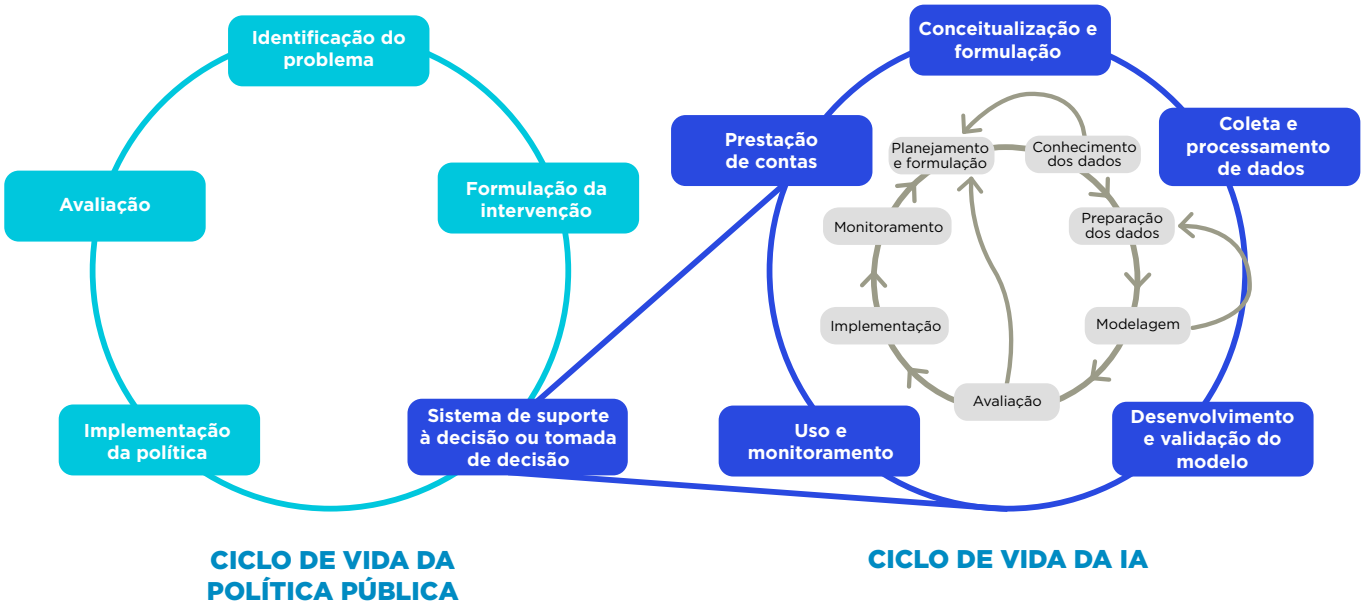
Sistemas de suporte à decisão: relacionados ao conceito de inteligência assistida ou aumentada, incluem os sistemas nos quais as informações geradas pelos modelos de AM são utilizadas como base para a tomada de decisão realizada por seres humanos.

Sistemas de tomada de decisão: relacionados ao conceito de inteligência automatizada e autônoma, as decisões finais e as ações decorrentes delas são tomadas sem intervenção humana direta. Isso significa que o sistema passa a realizar tarefas que antes eram executadas por pessoas. Em muitos contextos, usa-se a sigla ADM para denominar esses sistemas (abreviação em inglês de *Automated Decision Making*).

Componentes de um sistema de IA para políticas públicas

A IA, ou a criação de um sistema de suporte à decisão/tomada de decisão com base nela, não substitui as políticas públicas, pois a IA por si só não resolve nenhum problema social. É uma ferramenta utilizada durante o ciclo de vida das políticas públicas para fornecer informações na forma de previsão, classificação e/ou segmentação, entre outras possibilidades, no contexto da formulação de uma intervenção ou ação de natureza social, como ilustrado na Figura 1:

Figura 1. Ciclo de vida das políticas públicas com IA



Fonte: González, Ortiz e Sánchez Ávalos, 2020.

O ciclo de vida das políticas públicas é uma ferramenta simplificada que busca representar a forma em que as políticas devem ser desenvolvidas e serve para planejar e analisar as diferentes fases do processo. O ciclo começa com a identificação e definição de um problema ou assunto a ser resolvido para, em seguida, formular os diferentes planos de ação. Nessa etapa, o governo deve avaliar as alternativas propostas, inclusive a de manter o status quo, ou seja, abster-se de adotar qualquer medida.

Ao longo desse processo, a tecnologia atua como um instrumento que pode ser útil para desenvolver as alternativas de política pública propostas com base no contexto e na viabilidade da solução. Assim que o governo escolher a alternativa e o instrumento para desenvolvê-la (por exemplo, por meio de um sistema de suporte à decisão ou tomada de decisão baseado em IA), iniciar-se-á a fase de implementação. Por último, na etapa final, busca-se avaliar a eficácia da política em termos de seus objetivos, resultados e impactos previstos (Giorgi, 2017).

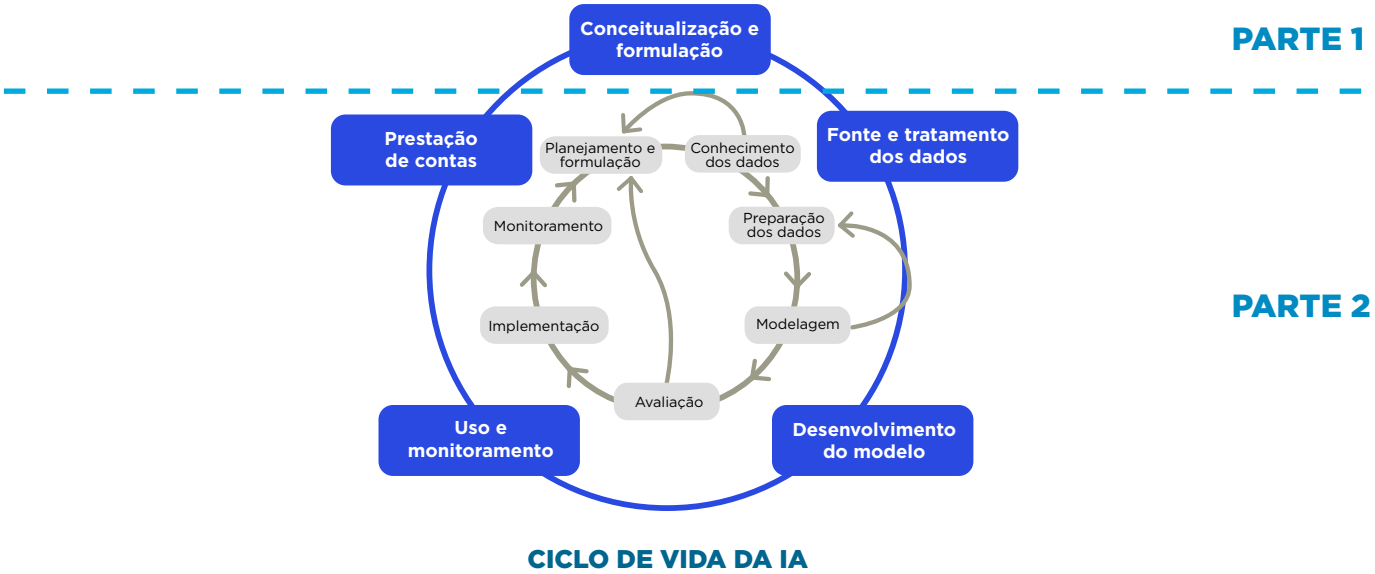
O ciclo de vida da IA inicia-se no momento em que ela é identificada como o instrumento adequado a ser incorporado à política pública para responder ao problema selecionado. As etapas do ciclo de vida da IA são:

1. **Conceitualização e formulação:** compreende os pontos-chave que precisam estar claros para os tomadores de decisão antes de iniciarem um projeto.
2. **Coleta e processamento de dados:** abrange o processo de limpeza e processamento de dados, bem como a identificação de deficiências e vieses que podem comprometer o desenvolvimento do modelo.
3. **Desenvolvimento do modelo e validação:** abrange os principais conceitos a serem seguidos para gerar sistemas de IA robustos e validados.
4. **Uso e monitoramento:** abrange a avaliação da ferramenta após o início da implementação.
5. **Prestação de contas:** supre a necessidade de fornecer informações e transparência para promover a compreensão da IA por parte dos cidadãos.

Na inter-relação entre esses dois ciclos, gera-se um conjunto importante de desafios ao desenvolvimento de uma IA robusta e responsável que devem ser avaliados e considerados durante o desenvolvimento e uso desses sistemas. Além disso, existem desafios transversais, inclusive a transparência, a prestação de contas e a governança, segurança e proteção dos dados pessoais. Por último, há desafios relacionados à elaboração da política pública e à definição da intervenção, bem como à aplicação de critérios de necessidade e proporcionalidade no uso da IA, que se apresentam ao longo do ciclo de vida da IA.

O papel da pessoa responsável pela tomada de decisão referente à política pública em questão (diretor do projeto) não é desenvolver a ferramenta baseada em IA, mas formular o projeto, comunicar-se com a equipe técnica, monitorar a mitigação de riscos, determinar a viabilidade da ferramenta e prestar contas pela execução.

Figura 2. Ciclo de vida da IA conforme a estrutura deste manual



Fonte: Modificado de González, Ortiz e Sánchez Ávalos, 2020.

Na medida em que afetam a gestão da instituição e de seus usuários, os projetos de IA precisam de uma pessoa responsável pela tomada de decisão e uma equipe multidisciplinar de profissionais responsáveis pela execução. Vale destacar que, nas diferentes etapas do ciclo de vida de cada projeto, serão exigidos conhecimentos de diversas áreas que podem ou não entrar no domínio da equipe responsável por sua implementação. Nessas circunstâncias, será necessário consultar especialistas dentro da instituição ou externamente.

As duas funções principais na formulação e execução do projeto podem ser descritas da seguinte forma:

- O responsável pela tomada de decisão ou diretor do projeto é responsável pela formulação da política pública e conduzirá o projeto em sua concepção e posterior implementação.
- A equipe técnica ficará encarregada de analisar as fontes de dados, desenvolver os modelos de IA e monitorar seu uso.

Para acompanhar o desenvolvimento de um sistema de suporte à decisão/tomada de decisão, são propostas ferramentas tanto para o diretor do projeto quanto para a equipe técnica. Esses formulários (Anexos 1 a 5) devem ser preenchidos simultaneamente e fazem parte de um processo de feedback durante todo o ciclo de desenvolvimento.

As ferramentas voltadas para o responsável pela tomada de decisão referente à política pública são:

- **Ficha de formulação e viabilidade do projeto:** esta ferramenta busca identificar os principais aspectos-chave de um projeto de IA para avaliar sua viabilidade, determinar se a IA é a solução correta para o problema e coletar as informações necessárias para a formulação, as quais devem ser compartilhadas com a equipe técnica e com toda a equipe multidisciplinar responsável pelo projeto (Anexo 1).

- **Matriz de maturidade dos dados:** permite uma estimativa inicial da qualidade e relevância dos dados que serão utilizados (Anexo 2).
- **Lista de verificação para o diretor do projeto:** ferramenta que reúne os principais receios por dimensão dos riscos do ciclo de vida da IA do ponto de vista do pessoa responsável pela tomada de decisão (Anexo 3).

As ferramentas para a equipe técnica são:

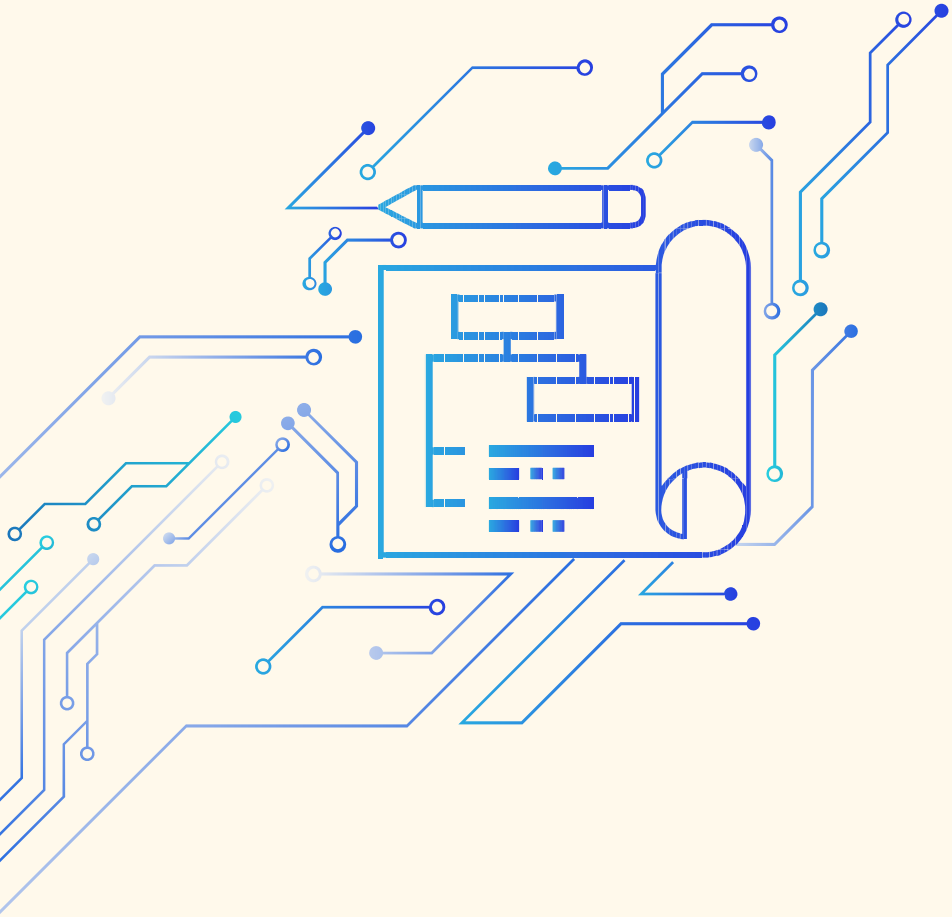
- **Perfil de dados:** contém os principais achados dos bancos de dados a serem utilizados com a ferramenta. Baseia-se na análise exploratória dos dados por parte da equipe técnica[7] (Anexo 4).
- **Perfil do modelo:** descrição final de um modelo de IA que satisfaça os requisitos da política pública e possa ser executado de acordo com os dados disponíveis[8] (Anexo 5).
- **Lista de verificação para a equipe técnica:** ferramenta que reúne os principais receios por dimensão dos riscos do ciclo de vida da IA do ponto de vista da equipe técnica.[9]

Como indicado no início, este manual foca-se principalmente nas ferramentas supramencionadas correspondentes ao diretor do projeto, a quem cabe a tomada de decisão referente à política pública. No entanto, ao longo do documento, também serão mencionadas as ferramentas da equipe técnica, na medida em que o diretor deverá supervisionar seu preenchimento. Para isso, sugere-se que a equipe técnica se baseie no manual "IA responsável: ciclo de vida da inteligência artificial" de González, Ortiz e Sánchez Ávalos (2020). Esse manual descreve as ferramentas técnicas mais detalhadamente, os principais riscos que surgem durante o desenvolvimento de um modelo e as medidas para mitigá-los.

No decorrer deste manual, o caso DART[10] (uma solução de detecção precoce de retinopatia diabética baseada em IA) será usado para exemplificar o que diz respeito tanto à conceitualização e formulação do projeto quanto à sua execução.

7 Parte do Manual Técnico - Ciclo de Vida da Inteligência Artificial.
8 Parte do Manual Técnico - Ciclo de Vida da Inteligência Artificial.
9 Parte do Manual Técnico - Ciclo de Vida da Inteligência Artificial.
10 DART (Chile): https://www.teledx.org/dart/?lang=es

PARTE 1. CONCEITUALIZAÇÃO E FORMULAÇÃO



PARTE 1. CONCEITUALIZAÇÃO E FORMULAÇÃO

A primeira etapa do ciclo de vida da IA é a conceitualização e formulação do projeto, que ficará a cargo de seu diretor, ou seja, a pessoa responsável por tomar as decisões e direcionar os esforços, que serão realizados em colaboração com uma equipe multidisciplinar da instituição. A liderança recai sobre essa pessoa em particular, pois ela possui o conhecimento especializado sobre o assunto a ser debatido e a visão da política pública que se pretende implementar. Também lhe cabe dirigir a execução do projeto a cargo da equipe técnica e aplicar a ferramenta à população destinatária.

Uma conceitualização e uma formulação adequadas do projeto garantirão sua viabilidade, sustentabilidade e valor público, além de mitigar os riscos envolvidos na aplicação da ferramenta de IA.

No processo de conceitualização e formulação, podem ser identificadas etapas-chave e perguntas essenciais às quais o responsável pela tomada de decisão deve responder antes de executar um projeto de IA.

1. **Definição do problema:** a primeira etapa de qualquer projeto é definir claramente o problema de política pública ao qual se deseja responder com a implementação de uma ferramenta de suporte à decisão e/ou tomada de decisão baseada em IA.

2. **Análise de pré-viabilidade:** depois de definir o problema e determinar se a IA é a ferramenta certa para apoiar a solução, e antes de dar continuidade ao projeto, há outras questões-chave que precisam ser respondidas (veja o Item 1.2 abaixo). O objetivo desta etapa é garantir a viabilidade do projeto a fim de não desperdiçar os recursos limitados da instituição.

3. **Definição dos objetivos:** uma vez declarada a viabilidade do projeto, é hora de definir os objetivos e suas métricas ou indicadores, que servirão para medir seu progresso. Essas métricas devem refletir o impacto esperado da aplicação da ferramenta à população destinatária. O atingimento dos objetivos deve ajudar a solucionar o problema identificado.

4. **Descrição das ações:** as ações são as atividades realizadas pela instituição pública que materializarão a resposta ao problema na forma de uma política pública. Podem fazer parte de programas de política pública que visam a tratar questões específicas ou de processos regulares da instituição, como contratações, pagamentos ou atendimento ao público. Embora essas ações geralmente existam independentemente do sistema de IA, esta ajudará a transformá-las para atingir os objetivos do projeto.

5. **Mapeamento de dados:** deve-se pesquisar se existem dados necessários e suficientes para a realização do projeto, se a instituição tem acesso aos bancos de dados ou se serão necessários convênios para obtê-los. Um projeto de IA pode ser baseado em dados tanto internos quanto externos, públicos ou privados. Cabe ressaltar que, na etapa de formulação e conceitualização, não é necessário realizar uma análise detalhada dos dados; isso será feito pela equipe técnica assim que se decidir dar continuidade ao projeto.

6. **Definição da análise e das ferramentas pertinentes:** nesta etapa, o diretor do projeto deve identificar preliminarmente que tipo de análise será necessária para solucionar o problema. O tipo de análise (ou a ferramenta a ser implementada) dependerá de sua natureza e ajudará a aprimorar os processos de atendimento ou resposta necessários. Nesta etapa, busca-se criar

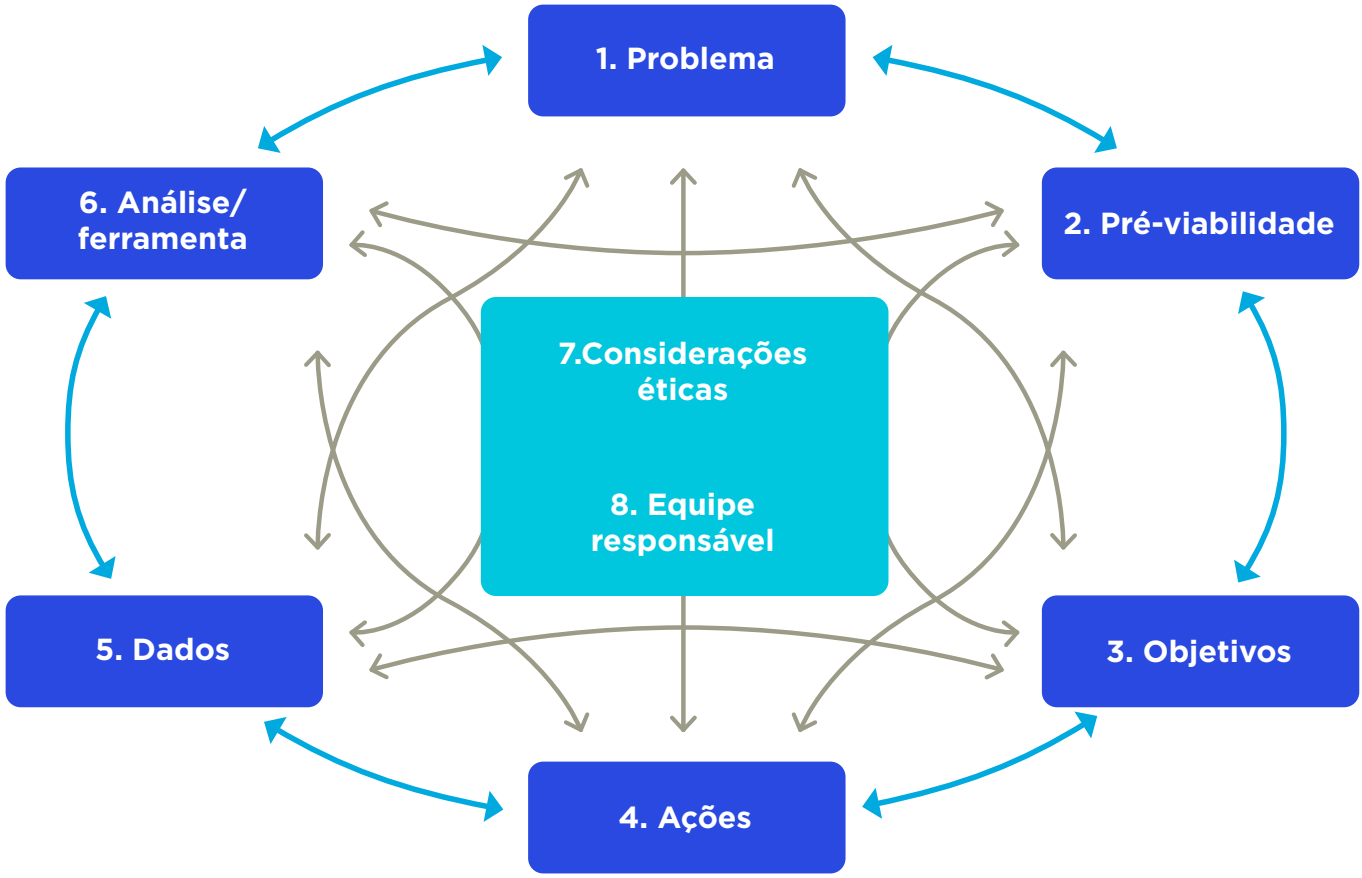
uma aproximação inicial que, posteriormente, deverá ser acordada com a equipe técnica do projeto.

7. **Considerações éticas, jurídicas e de governança:** antes mesmo de iniciar a execução do projeto, os desafios éticos e jurídicos que podem surgir durante a implementação devem estar claros para o diretor. Isso permitirá antecipar situações que possam comprometer o projeto e adotar as medidas de mitigação adequadas.

8. **Formação da equipe responsável:** na condição de diretor do projeto, a pessoa responsável pela tomada de decisão deverá formar a equipe que ficará encarregada de realizá-lo. Os projetos de IA não envolvem apenas o tomador de decisão e a equipe técnica; é necessária a participação de diversas áreas da instituição (por exemplo, a equipe jurídica) e até mesmo de instituições externas (por exemplo, instituições que possuam bancos de dados úteis para o projeto).

O processo de conceitualização e formulação de um projeto deve ser iterativo. Embora a ideia seja começar com uma definição sólida do escopo do problema, ela pode mudar, se, por exemplo, a instituição não tiver a capacidade necessária para adotar as devidas medidas, ou pode precisar ser repensada, se os dados necessários estiverem indisponíveis ou forem insuficientes.

Figura 3. Processo de conceitualização e formulação



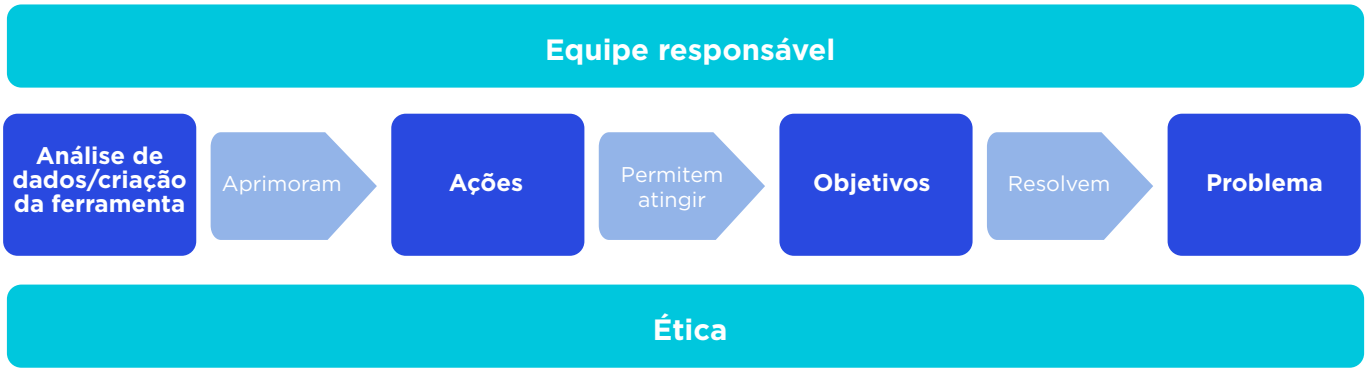
Fonte: elaboração própria.

Uma vez formulado o projeto, deve ser possível visualizar uma relação linear entre as etapas conforme mostrado a seguir (Figura 4):

1. A análise ou ferramenta dependerá dos dados disponíveis.
2. O resultado da análise ou da ferramenta deve aprimorar as ações da instituição.
3. O aprimoramento das ações permitirá atingir o objetivo definido e o estado desejado da situação.
4. Atingir o objetivo ajudará a solucionar o problema.

Aa considerações éticas acompanham todo o pro

Figura 4. Relação entre as etapas de conceitualização e projeção



Fonte: elaboração própria.

A seguir, serão detalhadas as fases essenciais da etapa de conceitualização e projeção. O Anexo 1 deste manual apresenta a ficha de formulação e viabilidade, que deve ser preenchida pelo diretor do projeto e na qual se baseiam os passos descritos aqui. Cada uma dessas fases será ilustrada com o exemplo do projeto DART executado no Chile, referente ao uso da IA na prevenção da cegueira, preenchendo-se as subseções correspondentes na ficha de formulação e viabilidade de um projeto.

Teoria

1.1 Definição do problema

O primeiro passo da conceitualização e formulação de um projeto que envolve IA é definir claramente o problema de política pública que uma instituição busca resolver; por esse motivo, esse passo também faz parte do ciclo de vida de uma política pública. A pessoa responsável por dirigir o projeto e tomar as decisões deve contar com conhecimentos especializados sobre o assunto a ser abordado e será quem priorizará sua resolução e comunicará o problema à equipe técnica.

As instituições públicas devem lidar diariamente com vários assuntos de natureza diversa, como tempos de resposta, alocação de recursos, distribuição de tarefas, aprimoramento do direcionamento de benefícios sociais, entre outros. Nesta etapa, é importante delimitar ao máximo o problema que se pretende resolver com a implementação da ferramenta de IA. O assunto priorizado pode afetar tanto os usuários ou beneficiários externos à organização quanto os que trabalham dentro dela.

A seguir, serão listadas as motivações mencionadas frequentemente para iniciar um projeto baseado em IA:

“Tenho muitos dados e quero fazer algo com eles”.

“Quero fazer um modelo preditivo”.

“Temos este software e devemos usá-lo”.

“Gostaria de usar a tecnologia X na instituição”.

No entanto, esses motivos não justificam iniciar um projeto baseado em IA. Embora seja verdade que as ferramentas de IA podem ser altamente benéficas em termos de economia de tempo e aumento da eficiência dentro da instituição, elas também são caras, considerando-se os recursos e o tempo necessários para desenvolvê-las. Nesse sentido, é fundamental identificar claramente um problema prioritário que a entidade tenha se comprometido a solucionar e para o qual uma ferramenta baseada em IA represente um valor agregado.

Uma vez definido o problema, é extremamente importante ter clareza sobre o estado atual de resposta da instituição, identificar suas limitações e áreas de oportunidade, e determinar como o sistema de suporte à decisão ou tomada de decisão baseado em IA pode melhorar o status quo. Uma boa prática recomendada é conduzir um estudo comparativo de experiências de implementação de ferramentas semelhantes em outras instituições ou outros países. Assim, serão obtidas informações sobre os desafios que tiveram que ser enfrentados, o que também contribuirá para determinar a viabilidade do projeto.

No âmbito da definição do problema, é importante que a equipe quantifique com precisão a quantidade de pessoas afetadas e o orçamento que a solução exigirá. Isso ajudará a dimensionar a magnitude e a gravidade do problema, e, portanto, a necessidade de solucioná-lo.

Por fim, deve haver vontade política de agir e solucionar o problema. As entidades, principalmente os órgãos públicos, enfrentam uma multiplicidade de problemas de diferentes naturezas e, portanto, devem priorizar o assunto a ser tratado para que o projeto obtenha o apoio institucional necessário.

O tema a ser abordado deve ser definido de forma clara e concisa, considerando-se todos os fatores mencionados acima. O diretor do projeto deve levar em consideração que não se trata apenas de comunicar o problema à equipe técnica, mas também à população destinatária. Por esse motivo, recomenda-se usar uma descrição simples que transmita sua importância, mantendo-se o foco da definição no problema em si e em como ele afeta as pessoas e as organizações, conforme os exemplos a seguir.

Exemplos de definições de problemas

Nome do projeto	Definição do problema
Desenvolvimento de modelos preditivos de risco para apoiar as decisões da linha telefónica de suporte para casos de maus-tratos infantis no condado de Allegheny (Pensilvânia)[11]	Para combater o problema de maus-tratos infantis, o Condado de Allegheny oferece duas linhas de emergência através das quais esses abusos podem ser denunciados: uma linha de emergência do condado e uma linha de emergência especializada, a *Childline*. As denúncias podem ser classificadas como CPS (serviços de proteção às crianças) e GPS (serviços de proteção geral). Assim que uma investigação exaustiva é ordenada, o lar em questão entra no sistema. De todas as denúncias classificadas como GPS, 48% entraram no sistema entre abril de 2010 e maio de 2016. O problema é que 53% das denúncias que não geraram uma intervenção se repetiram em um prazo de dois anos. Esse indicador mostra que o sistema de tomada de decisão atual, tal como está, deixa de fora crianças que precisam de apoio do condado. Dado o elevado fluxo de ligações e a assimetria de informações entre as pessoas que realizam as denúncias e os operadores das linhas de emergência, estes últimos precisam tomar decisões em pouco tempo e não conseguem levar em consideração todas as informações disponíveis.
Detecção de risco de suicídio em aplicativos de chat	No Chile, o suicídio é a principal causa de morte entre jovens de 15 a 24 anos. Segundo dados de 2015, a cada 2,8 dias, um menor de idade comete suicídio. A Fundação Todo Mejora[12] é uma ONG cujo objetivo é promover o bem-estar de crianças e adolescentes que sofrem bullying e apresentam comportamentos suicidas em decorrência da discriminação com base em sua orientação sexual, identidade e expressão de gênero. A Todo Mejora oferece um serviço de ajuda remota denominado “Hora segura”, iniciado por meio do serviço de mensagens do Facebook e cujos objetivos são prevenir o suicídio, orientar e, em situações específicas, encaminhar casos para instâncias mais avançadas de resolução. O problema atual é que o número de usuários que precisam de ajuda às vezes ultrapassa a quantidade de voluntários disponíveis para essa tarefa, não sendo possível priorizar o contato com crianças e adolescentes com um alto risco de suicídio. Concretamente, a Todo Mejora conta com uma equipe de 60 a 80 profissionais, que devem atender entre 8.000 e 9.000 pedidos de atendimento telefónico por ano. Anualmente, cerca de 60% das consultas são recebidas de adolescentes da comunidade LGBTQIA+ e 40% de adolescentes cisgênero heterossexuais. 68% das pessoas que procuram a “Hora segura” haviam apresentado comportamento suicida nos últimos dois meses.

11 Allegheny Family Screening Tool (EUA): https://www.alleghenycounty.us/

12 Fundación Todo Mejora (Chile): https://todomejora.org/

"Existe o risco adicional de desenvolver os projetos de IA com base na tecnologia, não no problema social em questão. Mesmo um projeto de IA necessário e funcional pode apresentar riscos se não for desenvolvida uma ação de política pública correta" (Cabrol, et. al., 2020).

Erros comuns na definição do problema

Embora a fase de identificação do problema pareça ser muito simples, talvez seja a mais crucial do processo de conceitualização e onde mais erros são cometidos, como os seguintes:

- Incluir descrições e conceitos excessivamente amplos e vagos (como "melhorar a tomada de decisão", "ser mais eficiente" etc.), em vez de manter o foco em um problema específico da instituição.
- Definir o problema como a ausência de uma ferramenta/modelo. Por exemplo: "o problema da minha instituição é que não possuímos uma ferramenta automática para direcionar a prestação de ajuda".
- Não especificar o estado atual e/ou as lacunas que precisam ser preenchidas.
- Não quantificar o número de pessoas afetadas e/ou o orçamento necessário para solucionar o problema.

Não se esqueça de...

1. **Definir** claramente o problema de política pública que se busca resolver, identificando e quantificando os grupos de pessoas afetadas, e determinando seu impacto no orçamento. A definição do problema deve ser facilmente compreendida por uma pessoa de fora da instituição.

2. **Entrar** em contato com as pessoas da sua instituição encarregadas de abordar o problema para entender como isso está sendo feito atualmente. Que informações elas podem fornecer sobre como aprimorar o sistema de resposta?

3. **Pesquisar** como outros órgãos (nacionais ou estrangeiros) com um problema semelhante têm implementado soluções baseadas em IA. O ideal seria contatá-los para saber mais sobre os desafios e dificuldades com os quais se depararam ao longo do caminho.

4. **Conversar** com a diretoria da sua instituição sobre a prioridade de solucionar o problema e obter adesão ao projeto no mais alto escalão. Se você fizer parte da diretoria, documente como essa prioridade é refletida nos planos estratégicos da instituição.

Atividade Uma vez finalizada esta seção, preencha a parte correspondente à **definição do problema** na Ficha de formulação e viabilidade de projetos.

VER ARQUIVO NO ANEXO 1

FICHA DE FORMULAÇÃO E VIABILIDADE DE PROJETOS

1 Definição do problema

Qual é o problema a ser solucionado?

Descreva a(s) população(ões) afetada(s) pelo problema (pessoas, grupos, entidades etc.)

Quantas pessoas/organizações/locais/etc. são afetados e em que medida?

Por que resolver esse problema é uma prioridade para a sua organização?

Você conhece algum caso de uso de IA semelhante que tenha sido implementado antes? Qual?

FICHA DISEÑO Y FACTIBILIDAD DE PROYECTO

EXEMPLO DART: Uuso da inteligência artificial na prevenção da cegueira

1 Definição do problema

Qual é o problema a ser solucionado?
O diabetes é uma doença crônica que ocorre quando o pâncreas não é capaz de produzir insulina suficiente ou quando o organismo não utiliza toda a insulina que produz. Existem dois tipos de diabetes: o diabetes tipo 1 e o diabetes tipo 2. O primeiro surge como consequência de um ataque do próprio sistema imunológico, não pode ser prevenido e não tem cura, enquanto o segundo é o resultado de estilos de vida que conduzem ao sedentarismo e à obesidade. Neste caso, o pâncreas continua a produzir insulina, embora em quantidades insuficientes, e em alguns casos pode ser curado. Ambos os tipos de diabetes podem causar cegueira, insuficiência renal, acidente vascular cerebral (AVC) e até levar à amputação das pernas.

A retinopatia diabética (RD) é a doença ocular mais comum entre os diabéticos no mundo todo. O diagnóstico da doença requer um exame de fundo de olho, no qual um tecnólogo médico capta uma imagem detalhada do globo ocular para análise posterior por um especialista.

Segundo as estimativas, no Chile, há um déficit anual de 39.168 horas de oftalmologistas, o que indica uma escassez do capital humano necessário para analisar todos os exames de fundo de olho (Hojman, 2014).

Descreva a(s) população(ões) afetada(s) pelo problema (pessoas, grupos, entidades etc.)
Esse problema afeta basicamente três grupos:

1. Pessoas diagnosticadas com diabetes que não podem realizar o exame de fundo de olho uma vez por ano, de acordo com as recomendações internacionais.
2. O Ministério da Saúde, que precisa gastar recursos adicionais com a retinopatia diabética por não conseguir fazer uma prevenção adequada.
3. Os oftalmologistas, cuja carga de trabalho é excessiva.

Quantas pessoas/organizações/locais/etc. são afetados e em que medida?
Em 2014, havia 422 milhões de pessoas com diabetes no mundo. Segundo estimativas da Organização Mundial da Saúde, em 2040, esse número terá subido para 600 milhões. No Chile, estima-se que uma em cada 10 pessoas tenha diabetes e que entre 15% e 20% desses pacientes apresentem algum grau de retinopatia diabética. Segundo a Pesquisa Nacional de Saúde de 2010, apenas 34,8% dos pacientes foram examinados por um oftalmologista.

Por que resolver esse problema é uma prioridade para a sua organização?
A Teledx foi criada com o único objetivo de apoiar tecnologicamente a tarefa oftalmológica do sistema de saúde e, assim, reduzir os índices de perda de visão e cegueira na população adulta. O foco principal é a retinopatia diabética, pois há uma necessidade urgente de aumentar a cobertura de diagnóstico.

A Estratégia Nacional de Saúde, elaborada pelo Ministério da Saúde para o decênio 2011-2020, estabelece diversos objetivos, entre os quais “Aumentar a proporção de pessoas com diabetes controlada”. A estratégia específica para atingir este objetivo consiste em melhorar o controle dos pacientes diabéticos por meio de uma cobertura mais ampla de diferentes exames clínicos, inclusive o exame de fundo de olho.

Você conhece algum caso de uso de IA semelhante que tenha sido implementado antes? Qual?
Sim, foi realizada uma revisão da literatura de outros estudos que buscaram detectar a retinopatia diabética (Arenas, 2012).

Teoria 1.2 Análise de pré-viabilidade

Depois de definido o problema, é necessário questionar a viabilidade da execução do projeto, além de seus aspectos técnicos a cargo da equipe de especialistas. Esse passo é fundamental antes de iniciar um projeto, pois evita um desperdício dos recursos financeiros e humanos limitados da instituição. Algumas das perguntas a serem respondidas nesse passo são:

1. **Está dentro da alçada da entidade agir sobre o problema? Será necessário firmar parcerias com outros órgãos públicos? Os recursos humanos e financeiros necessários para realizar o projeto estão disponíveis?**
   Este ponto refere-se às competências tanto jurídica quanto regulatória para responder ao problema e à disponibilidade de recursos humanos (internos) e financeiros que podem ser alocados para criar a ferramenta. A instituição precisa poder responder ao problema dentro dos âmbitos jurídico e orçamentário.

2. **Há dados relevantes (suficientes para mudar a forma atual de responder ao problema)? É possível ter acesso a esses dados?** Os dados devem ser suficientes para poder criar uma ferramenta de IA que ajude a resolver o problema. Por exemplo, se a questão a ser solucionada é que a autoridade competente não consegue prever eventos ambientais de má qualidade do ar, o que se traduz em um aumento das consultas médicas devido a doenças respiratórias, é necessário, pelo menos, dispor de informações históricas sobre a presença de partículas poluentes na atmosfera nos últimos anos. Se os dados estiverem agregados por meses, é pouco provável que um modelo preditivo possa ser desenvolvido.

3. **Quais são os riscos do projeto (éticos, licença social, implementação etc.)?** Um projeto de um órgão público baseado em IA buscará gerar valor para a população. No entanto, como as ferramentas de IA e sua execução apresentam alguns riscos, deve ser feita uma análise que compare o efeito positivo esperado com os riscos em potencial das ferramentas de IA (posteriormente, esses temas serão tratados com maior profundidade). Alguns desses riscos são:

   Riscos éticos: os sistemas de resposta automatizados podem reproduzir os vieses presentes nos dados ou podem estar enviesados em decorrência de um modelo inadequado. Como há uma probabilidade de a ferramenta apresentar algum desses riscos, cabe ao diretor do projeto adotar as medidas necessárias para mitigá-los.

   Licença social: refere-se à aprovação concedida pela população destinatária para usar a ferramenta. Ou seja, se essa população conhecesse o projeto, ela aprovaria o uso da IA? Para obter a licença social, é necessário, primeiro, que o benefício a ser gerado pela aplicação da ferramenta esteja claro para a população e, segundo, que haja transparência sobre a proteção dos dados, o modo de utilização da ferramenta e as medidas que serão adotadas para mitigar vieses, caso sejam detectados.

   Riscos de implementação: refere-se aos riscos que podem surgir a partir do início da execução por parte da instituição. O primeiro risco é de o modelo não funcionar suficientemente bem e, portanto, não resultar em um aprimoramento do trabalho institucional, tanto internamente quanto perante o público; o segundo refere-se a ele

nunca ser implementado devido a falhas em um dos componentes-chave (adesão da autoridade, falta de recursos e/ou resistência às mudanças por parte das pessoas que usarão a ferramenta); e o terceiro é o risco de ser implementado, mas nunca usado.

Responder a essas perguntas permite identificar alguns aspetos fundamentais que conduzirão a uma formulação e uma execução bem-sucedidas do projeto, uma vez que esse exercício permite uma melhor compreensão de suas limitações e riscos.

A ideia é que, desde o início, a necessidade do projeto seja plenamente justificada, as capacidades de projetá-lo e executá-lo com eficácia estejam disponíveis, e tanto os riscos que o projeto pode apresentar quanto as medidas para mitigá-los sejam identificados.

Não se esqueça de...

1. **Obter** a adesão das pessoas que executarão a ferramenta e incorporá-las à equipe de trabalho.
2. **Identificar** os possíveis riscos éticos e jurídicos do modelo. Conversar com a equipe jurídica e incorporá-la à equipe do projeto.
3. **Assegurar** os recursos financeiros e humanos que permitam sustentar a ferramenta com o passar do tempo.

Actividade

Uma vez finalizada esta seção, preencha a parte correspondente à **análise de pré-viabilidade** na Ficha de formulação e viabilidade de projetos.

VER ARQUIVO NO ANEXO 1

FICHA DE FORMULAÇÃO E VIABILIDADE DE PROJETOS

2 Análise de pré-viabilidade

Está dentro da alçada da entidade agir sobre o problema? Será necessário firmar parcerias com outros órgãos públicos? Os recursos humanos e financeiros necessários para realizar o projeto estão disponíveis?

Há dados relevantes (suficientes para mudar a forma atual de responder ao problema)? É possível ter acesso a esses dados?

Quais são os riscos do projeto (éticos, licença social, implementação etc.)?

FICHA DISEÑO Y FACTIBILIDAD DE PROYECTO

EXEMPLO DART: Uso da inteligência artificial na prevenção da cegueira

2 Análise de pré-viabilidade

Está dentro da alçada da entidade agir sobre o problema? Será necessário firmar parcerias com outros órgãos públicos? Os recursos humanos e financeiros necessários para realizar o projeto estão disponíveis?
Requer parcerias com outros órgãos, principalmente com o Ministério da Saúde, para que a ferramenta possa ser implementada nos centros de saúde. Há recursos humanos disponíveis, mas não financeiros, sendo necessário, portanto, candidatarmo-nos para ter acesso a algum tipo de financiamento.

Há dados relevantes (suficientes para mudar a forma atual de responder ao problema)? É possível ter acesso a esses dados?
Sim, mas um número maior de centros de saúde deve ser contatado para solicitar mais imagens de exames e, assim, aumentar o tamanho da amostra (N) disponível.

Quais são os riscos do projeto (éticos, licença social, implementação etc.)?
Os resultados dos exames de fundo de olho são dados pessoais e, por isso, os pacientes precisam autorizar sua realização. Um dos principais riscos do projeto decorre da implementação, pois é necessário obter a adesão dos centros de saúde, tecnólogos médicos e oftalmologistas que interagirão com o modelo, para que usem a ferramenta de forma eficaz. Como o projeto busca reduzir o tempo de espera, o que também reduzirá o risco de cegueira, é muito provável que a população em geral conceda a licença social.

Teoria

1.3 Definição dos objetivos

Os objetivos serão definidos como o estado que se pretende atingir em um determinado assunto. Por exemplo, se, em um determinado momento, um órgão público demora, em média, cinco dias para responder às solicitações dos cidadãos, um dos objetivos do projeto baseado em IA pode ser reduzir o tempo de resposta para três dias. Assim, os objetivos expressam avanços concretos de uma situação de deficiência para uma de satisfação. Portanto, todo indicador declarado nos objetivos também deve ser descrito na definição do problema.

Uma boa metodologia para definir objetivos é a conhecida pela sigla em inglês **SMART**, segundo a qual os objetivos precisam ser e**S**pecíficos, **M**ensuráveis, **A**tingíveis, **R**ealistas e com um **T**empo definido.

Como indicado acima, um dos erros mais comuns é definir como objetivo "criar um sistema automático de resposta" ou "construir um modelo preditivo", pois nenhum deles realmente responde à pergunta central desta etapa: "Atingir o objetivo ajuda a solucionar o problema?" A resposta nesses dois casos seria "não", pois o simples fato de contar com um sistema preditivo não significa que o problema em questão será solucionado. Em outras palavras, a eficácia da ferramenta ou do modelo dependerá não apenas de sua precisão, mas também de como será aplicado a seus beneficiários em potencial. As atividades auxiliares do projeto, como "explorar os dados" ou "entender o comportamento histórico dos usuários do programa", também não constituem objetivos do projeto.

Na hora de definir os objetivos, uma boa prática é perguntar-se "para quê": um modelo preditivo é necessário para quê? Implementar um chatbot no site é necessário para quê? Conhecer o comportamento dos usuários é necessário para quê? Ao repetir constantemente esse "para quê", o objetivo real do projeto será finalmente atingido. Na tabela sinóptica abaixo, serão apresentados três exemplos de projetos com uma breve descrição do problema, um objetivo inicial e um objetivo aprimorado segundo os parâmetros abordados nesta seção.

Caso	Problema	Objetivo inicial	Fraqueza	Objetivo aprimorado
Prevenção do envenenamento por chumbo.[13]	X crianças têm um certo nível de chumbo no sangue. Não há níveis aceitáveis de chumbo no sangue. A principal fonte de intoxicação por chumbo em crianças está nas paredes de casas antigas, cobertas com pintura que contém chumbo. O chumbo é detectado por meio de um exame de sangue realizado quando as crianças ingressam no sistema educacional. Quando o chumbo é detectado, um inspetor é enviado à casa da criança, sendo iniciadas ações para remover esse metal pesado, para as quais há subsídios do estado disponíveis. No entanto, não há recursos suficientes para consertar todas as casas cobertas com pintura com chumbo nas cidades.	Prevenir o envenenamento por chumbo em crianças.	O objetivo é muito amplo. Embora ajude a resolver o problema (envenenamento por chumbo em crianças), não é específico nem está relacionado a ações concretas que a cidade pode tomar. Também não é oferecida nenhuma métrica de sucesso; portanto, não é possível saber se foi atingido ou não.	Para o ano de 2019, espera-se que X% do total de inspeções detecte chumbo nas casas habitadas por mulheres gestantes.
Detecção da retinopatia diabética (RD).[14]	Atualmente, há um atraso de X meses na entrega dos resultados dos exames de fundo de olho (para estabelecer se um paciente sofre ou não de RD). Uma retinopatia diabética não diagnosticada, ou com um diagnostico tardio, pode causar cegueira total. Atualmente, não há capacidade humana (oftalmologistas e tecnólogos médicos) necessária para cumprir um prazo prudente de entrega de resultados dos exames de fundo de olho.	Analisar os exames de fundo de olho de todos os pacientes do sistema de saúde.	O objetivo não se refere ao problema, que é a demora na análise dos exames de fundo de olho. Isso aumentaria as taxas de retinopatia diabética, que, sem tratamento, aumentaria os casos de cegueira.	Reduzir, dentro de até dois anos, o tempo de entrega dos resultados a pacientes com alta probabilidade de sofrer de retinopatia diabética em X%.
Fiscalização de importações.	Estima-se que, atualmente, X% das importações que entram no país sejam fraudulentas e não sejam submetidas ao controle aduaneiro.	Desenvolver um sistema preditivo de importações fraudulentas.	A solução técnica não é um objetivo.	Aumentar a porcentagem de apreensões de importações fraudulentas sobre o total de importações fiscalizadas no ano fiscal seguinte.

13 Lead poisoning prevention (EUA): http://www.datasciencepublicpolicy.org/projects/public-health/poison-prevention/

14 DART (Chile): https://www.teledx.org/dart/?lang=es

Embora a concretização dos objetivos conduza à resolução dos problemas identificados e permita atingir o estado desejado, não se deve esquecer que os projetos sofrem de limitações que os podem afetar e colocar em risco os objetivos estabelecidos. Entre as limitações típicas, estão a falta de recursos humanos e/ou financeiros e a resistência às mudanças, entre outros. Identificar as possíveis limitações permitirá ao diretor do projeto antecipar o surgimento desses obstáculos na etapa de implementação.

Não se esqueça de...

1. **Definir** objetivos conforme o problema de política pública a ser resolvido.

2. **Definir** objetivos **SMART**: e**S**pecíficos, **M**ensuráveis, **A**tingíveis, **R**ealistas e com um **T**empo definido.

Atividade uma vez finalizada esta seção, preencha a parte correspondente aos **objetivos** na Ficha de formulação e viabilidade de projetos.

VER FICHA EM ANEXO 1

FICHA DE FORMULAÇÃO E VIABILIDADE DE PROJETOS

3 Objetivos

	Objetivo	Limitações
1		
2		

FICHA DE FORMULAÇÃO E VIABILIDADE DE PROJETOS

EXEMPLO DART: Uso da inteligência artificial na prevenção da cegueira

3 Objetivos

	Objetivo	Limitações
1	Reduzir em X%, dentro de dois anos, o tempo de entrega dos resultados aos pacientes com alta probabilidade de sofrer de retinopatia diabética.	Há um número limitado de oftalmologistas para analisar os exames de fundo de olho.
2	Reduzir os gastos públicos associados ao tratamento da retinopatia diabética em 50% dentro de 10 anos.	As unidades de atenção básica à saúde devem atingir as metas de cobertura do exame anual de fundo de olho para pacientes diabéticos, a fim de detectar precocemente e prevenir a retinopatia.

Teoria

1.4 Descrição das ações

Após definir os objetivos do projeto, cabe perguntar: o que a instituição está fazendo atualmente para solucionar o problema? Como sua intervenção pode ser aprimorada com a implementação da ferramenta de IA? É nesta etapa que devem ser claramente identificados os processos atuais de atendimento ou resposta da instituição em relação ao problema detectado e que podem ser aprimorados com a implementação de análises ou de ferramentas baseadas em IA. Algumas ações das instituições são:

- **Implementação de produtos ou programas estratégicos:** concessão de subsídios de moradia, alocação de subsídios de alimentação para estudantes, planejamento de fiscalizações de importações etc.
- **Processos de suporte da instituição:** atendimento aos usuários, contratação de pessoal, planejamento de rotas de transporte etc.

Nesta etapa, é importante identificar com precisão as pessoas que estão realizando essas atividades atualmente, pois provavelmente precisarão ser incluídas na equipe do projeto, bem como através de que canais essas atividades estão sendo executadas.

Isso já havia sido delineado na Pergunta 2 da análise de pré-viabilidade, Item 1.2: A organização tem capacidade para adotar medidas para solucionar o problema? Está dentro da alçada da organização? Os recursos humanos e financeiros necessários estão disponíveis? Aqui, espera-se que o diretor do projeto realize uma análise profunda das ações que a instituição está realizando atualmente para resolver o assunto em questão.

A importância de definir corretamente essas ações decorre do fato de que uma instituição executa diversas intervenções e/ou programas, mas nem todos estão relacionados ao problema identificado. **Nesta etapa, devem ser definidas apenas as ações que podem ter um impacto no problema em questão.** Por exemplo, se o que deve ser solucionado é o tempo excessivo decorrido entre o momento em que chega uma solicitação de um cidadão e a resposta da autoridade pública competente, uma ação relevante seria determinar quais são os impedimentos durante a análise do requerimento ou identificar se o sistema de notificações está funcionando com atrasos.

Em outros casos, um mesmo problema pode ser de responsabilidade de diferentes instituições, mas como ele será respondido dependerá da missão de cada uma. Por exemplo, a luta contra a pobreza é um problema muito amplo; portanto, diferentes ministérios estarão envolvidos de várias formas em sua solução. O Ministério do Desenvolvimento, por exemplo, pode conceder um auxílio às famílias mais vulneráveis durante os meses de inverno para ajudar a pagar as despesas que aumentam nessa época do ano (aumento das consultas médicas e aumento dos preços dos alimentos, por exemplo), enquanto o Ministério da Educação pode se concentrar no acesso universal à educação, apelando para a mobilidade social que acarreta um nível mais alto de ensino. Portanto, os sistemas de resposta variarão de acordo com as capacidades e competências de cada instituição.

Outro caso seria um projeto para aumentar a porcentagem de apreensões de importações fraudulentas. Aqui, a ação principal é a fiscalização, o que exige uma explicação clara de como a decisão de quais importações são submetidas ao procedimento é tomada atualmente. Por exemplo, é possível que, a cada dia, seja decidido quais cargas serão fiscalizadas com base no histórico de cumprimento

do importador, ou que os fiscais responsáveis tomem a decisão na hora, sem que exista um planejamento. É importante considerar como o sistema opera atualmente e como a mudança pode afetar as pessoas encarregadas de gerenciá-lo. Neste exemplo, o que aconteceria se os inspetores recebessem um bônus por cada carga fiscalizada? Haveria, então, um incentivo para controlar uma grande quantidade de cargas, independentemente do resultado. Portanto, um modelo preditivo que reduza o número de cargas a serem inspecionadas pode gerar resistência na equipe responsável pela tarefa, pois obrigaria seus integrantes a inspecionarem menos importações em busca de uma maior eficiência, o que, por sua vez, reduziria os bônus. Portanto, para garantir a viabilidade do projeto, é fundamental conhecer profundamente como a instituição aborda o problema atualmente e incluir na equipe do projeto alguém que seja responsável por essas ações (chefe de fiscalizações, por exemplo), para que detecte e saliente possíveis fraquezas.

Um dos erros mais comuns cometidos nesta etapa é manter o foco em tarefas auxiliares do projeto, como coletar informações, explorar os dados ou vincular diferentes bancos de dados. Embora essas atividades devam estar a serviço do projeto, não devem se tornar sua atividade principal. Para evitar esse erro, é necessário retornar ao assunto a ser solucionado e relembrar a cadeia de relacionamento entre as diferentes etapas. Assim, um melhor sistema de resposta (informado pela análise ou ferramenta) deve ajudar a atingir os objetivos, que, por sua vez, contribuirão para a resolução do desafio. A definição das ações, juntamente com o contexto em que são realizadas e as pessoas envolvidas, é fundamental para o sucesso do projeto na etapa de implementação. No entanto, às vezes surgem imprevistos que dificultam o cumprimento dos objetivos e precisam ser eliminados para que não se transformem em obstáculos.

Exemplo: **Nem tudo pode ser previsto**

Na T ailândia, com a ajuda do Google, foi implementado um sistema automático de detecção de retinopatia diabética por meio da análise de fundo de olho. O modelo funcionou muito bem no laboratório, mas, quando foi implementado, os locais onde os exames eram realizados não contavam com uma iluminação suficiente. Assim, as imagens eram de qualidade inferior às usadas para treinar o modelo, e não foi possível fornecer um resultado confiável sobre a presença dessa doença. Exigiu-se, então, que os pacientes se dirigissem a centros de atendimento com melhores condições, o que raramente acontecia devido às dificuldades de deslocamento. Por outro lado, a baixa qualidade da conexão com a internet fez com que a espera pelo resultado fosse muito longa. O problema neste exemplo é que o Google não incluiu a equipe de enfermagem (responsável pela execução da ação) na etapa de conceitualização do projeto; portanto, os enfermeiros não estavam cientes dos problemas no sistema de atendimento, ou poderiam ter incorporado ações que pudessem mitigar os efeitos da mudança na iluminação ou da conexão fraca com a internet antes de o projeto ser implementado (Heaven, 2020).

Caso não exista a capacidade de adotar medidas referentes ao assunto prioritário selecionado, e dado o processo iterativo de formulação de um projeto, pode ser necessário retornar ao problema para uma reformulação completa. Daí a importância da análise de pré-viabilidade, pois evita que essa situação aconteça posteriormente.

Não se esqueça de...

1. **Conversar** com as pessoas atualmente encarregadas do problema a ser resolvido para determinar como as decisões são tomadas e como podem ser afetadas pelas mudanças causadas pela implementação da ferramenta de IA.

2. **Comunicar** de forma eficaz as implicações e os benefícios da implementação de uma ferramenta baseada em IA para obter a adesão das pessoas que a utilizarão.

3. **Incorporar** à equipe do projeto pelo menos uma das pessoas encarregadas de responder ao problema antes de implementar o sistema de IA.

Atividade Uma vez finalizada esta seção, preencha a parte correspondente **às ações** na Ficha de formulação e viabilidade de projetos.

VER FICHA EM ANEXO 1

FICHA DE FORMULAÇÃO E VIABILIDADE DE PROJETOS

4 Descrição das ações

Problema para resolver	Ação 1	Ação 2
Ações		
Quem executa a ação?		
Para quem ou o que a ação está sendo executada?		
Com que frequência é tomada a decisão de realizar a ação?		
Que canais estão sendo utilizados ou podem ser utilizados para realizar a ação?		
Outras informações úteis sobre a ação		

FICHA DISEÑO Y FACTIBILIDAD DE PROYECTO

EJEMPLO DART: Uso da inteligência artificial na prevenção da cegueira

4 Descripción de acciones

Problema para resolver	**Ação 1**	**Ação 2**
Ações	Realização do exame de fundo de olho.	Diagnóstico de retinopatia diabética por parte de um oftalmologista.
Quem executa a ação?	Tecnólogos médicos em unidades de atenção básica à saúde.	Oftalmologista.
Para quem ou o que a ação está sendo executada?	Pessoas diagnosticadas com diabetes.	Pessoas diagnosticadas com diabetes.
Com que frequência é tomada a decisão de realizar a ação?	Anualmente, por pessoa.	Depois de realizado o exame.
Que canais estão sendo utilizados ou podem ser utilizados para realizar a ação?	Unidades de atenção que contam com os suprimentos necessários.	Unidades de atenção.
Outras informações úteis sobre a ação		

Teoria

1.5 Mapeamento de dados

A realização da análise ou da ferramenta de IA depende da existência dos dados que as alimentam. Portanto, antes de iniciar o projeto, o diretor deve explorar a disponibilidade e qualidade dos dados internos e externos que possam ser necessários. Com relação a esses dados, deve-se consultar se é possível obter acesso a eles e qual seria o custo. É possível que algumas ferramentas de IA exijam apenas dados abertos. Se necessário, nesta etapa, o diretor deve consultar as áreas da instituição que coletam ou administram os dados sobre alguns aspectos-chave dos bancos de dados. Não é necessário se aprofundar, pois essa tarefa caberá à equipe técnica na etapa de execução.

Os dados devem ser suficientes para a análise e apresentar a granularidade exigida. Uma ferramenta útil para saber se um banco de dados é maduro (bom) é a matriz de maturidade de dados. No Anexo 2 deste manual, será apresentada uma adaptação do Quadro de Maturidade dos Dados (Data Maturity Framework) da Universidade de Chicago.[15] Esta ferramenta procura ajudar o diretor do projeto a analisar, em cada um dos aspectos para os quais se pretende adotar medidas, o estado de progresso de cada um dos bancos de dados que serão utilizadas na ferramenta de IA. É possível que, neste passo, seja necessário consultar diretamente os responsáveis pelos dados sobre suas características e problemas.

Uma explicação das caraterísticas usadas na matriz será apresentada abaixo.

Acesso: independentemente de um banco de dados ser interno ou externo, a facilidade de acesso será um grande propiciador do projeto de IA.

Armazenamento: as instituições detêm quantidades imensas de dados, mas eles podem ser armazenados de várias maneiras. Embora, idealmente, os dados devam ficar em um banco de dados digital, muitas vezes, o diretor do projeto pode deparar-se com registros em formato físico. Isso exigirá um passo anterior à implementação, no qual os dados serão digitalizados para que possam ser lidos por algum software e utilizados na análise pertinente. Como isso pode atrasar a execução do projeto, é extremamente importante determinar o estado dos dados antes desse passo.

Integração: embora as instituições geralmente detenham uma grande quantidade de dados, muitas vezes, cada banco de dados é tratado de forma isolada e independente dos outros. O cenário ideal é que haja interoperabilidade entre os bancos de dados (internos e externos), de forma que seja mais fácil acessar as informações necessárias e unificar as variáveis ou os rótulos-chave.

Relevância: a relevância dos dados refere-se à sua pertinência ao problema a ser abordado.

Qualidade: refere-se à possibilidade de o banco de dados estar incompleto ou conter erros. Um banco de dados que não contenha os dados de um sujeito ou observações sobre ele (região, bairro ou idade, entre outros) pode ter consequências importantes no modelo resultante, pois pode incluir vieses associados a uma parte da população. Um banco de dados com variáveis ausentes gera grandes desafios, pois a equipe técnica, liderada pelo diretor do projeto, terá que tomar a decisão de eliminar totalmente a observação ou imputar-lhe dados. Com frequência, os bancos de dados também apresentam erros de digitação, que ocorrem no momento da coleta ou transcrição das informações. Aqui, novamente, a equipe técnica deve decidir, junto com o diretor de projeto,

15 http://dsapp.uchicago.edu/resources/datamaturity/

o que fazer com esses erros para não comprometer a qualidade da ferramenta final. O ideal seria que os bancos de dados não apresentassem nenhum dos problemas descritos acima, o que é incomum.

Frequência: refere-se à regularidade com que os dados são coletados, não à regularidade dos dados em si (observações anuais, mensais ou diárias). Os dados podem ser coletados uma vez para uma política pontual, por exemplo, ou podem ser coletados mensalmente, diariamente ou em tempo real, ou seja, à medida que forem surgindo. É mais importante aqui determinar qual é a política de coleta de informações que saber exatamente quantas vezes elas foram coletadas.

Granularidade: refere-se ao nível de detalhamento dos dados. É possível ter observações por país, região, bairro, evento (toda vez que ocorrer) ou indivíduo. Um banco de dados mais maduro terá um nível de granularidade maior, ou seja, terá observações cada vez mais específicas. Entender o nível de granularidade dos bancos de dados é fundamental para o sucesso do projeto. Por exemplo, um modelo preditivo de abandono escolar precisará de dados de cada um dos alunos; não basta ter informações por área geográfica, pois isso não permitiria o devido direcionamento das ações de prevenção, que podem ser realizadas no âmbito individual.

Histórico: os dados de uma observação podem mudar com o passar do tempo. Por exemplo, uma pessoa pode estar solteira aos 25 anos, mas casada 10 anos depois; portanto, na nova coleta de dados, a variável "estado civil" será diferente. Existem várias maneiras de armazenar esses dados históricos: a mais deficiente é limpar o histórico de dados antigos e substituí-lo pelas novas informações, enquanto a mais avançada é salvar os novos dados, registrando sua data e relacionando-os com os anteriores.

Privacidade: além do tratamento de dados pessoais ou sensíveis, o uso, acesso e manipulação desses tipos de informação devem ser regulados por uma política de proteção. Estabelecer uma política de privacidade implica regular quem (funções) tem acesso a quais tipos de dado e para quais tipos de projeto. Isso pode ser feito por meio de um processo ad hoc ou de um software (geralmente, um sistema administrador de bancos de dados) responsável por gerenciar a política de acesso.

Documentação: normalmente, os bancos de dados usam códigos nas variáveis; nesses casos, é imprescindível dispor de um livro de códigos ou dicionário de dados que explique o que significa cada uma dessas variáveis. Da mesma forma, uma documentação adequada deve conter informações sobre os metadados do conjunto de dados: sua descrição, o modo em que foram extraídos e se estão relacionados a outros conjuntos de dados, entre outros aspectos.

Vale destacar que o fato de um banco de dados ser classificado como básico na matriz de maturidade dos dados não significa que seja insuficiente para treinar a ferramenta. No entanto, é necessário levar em consideração que a equipe técnica precisará investir mais tempo e recursos em seu gerenciamento. Por esse motivo, pode ser que o diretor do projeto precise suspender momentaneamente o desenvolvimento da ferramenta de IA para implementar políticas mais avançadas de coleta, armazenamento e tratamento de dados antes de prosseguir com a implementação.

Se os dados necessários para criar a ferramenta não existirem ou se a granularidade for insuficiente, será impossível resolver o problema selecionado por meio do uso da inteligência artificial. Isso exigirá reorientar a solução do problema ou voltar à etapa de definição para revisar detalhadamente a disponibilidade dos dados em outros casos semelhantes que tenham usado a IA.

Não se esqueça de...

1. **Identificar** claramente todos os bancos de dados necessários para realizar o projeto.
2. **Conversar** com os responsáveis sobre as características fundamentais dos bancos de dados para avaliar seu nível de maturidade e se são suficientes para executar o projeto.
3. No caso de solicitar dados de outras instituições, **negociar** o mais rápido possível os acordos necessários para garantir o acesso em tempo hábil.
4. **Incorporar** os responsáveis pelos dados à equipe do projeto.

Atividade Uma vez finalizada esta seção, preencha a parte correspondente ao **mapeamento de dados** na Ficha de formulação e viabilidade de projetos.

VER FICHA EM ANEXO 1

FICHA DE FORMULAÇÃO E VIABILIDADE DE PROJETOS

5 Mapeamento de dados

Que dados estão disponíveis internamente? Não há dados disponíveis.

Que dados podem ser obtidos de fontes externas, privadas ou públicas?

	Exame de fundo de olho	Dados do paciente
O que contém?		
Qual é o nível de granularidade?		
Com que frequência as informações são coletadas e/ou atualizadas depois de captadas?		
Há identificadores exclusivos e confiáveis que podem ser vinculados a outras fontes de dados?		
Quem é o responsável interno pelos dados?		
Como são armazenados?		
Comentários adicionais		

FICHA DE FORMULAÇÃO E VIABILIDADE DE PROJETOS

EXEMPLO DART: Uso da inteligência artificial na prevenção da cegueira

5 Mapeamento de dados

Que dados estão disponíveis internamente? Não há dados disponíveis.

Que dados podem ser obtidos de fontes externas, privadas ou públicas?

	Exame de fundo de olho	Dados do paciente
O que contém?	Imagens de exame de fundo de olho.	Ficha médica do paciente, sexo, idade, data do diagnóstico etc.
Qual é o nível de granularidade?	Individual.	Individual.
Com que frequência as informações são coletadas e/ou atualizadas depois de captadas?	Em tempo real.	Na hora do exame.
Há identificadores exclusivos e confiáveis que podem ser vinculados a outras fontes de dados?	Sim, nome e RUT (Registro Único Tributário).	Sim, nome e RUT.
Quem é o responsável interno pelos dados?	Unidades de atenção (Ministério da Saúde).	Unidades de atenção (Ministério da Saúde).
Como são armazenados?	Imagens.	Banco de dados estruturado.
Comentários adicionais		

Teoria

1.6 Análise/ferramenta

Uma vez identificados os dados disponíveis e dependendo da natureza do problema, é hora de considerar as análises ou ferramentas de IA que serão necessárias para solucioná-lo. O papel do diretor do projeto é identificar os dados que julga mais adequados, levando em consideração que eles devem informar e aprimorar o processo de resposta atual, que, por sua vez, deve ajudar a atingir o objetivo, ou seja, materializar a solução proposta.

A IA permite realizar vários tipos de análise (descritiva, preditiva e de detecção), sendo exatamente este seu valor: possibilitar a realização de análises que antes não podiam ser realizadas com um processador de dados simples.

A seguir, serão apresentados alguns exemplos de uso da IA na América Latina e no Caribe, de acordo com a tarefa realizada pelo sistema de IA e com base nas categorias da OCDE (OCDE, a ser publicado em breve):[16]

Reconhecimento: baseia-se principalmente em uma categorização de imagens, textos e vídeos por meio da identificação de suas principais características, entre outros aspectos. Na América Latina, existem alguns exemplos de projetos em execução: o DART (Chile), que, como já vimos, detecta a retinopatia diabética por meio da análise de exames de fundo de olho; o Dymaxion Lab (Argentina), que monitora o povoamento humano, as inundações e o uso do solo por meio de imagens de satélite; e o MIDIS-Detecção precoce de anemia (Peru), que detecta casos de anemia por meio da análise de uma fotografia da conjuntiva ocular.

Detecção de eventos: o objetivo aqui é detectar padrões e anomalias. Um exemplo é a LAURA (Brasil), uma plataforma de IA usada para identificar pacientes com deterioração clínica com maior probabilidade de sofrer sepse. Com isso, a equipe clínica conta com as informações necessárias sobre os pacientes, o que permite focar sua atenção nos casos de maior risco.

Predição: o objetivo é prever um estado futuro com base em dados históricos. Alguns exemplos na região são a NotCo (Chile), empresa que utiliza um algoritmo para indicar os ingredientes ideais para criar produtos de origem vegetal que, geralmente, são de origem animal (maionese, leite etc.); o Traive (Brasil), um sistema de crédito alternativo que prevê o desempenho futuro do solicitante; e o Carabineros (Chile), que utiliza modelos de previsão de crimes para prever onde é mais provável que crimes sejam cometidos, informações que também são usadas na programação de turnos do pessoal.

Personalização: o objetivo é desenvolver perfis de usuários que, com base nos dados gerados por suas próprias ações, melhorem com o tempo. As redes sociais e os serviços de streaming usam esta ferramenta para fazer sugestões personalizadas. Um exemplo é o Livox (Brasil), um programa que permite que pessoas com deficiência verbal se comuniquem por meio de uma seleção de imagens. Com o tempo, o aplicativo começa a sugerir imagens prováveis aos usuários com base em suas interações anteriores.

16 Os casos listados aqui são uma seleção dos casos registrados no documento "Inteligência artificial a serviço do bem social na América Latina e no Caribe: panorama da região e retrato de doze países", da Iniciativa FAIRLAC do Banco Interamericano de Desenvolvimento (Gómez e outros, 2020).

Suporte à interação: esta é uma tarefa relacionada ao suporte à interação entre humanos e máquinas (chatbots, assistentes virtuais e outros). Existem alguns exemplos na região, como o Amanda Care (Argentina), uma assistente virtual que acompanha os pacientes durante seus tratamentos médicos, e o SpeakLiz (Equador), um software que transforma a língua de sinais em voz.

Otimização com foco em objetivos: trata-se de otimizar os processos dentro de uma instituição por meio de simulações de casos específicos. Por exemplo, o Kilimo, na Argentina, tem como objetivo ajudar os produtores agrícolas do país a otimizarem o uso da água em suas atividades. Para isso, ele coleta dados de campo e os combina com informações históricas e imagens de satélite em um software patenteado, por meio do qual fornece: 1) estimativas do consumo hídrico da cultura em sete dias; 2) conselhos periódicos resumidos em linguagem simples para os produtores sobre a quantidade da irrigação necessária para atingir suas metas de produção; e 3) informações relacionadas a riscos para fins de inteligência empresarial.

Raciocínio com estruturas de conhecimento: por meio deste tipo de análise, estabelece-se uma relação causal entre os dados disponíveis e um evento futuro inexistente. Ao contrário da previsão simples, este tipo de análise concentra-se na causalidade entre as variáveis. Um exemplo é o Portal de Telemedicina (Brasil), por meio do qual se busca prever diagnósticos médicos com base no histórico dos pacientes e a partir de uma análise causal. É, também, uma plataforma de colaboração entre as clínicas de diagnóstico e as equipes médicas.

Os tipos de análise resumidos aqui são apenas uma seleção de casos de uso da IA na região. A ideia é que sirvam como material para quando o diretor do projeto comunicar à equipe técnica qual deles pode ser mais adequado para solucionar o problema de política pública ao qual se pretende atender. É importante salientar que é comum que o sistema de IA utilizado em um projeto combine tarefas diferentes. É o caso dos chamados sistemas compostos, que detectam eventos anômalos e, a partir daí, fazem previsões.

As análises baseadas em IA também utilizam instrumentos variados, que incluem regressões, árvores de decisão e redes neurais. A decisão sobre a técnica de modelagem a ser utilizada será tomada pela equipe técnica, cujos membros devem explicar ao diretor do projeto como a análise foi realizada e quais foram os erros e dificuldades encontrados, antes de adotar a decisão sobre qual será a técnica a ser implementada com a ferramenta. Esses aspectos serão aprofundados no Item 2.2, que trata do desenvolvimento e validação do modelo.

Tarefas de um sistema de IA	Possível técnica de análise*
Reconhecimento	• redes neurais • máquinas de vetor de suporte (SVMs, na sigla em inglês)
Detecção de eventos	• redes neurais • máquinas de vetor de suporte
Predição	• regressão linear • árvores de regressão • redes neurais
Personalização	• filtros colaborativos
Suporte à interação	• redes neurais

Otimização com foco nos objetivos	• programação linear, não linear, dinâmica etc.
Raciocínio com estruturas de conhecimento	• redes bayesianas

*Lista não exaustiva.

Não se esqueça de...

1. **Identificar** o tipo de ferramenta e/ou análise mais adequado para solucionar o problema de política pública selecionado (regressões logísticas, árvore de decisão, redes neurais). Não é necessário decidir já no início a técnica de análise específica a ser usada na modelagem.

2. **Reunirse** com a equipe técnica (responsável pela modelagem) para discutir o problema de política pública e chegar a um acordo sobre o tipo de análise necessária.

3. **Incluir** a equipe técnica na equipe multidisciplinar do projeto.

Atividade Uma vez finalizada esta seção, preencha a parte correspondente à **análise/ferramenta** na Ficha de formulação e viabilidade de projetos.

VER FICHA EM ANEXO 1

FICHA DE FORMULAÇÃO E VIABILIDADE DE PROJETOS

6 Análise/ferramenta

	Análise 1	Análise 2
Tipo de análise (descrição, previsão, detecção, mudança de comportamento).		
Finalidade da análise (por exemplo, compreender o comportamento histórico das pessoas, estimar o risco de doença de um paciente).		
A qual ação dará respaldo esta análise?		
Como a análise será validada com base nos dados existentes? (por exemplo, usando-se dados históricos, realizando-se um estudo controlado e randomizado).		

FICHA DE FORMULAÇÃO E VIABILIDADE DE PROJETOS

EXEMPLO DART: uso de inteligência artificial na prevenção da cegueira

6 Análise/ferramenta

	Análise 1	Análise 2
Tipo de análise (descrição, previsão, detecção, mudança de comportamento).	Detecção.	
Finalidade da análise (por exemplo, compreender o comportamento histórico das pessoas, estimar o risco de doença de um paciente).	Obter automaticamente os casos anômalos que podem ser traduzidos em retinopatia diabética.	
A qual ação dará respaldo esta análise?	Ação 2 (veja o Item 6 correspondente à "Descrição das ações"). Agora, o oftalmologista terá à sua disposição informações filtradas no momento de diagnosticar o paciente.	
Como a análise será validada com base nos dados existentes? (por exemplo, usando-se dados históricos, realizando-se um estudo controlado e randomizado).	Treinamento e avaliação com dados históricos de exames de fundo de olho e com o diagnóstico correspondente.	

Teoria

1.7 Considerações éticas, jurídicas e de governança

Nesta etapa, é necessário fazer um diagnóstico das considerações éticas, jurídicas e de governança do projeto. Isso permitirá dimensionar seus riscos e traçar um plano adequado de acordo com sua complexidade. Cabe advertir, mais uma vez, que os algoritmos e as regras automatizadas para a tomada de decisão podem reproduzir vieses e, portanto, os cidadãos podem resistir à sua implementação. Também é possível que existam nos modelos grupos populacionais sub-representados e, portanto, prejudicados pela implementação da ferramenta. A fAIr LAC identificou os seguintes desafios transversais para todos os projetos de IA:

Proteção de dados pessoais: ao implementar um sistema de suporte à decisão e/ou tomada de decisão baseado em IA, é provável que se trabalhe com dados pessoais[17] sensíveis, como nível de renda, diagnósticos médicos ou outras informações de saúde, bem como identificadores exclusivos, como o documento nacional de identidade (DNI) ou dados biométricos.

O diretor do projeto deve assegurar que o tratamento dado a essas informações cumpra as normas nacionais de proteção de dados e seja orientado pelas boas práticas internacionais.

O tratamento dos dados pessoais deve ter sempre um fundamento legal, ou seja, um suporte que dê legitimidade jurídica ao que será feito com esses dados. Dependendo do tipo de projeto e dos regulamentos relativos a dados pessoais de cada país, a base jurídica pode variar. O mais comum é o consentimento informado, bem como as leis orgânicas dos órgãos públicos, o interesse legítimo, o interesse vital ou a existência de um contrato. Por esse motivo, é importante familiarizar-se com a lei de proteção de dados do país em questão.[18]

Uma boa prática é, sempre que possível, obter o consentimento das pessoas em estudo em relação ao uso de seus dados. Consentimento é entendido como a "expressão livre, específica, inequívoca e informada da vontade do titular, por meio da qual ele aceita e autoriza o tratamento dos dados pessoais que lhe correspondam" (Rede Ibero-americana de Proteção de Dados, 2017).

Para que o consentimento seja informado, os indivíduos precisam conhecer, no mínimo, a identidade do responsável pelo tratamento dos dados e as finalidades para as quais os dados pessoais serão usados. Também precisam conhecer a política de segurança adotada no tratamento e armazenamento de seus dados, as vantagens associadas à sua utilização, os riscos em potencial e as medidas de mitigação.

Uma boa prática é anonimizar ou pseudonimizar as observações pessoais; no entanto, o diretor do projeto não pode se esquecer de que "os dados pessoais não se limitam a nomes e sobrenomes, mas também incluem qualquer elemento que possa levar à identificação do sujeito específico" (Buenadicha et al., 2019). Por esse motivo, deve-se ter cuidado não apenas para eliminar personalizações óbvias nos dados, mas também para codificar os possíveis caminhos e cruzamentos de dados que permitiriam a identificação de um indivíduo.

17 Os dados pessoais são aqueles que permitem identificar direta ou indiretamente uma pessoa (Rede Ibero-americana de Proteção de Dados, 2017).

18 Para mais informações sobre as leis de proteção de dados pessoais em diferentes países da América Latina, consulte Bojalil e Vela-Treviño (2018).

Tanto o conhecimento sobre o uso de dados pessoais e sensíveis quanto a necessidade de obter o consentimento informado devem ser levados em consideração no processo de conceitualização e formulação do projeto. Se, durante sua implementação, o uso dos dados mudar, isso deve ser devidamente comunicado aos indivíduos, para que deem seu consentimento referente ao novo uso.

Responsabilidade e transparência: o diretor do projeto será o responsável final pela implementação da ferramenta de IA, devendo informar os cidadãos sobre seus benefícios, riscos e medidas de mitigação correspondentes. Responsabilidade é entendida como prontidão para prestar contas sobre suas próprias ações ou atividades, criações ou pessoas sob sua responsabilidade, aceitando as consequências desses atos (Oliver, 1994). No caso da IA, implica a responsabilidade pelas decisões tomadas pelo sistema e pelos efeitos que este possa causar. Para isso, o diretor do projeto deve criar mecanismos simples para que os afetados possam apresentar pedidos de revisão das decisões ou queixas sobre o funcionamento da ferramenta.

Em relação à transparência algorítmica, é um princípio ético essencial na IA, porque funciona como um facilitador de outros princípios: permite saber quais dados são usados, quem os usa, como são usados e como isso afeta as decisões relativas às políticas públicas (Sangüesa, 2018). Quando um sistema é transparente, é possível saber, por exemplo, se os dados estão protegidos e se os resultados são equitativos.

Os sistemas de tomada de decisão e/ou suporte à decisão baseados em IA muitas vezes são vistos como caixas-pretas: a população em geral não sabe quais de seus dados estão sendo usados, como são usados, qual são as implicações de as decisões agora estarem sendo tomadas com base em um sistema automatizado e qual é o impacto final que isso terá na sociedade. Aqui, o papel do diretor do projeto é fornecer aos cidadãos as informações necessárias e suficientes para que entendam os mecanismos gerais da ferramenta e se sintam seguros em relação à proteção de seus dados e a como as decisões serão tomadas.

Além da transparência do modelo em si, caberá ao diretor do projeto a responsabilidade de informar os cidadãos em geral e as partes envolvidas sobre os diferentes passos e as decisões tomadas na etapa de implantação, e até mesmo durante a avaliação do modelo em campo. Uma boa prática é ir documentando todas as decisões adotadas para manter registros e informar adequadamente, quando necessário.

Interpretabilidade e explicabilidade: este desafio advém da necessidade de entender os modelos de suporte à decisão e tomada de decisão baseados em IA. Visto que, como já indicado, os algoritmos costumam ser vistos como caixas-pretas, o papel do diretor do projeto aqui é explicar à população o funcionamento geral da ferramenta.

A explicabilidade e a transparência algorítmica de um modelo, bem como a quantidade de informações que podem ser fornecidas sobre o funcionamento de um algoritmo, dependerão do nível de opacidade (Burrell, 2016; Buenadicha et al., 2019). Existem três níveis:

> ***Opacidade intencional:*** haverá situações em que o algoritmo não poderá ser explicado aos cidadãos, pois isso comprometeria a efetividade do próprio modelo. Por exemplo, se um modelo for aplicado para prever a evasão fiscal, os cidadãos não devem saber como ele funciona especificamente, pois alguns podem tentar "enganar" o modelo. Essa opacidade intencional também pode ocorrer por motivos de propriedade intelectual.

Opacidade analfabeta: ocorre quando as pessoas a quem o algoritmo é explicado não entendem como ele funciona por falta de conhecimento em assuntos técnicos. Nesse caso, é recomendável capacitar os funcionários da linha de frente da instituição, para que possam explicar, em linguagem clara, como o sistema funciona.

Opacidade intrínseca: às vezes, os algoritmos usados são extremamente complexos; por isso, é difícil mostrar uma relação causal entre os valores de entrada e saída. Se este tipo de opacidade existir, será impossível explicar o modelo.

Algumas ferramentas são mais explicáveis (mais fáceis de ser entendidas pela população em geral) que outras. Por isso, o diretor do projeto deve informar claramente a equipe técnica sobre os requisitos de explicabilidade à luz dos regulamentos nacionais e da definição do problema, pois tudo isso terá um impacto no modelo que a equipe desenvolver. Embora nem sempre seja necessário fornecer todas as informações, os cidadãos devem estar cientes de como as decisões relativas às políticas públicas estão sendo tomadas.

Além dos desafios descritos acima, desde a fase de formulação de um projeto, é importante questionar-se sobre a possibilidade de a implementação da ferramenta gerar discriminações negativas. Elas podem ter origem em:

Vieses: o erro do sistema é a diferença entre o valor previsto resultante do modelo e o valor real da variável que está sendo estimada. Se o erro for sistemático em uma direção ou em um subconjunto específico de dados, ele é chamado de **viés**. Por exemplo, se uma balança sempre pesa um quilo a mais, ela tem um viés. Da mesma forma, se um valor é sempre menor (como o salário das mulheres em contraste com o dos homens para um trabalho equivalente), a variável "salário" tem um viés. Quando o erro não é sistemático, mas aleatório, é chamado de **ruído** (González, Ortiz e Sánchez Ávalos, 2020).

O viés de um sistema de IA pode ter implicações éticas quando seus resultados são usados para tomar decisões de políticas públicas que levam a ações que possam ser consideradas injustas ou discriminatórias para alguns subgrupos da população destinatária. Por exemplo, um modelo ou ferramenta de IA pode favorecer ou prejudicar constantemente o mesmo setor da população. Existem diferentes tipos de viés, alguns dos quais são causados por problemas intrínsecos aos dados. Existem os **vieses históricos** ou estados indesejados, que são registrados quando surgem no mundo padrões que não se deseja reproduzir ou propagar no modelo; **vieses de representação**, que ocorrem quando as informações estão incompletas, seja por atributos faltantes, pelo desenho da amostra ou pela ausência total ou parcial de subpopulações; e os **vieses de mensuração**, oriundos do uso ou omissão de variáveis que serão utilizadas nos modelos (Harini Suresh, 2019).

Para treinar uma ferramenta de IA, são necessários dados históricos com os quais ela "aprenderá" e mostrará resultados. Cabe ao diretor do projeto, dado seu conhecimento especializado sobre o problema e a situação atual, antecipar quais setores da população podem estar sub-representados nos dados ou quais problemas de viés podem afetá-los. Os vieses podem ser mitigados com alguma medida de reparação ou com uma calibração da ferramenta de IA pela equipe técnica na etapa de execução.

Se os dados estiverem distorcidos, a ferramenta também pode estar. Uma ferramenta enviesada coloca em risco a viabilidade de um projeto de IA. Por exemplo, a Amazon desenvolveu

uma ferramenta de recrutamento baseada em IA, mas precisou tirá-la do ar, pois favorecia significativamente o recrutamento de homens. Isso ocorreu porque a indústria de tecnologia foi historicamente dominada por eles, o que enviesou os dados históricos de contratação. Dada a discriminação arbitrária da ferramenta, ela precisou ser descartada (Dastin, 2018).

Desigualdades no processo no qual se atuará: é provável que a situação atual, sem um projeto de IA, registre desigualdades entre diferentes subgrupos da população. Por exemplo, as taxas de suicídio são mais altas entre os jovens LGBTQIA+ que entre a população jovem em geral. No caso da violência doméstica, a maior proporção de vítimas são mulheres. O diretor do projeto deve estar ciente dessas linhas de referência para incorporar as considerações do caso e analisar se uma ferramenta de IA pode ou não ajudar a solucionar o problema em questão.

Subgrupos da população para os quais se busca garantir a igualdade: embora seja verdade que um banco de dados pode estar livre de vieses, muitas vezes, ocorre que alguns grupos minoritários da população perdem representação no modelo de IA. Isso porque, ao usar dados históricos, o modelo dará o mesmo peso a todos os indivíduos ou observações, o que, por sua vez, leva a uma menor representação dos grupos que constituem uma porcentagem menor da população. Como especialista no problema de política pública a ser abordado, o diretor do projeto é quem tem clareza sobre como a implementação da ferramenta de IA pode afetar grupos minoritários de maneiras diferentes e quem, portanto, deve garantir a equidade na intervenção e implementação da ferramenta, definindo grupos ou atributos protegidos. Por exemplo, caso se deseje implementar um sistema de detecção de doenças, não seria aceitável que ele detectasse a patologia com taxas de precisão diferentes para homens e mulheres. Ou, se uma determinada instituição aplicasse um modelo automático de promoção, isso poderia favorecer pessoas com mais experiência em detrimento dos jovens.

Caberá ao diretor do projeto informar a equipe técnica sobre a discriminação que pode advir dos dados, identificando claramente os subgrupos da população para os quais se pretende garantir a igualdade e tomando medidas que garantam a justiça algorítmica (ou seja, que os algoritmos não prejudiquem alguns subgrupos da população e que os resultados sejam equitativos para todos). Atualmente, existem métricas que permitem determinar as iniquidades que podem surgir na etapa de modelagem. O diretor do projeto deve solicitar à equipe técnica a realização de diversos testes que permitam medir o nível de justiça algorítmica da ferramenta (seguindo definições como paridade demográfica, igualdade de possibilidades ou justiça contrafactual, entre outras).[19] Considerando-se que haverá casos nos quais a ferramenta não poderá intervir, poderão ser adotadas medidas para mitigar vieses e discriminações não por meio da IA, mas das próprias políticas públicas.

Além dos pontos mencionados acima, nesta fase, o diretor do projeto deve perguntar-se se obteve a **licença social**. Licença social é entendida como a aceitação, pelos cidadãos, da implementação da ferramenta de IA para um sistema de suporte à decisão e/ou tomada de decisão (Data Futures Partnership, 2017). Não basta estar em conformidade com os quadros jurídicos dos países; é necessário um passo adicional, que é obter a aceitação da população. No início, a IA foi muito bem recebida, pois significava um aumento da produtividade e uma diminuição do tempo de trabalho, deslocamento etc. No entanto, mais recentemente, as pessoas estão tomando cada vez mais consciência dos efeitos adversos em potencial da IA em suas vidas; portanto, obter a licença social é fundamental quando se pretende implementar um projeto dessa natureza, quer isso seja feito por uma instituição pública ou privada (Hewitt, 2019). Para saber se um projeto possui licença social, seu diretor pode consultar estudos de opinião sobre o valor público que se espera que a ferramenta

19 Para mais detalhes, consulte González, Ortiz y Sánchez Ávalos (2020).

gere em contraste com os riscos que ela pode acarretar, além de realizar reuniões com os grupos de interesse e comunidades onde ela será implementada para saber quais são suas percepções a respeito dela.

Governança e segurança: os desafios de governança e segurança são transversais aos projetos de IA e estão relacionados à regulamentação e à segurança da infraestrutura na qual a ferramenta é desenvolvida. Três desafios são identificados aqui:

Normas e regulamentos: o escopo de um projeto de IA será enquadrado nas normas e regulamentos dos países onde ele estiver sendo implementado e do setor da política pública para a qual a ferramenta for utilizada. Espera-se que o projeto seja regido pelas leis de proteção de dados pessoais e acesso à informação. Vale notar que existem setores específicos, como o da saúde, que são regulados por normas especiais e que vários países possuem leis antidiscriminação. Portanto, sempre que um sistema de IA for implementado, será necessário compilar todas as normas pertinentes. Dispor dessas informações na fase de conceitualização contribuirá para a viabilidade da implementação. Recomenda-se procurar o departamento jurídico da instituição e incluí-lo na equipe do projeto.

Cibersegurança: os sistemas de IA são treinados com dados; portanto, tanto os dados pessoais quanto os resultados dos algoritmos devem ser devidamente protegidos para evitar vazamentos de informações sensíveis pertencentes aos cidadãos. Por esse motivo, o diretor do projeto deve estabelecer protocolos de segurança para o armazenamento e tratamento dos dados, bem como para possíveis casos de furto de informações.

Penetração ilegal da IA e ataques adversários: além do furto de informações sensíveis, o sistema pode ser alvo de ataques que buscam confundir o algoritmo, nos quais o agressor se faz passar por usuário da ferramenta. Aqui também devem ser instalados sistemas de segurança para evitá-los, além de protocolos de resposta caso ocorram.

Não se esqueça de...

1. **Fazer** um inventário das leis do país aplicáveis ao seu projeto para cumprir os requisitos legais.
2. **Obter** as autorizações legais necessárias para utilizar informações que podem ser pessoais e/ou sensíveis.
3. **Determinar** todos os riscos éticos que podem surgir durante a implementação da ferramenta.
4. **Definir** claramente os grupos populacionais para os quais se pretende garantir a igualdade na intervenção (justiça algorítmica).
5. **Avaliar** se o projeto possui licença social, ou seja, se a população aceita o uso proposto da IA.

Atividade Uma vez finalizada esta seção, preencha a parte correspondente às **considerações éticas e jurídicas** na Ficha de formulação e viabilidade de projetos.

VER FICHA EM ANEXO 1

FICHA DE FORMULAÇÃO E VIABILIDADE DE PROJETOS

7	Considerações éticas e jurídicas	
Proporcio-nalidade	Você acha que um sistema de ciência de dados/IA é o meio certo para solucionar o problema? Por quê? Já avaliou alternativas?	
	Que impactos negativos o seu projeto pode ter? Estude casos de uso semelhantes identificados na seção "Definição do problema".	
Licença social	Você acha que a população aceitará o uso dos dados propostos para solucionar o problema? Por quê?	
	Se a população do projeto tomar ciência do uso da IA para os propósitos declarados, ela o aprovará? Por quê?	
Proteção de dados	Você está trabalhando com dados pessoais e/ou sigilosos identificáveis individualmente? Quais?	
	Você identificou a justificativa ou base jurídica para trabalhar com esses dados?	
	Você identificou os regulamentos que podem afetar o projeto?	
	São necessários mecanismos (de acesso, eliminação ou retificação, por exemplo) para garantir a qualidade dos dados pessoais?	

Transpa-rência	Quais partes interessadas devem estar cientes do projeto? (Responsáveis pela formulação de políticas públicas, trabalhadores da linha de frente, organizações da sociedade civil, órgãos públicos, pessoas que serão afetadas pelas ações etc.) Mencione organizações/ indivíduos específicos.	
	Você considerou um mecanismo para que as partes interessadas entrem em contato com a instituição para obter informações sobre o projeto?	
	Será necessário explicar os mecanismos de tomada de decisão ou as análises a serem implementadas? Por quê?	
Discrimi-nação/equi-dade	Que desigualdades estruturais existem no processo ou ambiente no qual o projeto está inserido?	
	Existem grupos específicos (vulneráveis) para os quais se busca garantir a equidade nos resultados ou a proteção de seus direitos? (Por exemplo, por gênero, idade, localização, classe social, escolaridade, origem urbana ou rural, etnia.)	
	Que vieses você acha que os dados podem ter?	
Prestação de contas	Caso seja apresentada uma solicitação de informações sobre o projeto, quem é responsável por responder?	
	Quem é responsável por quaisquer erros cometidos pelo sistema?	
	Existe uma previsão de mecanismos de monitoramento, controle e avaliação? Como e com que frequência eles serão documentados?	

FICHA DE FORMULAÇÃO E VIABILIDADE DE PROJETOS		
EXEMPLO DART:	uso da inteligência artificial na prevenção da cegueira	
7	**Considerações éticas e jurídicas**	
Proporcio-nalidade	**Você acha que um sistema de ciência de dados/ IA é o meio certo para solucionar o problema? Por quê? Já avaliou alternativas?**	Sim, visto que não é viável aumentar substancialmente o número de oftalmologistas no curto prazo e que o projeto permite ampliar a cobertura. Isso, por sua vez, contribuirá para aumentar a detecção precoce da retinopatia diabética e preveni-la, levando a uma redução de custos para o Ministério da Saúde. Da mesma forma, uma ferramenta é necessária para aumentar significativamente o número de exames de fundo de olho que são informados. A opção de formar mais oftalmologistas é uma solução muito a longo prazo.
	Que impactos negativos o seu projeto pode ter? Estude casos de uso semelhantes identificados na seção "Definição do problema".	Um possível risco é que a qualidade das imagens seja insuficiente para realizar a análise. Entre os possíveis impactos adversos, os resultados dos exames podem ser falsos negativos, não sendo, portanto, avaliados por um oftalmologista. Isso causaria danos aos pacientes cujos resultados estão incorretos e que sofrem dessa doença.
Licença social	**Você acha que a população aceitará o uso dos dados propostos para solucionar o problema? Por quê?**	Sim, porque os pacientes terão um benefício claro, isto é, a possibilidade de revisão de seus exames, o que aumenta a cobertura do check-up anual e poupa o tempo dos oftalmologistas, que agora podem se concentrar nos casos mais extremos. Quanto a possíveis resistências de oftalmologistas e tecnólogos médicos, isso pode advir do fato de que precisarão mudar sua forma de realizar os exames e diagnosticar os casos. É importante incluí-los no projeto para obter sua adesão à implementação.
	Se a população do projeto tomar ciência do uso da IA para os propósitos declarados, ela o aprovará? Por quê?	Sim, pois significa um claro benefício para todos, uma vez que libera recursos humanos e financeiros dos centros de saúde e foca os existentes nos casos que os necessitam. De qualquer forma, é importante relatar os benefícios de longo prazo do uso da ferramenta nos centros de saúde, dado o investimento inicial que o ministério da área deve fazer. No entanto, pode haver alguma resistência devido ao fato de que o resultado do exame é gerado por um algoritmo. É necessário, portanto, que o fator humano seja incorporado e que um especialista confirme os casos de retinopatia diabética encontrados automaticamente, para que a equipe de saúde possa, então, comunicar o diagnóstico ao paciente. No momento da implementação, será necessário ajustar as expectativas dos pacientes, pois a ferramenta foi projetada para detectar apenas a retinopatia diabética, não outras doenças oculares.

Proteção de dados	**Você está trabalhando com dados pessoais e/ou sigilosos identificáveis individualmente? Quais?**	Sim, são dados sigilosos porque se referem ao estado de saúde das pessoas.
	Você identificou a justificativa ou base jurídica para trabalhar com esses dados?	Sim, a proteção de dados pessoais, de acordo com a lei 19.628 - Lei de direitos e deveres do paciente (n.º 20.584) e o regulamento correspondente, o CSAN, DFL 1 Salud de Chile.
	Você identificou os regulamentos que podem afetar o projeto?	Sim, a proteção de dados pessoais, de acordo com a lei 19.628 - Lei de direitos e deveres do paciente (n.º 20.584) e o regulamento correspondente, o CSAN, DFL 1 Salud de Chile.
	São necessários mecanismos (de acesso, eliminação ou retificação, por exemplo) para garantir a qualidade dos dados pessoais?	Os pacientes devem poder acessar as imagens de seus exames de fundo de olho e os resultados da revisão do oftalmologista, se necessário, por exemplo, para realizar uma consulta médica particular. Caso contrário, eles podem se recusar a fazer o exame.
Transpa-rência	**Quais partes interessadas devem estar cientes do projeto? (Responsáveis pela formulação de políticas públicas, trabalhadores da linha de frente, organizações da sociedade civil, órgãos públicos, pessoas que serão afetadas pelas ações etc.) Mencione organizações/indivíduos específicos.**	Os centros de saúde e clínicas que implementarem o software, o Ministério da Saúde, a Sociedade Chilena de Oftalmologia (SOCHIOF), o Colégio de Tecnólogos Médicos do Chile, a Confederação Nacional de Funcionárias e Funcionários da Saúde Municipal (Confusam), a Associação de Diabéticos do Chile (Adich) e a Fundação Diabetes Juvenil do Chile.
	Você considerou um mecanismo para que as partes interessadas entrem em contato com a instituição para obter informações sobre o projeto?	Sim, por meio de reuniões periódicas com os responsáveis do Ministério da Saúde. O Ministério da Saúde, por sua vez, também coleta informações de percepção nos centros de saúde.
	Será necessário explicar os mecanismos de tomada de decisão ou as análises a serem implementadas? Por quê?	Sim. Embora não o algoritmo em si, pois é altamente técnico, deve ser possível explicar sua validação clínica.

Discriminação/equidade	**Que desigualdades estruturais existem no processo ou ambiente no qual o projeto está inserido?**	Existem desigualdades no que refere aos fatores de risco de diabetes (Sapunar, 2016). O risco de diabetes é maior em pessoas mais velhas, do sexo feminino, que vivem em áreas urbanas e têm um nível de escolaridade mais baixo. Também se descobriu que o risco é menor entre os povos originários.
	Existem grupos específicos (vulneráveis) para os quais se busca garantir a equidade nos resultados ou a proteção de seus direitos? (Por exemplo, por gênero, idade, localização, classe social, escolaridade, origem urbana ou rural, etnia.)	Não. Serão analisados os exames de todos os pacientes diabéticos. É possível que o algoritmo apresente um desempenho diferente para alguns subgrupos da população diabética; por isso, seria aconselhável realizar uma análise de disparidade.
	Que vieses você acha que os dados podem ter?	Nenhum, pois os fatores de risco descritos anteriormente não devem influenciar a qualidade da imagem do fundo de olho.
Prestação de contas	**Caso seja apresentada uma solicitação de informações sobre o projeto, quem é responsável por responder?**	O Ministério da Saúde e a Teledx.
	Quem é responsável por quaisquer erros cometidos pelo sistema?	O Ministério da Saúde.
	Existe uma previsão de mecanismos de monitoramento, controle e avaliação? Como e com que frequência eles serão documentados?	Sim. Um estudo observacional está previsto para assim que a ferramenta for implementada em centros suficientes, a fim de obter uma amostra suficientemente grande. Da mesma forma, será feito um monitoramento periódico da ferramenta quanto à qualidade das previsões, vieses algorítmicos, experiência do usuário e análise de lacunas no acesso aos exames.

Teoria

1.8 Formação da equipe

Uma parte crucial do processo de conceitualização e formulação do projeto consiste em identificar as pessoas da instituição (ou externas) que devem fazer parte da equipe multidisciplinar responsável e envolvê-las desde o início. A participação ativa dessas pessoas na formulação ajudará a garantir a sustentabilidade da ferramenta durante a implementação.

Ao longo de todo esse processo, o diretor do projeto conversa com pessoas de dentro e fora da instituição sobre a iniciativa proposta, e incorpora seu aprendizado à formulação. Por exemplo, se for um projeto para otimizar as inspeções aduaneiras, certamente deverão ser realizadas reuniões com a diretoria responsável para entender como essas inspeções são conduzidas atualmente, como decidem o que inspecionar, quanto tempo demoram e quais problemas surgem. Também é provável que o departamento de estatística seja contatado para solicitar informações sobre os dados existentes e avaliar sua maturidade. Todas essas pessoas devem fazer parte da equipe do projeto, pois o impacto também as afetará. No caso da equipe encarregada das inspeções, serão seus membros que utilizarão a ferramenta de IA; portanto, é importante que tenham plena consciência das dificuldades atuais e futuras relativas à implementação. O departamento de estatística certamente se encarregará de gerenciar internamente os dados; por isso, seus especialistas ficarão encarregados de fornecê-los à equipe técnica.

A seguir, serão descritos alguns cargos dos possíveis membros de um projeto de IA e suas funções. A lista é apenas uma sugestão, ou seja, a formação da equipe técnica não será abordada de forma detalhada, pois isso dependerá da natureza do projeto.

Diretor do projeto (responsável pela tomada de decisão)	Desempenha o papel principal na etapa de conceitualização e formulação, sendo responsável por supervisionar todos os passos dados pela equipe técnica na etapa de execução.
Referência do processo	Participa do processo de implementação da ferramenta de IA; é respeitado por seus colegas por sua experiência e conhecimento. Seus conhecimentos especializados e reconhecimento contribuirão para a adoção da tecnologia.
Responsáveis pelos dados (internos e/ou externos)	Fornecem informações sobre a qualidade dos dados necessários para a implementação da ferramenta. Determinam as necessidades em termos de coleta, armazenamento e manuseio. Aplicam seu conhecimento especializado para detectar possíveis anomalias nos dados.
Analistas de dados	Realizam a análise exploratória dos dados e desenvolvem a ferramenta. Caso essas capacidades não existam dentro da instituição, o diretor do projeto deve contratar esse serviço de uma instituição externa.
Equipe jurídica	Assegura o cumprimento das normas jurídicas vigentes em cada país em matéria de proteção de dados pessoais, transparência e regulamentos de cada setor.

Linha de diretoria da instituição	Apoia e promove o desenvolvimento da ferramenta. É fundamental para viabilizar o projeto e permitir sua implementação.
Equipe que executa a ação	Possui conhecimento especializado de como o problema está sendo respondido atualmente e como essa resposta pode ser aprimorada. Apresenta receios sobre como os funcionários interagirão com a ferramenta e as possíveis consequências disso.
Comunicações	Comunica interna e externamente os principais aspectos da implementação da ferramenta.
Departamento de estudos ou avaliação	Contribui para a elaboração dos aspectos metodológicos da avaliação do projeto.

Não se esqueça de...

1. **Considerar** todas as partes interessadas relevantes na hora de formar a equipe.
2. **Obter** a adesão dos membros da equipe ao projeto.
3. **Certificar-se** de que todos os membros da equipe participem da conceitualização e formulação do projeto.

Atividade Uma vez finalizada esta seção, preencha a parte correspondente à **formação da equipe** na Ficha d formulação e viabilidade de projetos.

VER FICHA EM ANEXO 1

FICHA DE FORMULAÇÃO E VIABILIDADE DE PROJETOS

8 Formação da equipe

Organização/ departamento	Nome/função da contraparte	Descrição da participação desejada

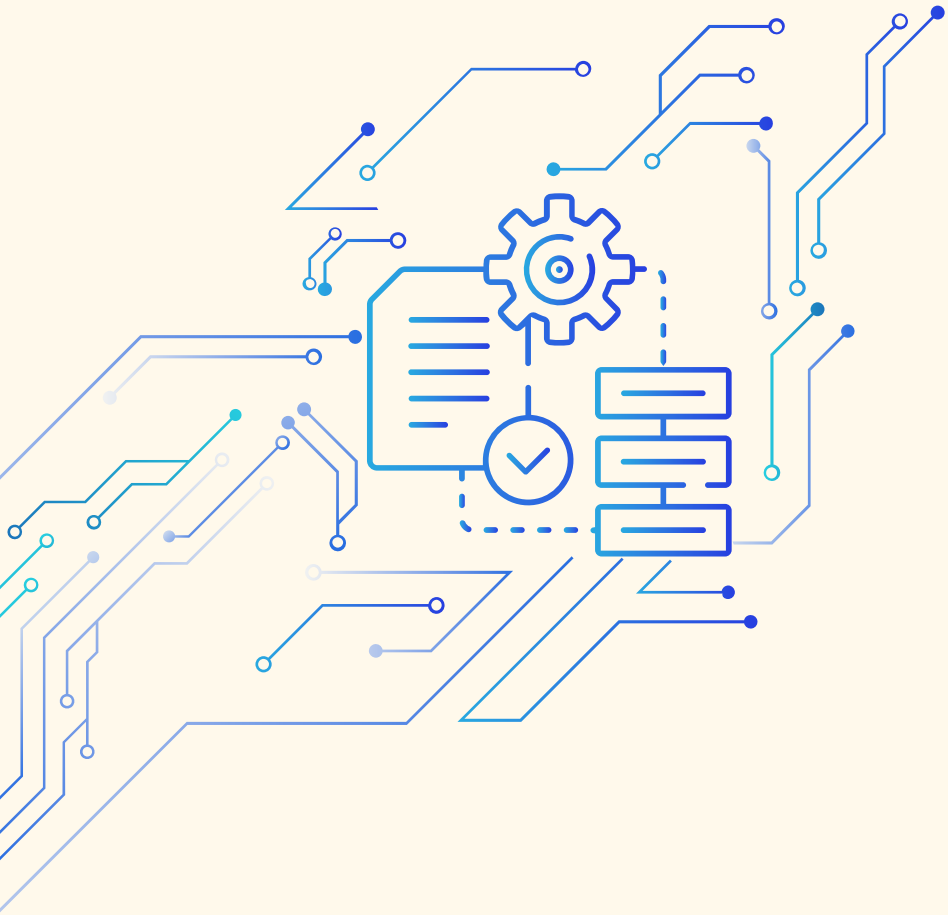
FICHA DE FORMULAÇÃO E VIABILIDADE DE PROJETOS

EXEMPLO DART: Uso da inteligência artificial na prevenção da cegueira

8 Formação da equipe

Organização/ departamento	Nome/função da contraparte	Descrição da participação desejada
Teledx	Analistas de dados	Responsáveis pela criação da ferramenta de análise de imagens
Ministério da Saúde	Subsecretaria de redes assistenciais	Implementação da ferramenta nos centros de saúde
Ministério da Saúde	Divisão jurídica	Assessoria para garantir a conformidade do projeto com as normas vigentes
Teledx/Ministério da Saúde	Equipe de treinamento	Treinamento de oftalmologistas e tecnólogos médicos no uso da ferramenta

PARTE 2. EXECUÇÃO

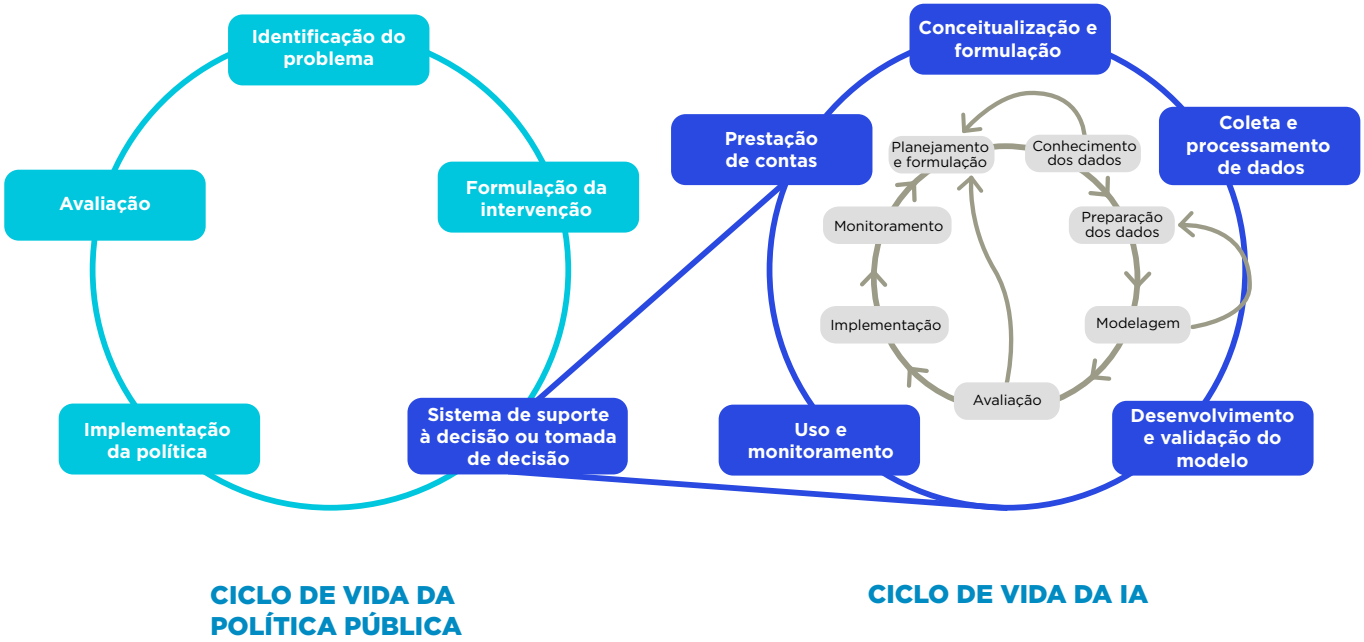


PARTE 2. EXECUÇÃO

Assim que o projeto for formulado e a instituição decidir realizá-lo, é hora de executá-lo. Como isso é feito? Quais aspectos precisam ser considerados durante esta fase? Até onde deve ir a transparência e a prestação de contas em relação às especificidades do modelo?

Este manual descreve os desafios que um projeto de IA implica tanto para seu diretor quanto para a equipe técnica, dependendo de seu ciclo de vida. Esses desafios também podem ser compreendidos com base nas diferentes fases que a equipe responsável precisa desenvolver, onde o diretor desempenha um conjunto de funções fundamentais (Figura 5):

Figura 5. Conexão entre políticas públicas e desenvolvimento da IA



Fonte: González, Ortiz y Sánchez Ávalos, 2020.

1. Coleta e processamento de dados: na etapa de conceitualização e formulação, o diretor do projeto já mapeou os diferentes bancos de dados e determinou seu grau de maturidade. Nesta fase, a equipe técnica iniciará uma análise exploratória dos dados para obter mais informações e identificar os desafios que podem surgir durante sua utilização.

2. Desenvolvimento e validação do modelo: depois de definidas as variáveis a serem utilizadas e compreendidos os desafios apresentados pelos dados, é hora de a equipe técnica começar a desenvolver o modelo, utilizando técnicas de modelagem específicas. Na etapa de conceitualização e formulação, o diretor do projeto deve ter comunicado à equipe técnica o objetivo da análise: "Que tipo de análise (ou ferramenta) é mais adequada para solucionar o problema?" Agora, a equipe técnica deve apresentar diferentes opções de modelagem ao diretor do projeto para, juntos, decidirem qual é a tecnologia mais adequada de acordo com critérios técnicos e de equidade. É possível que, durante o desenvolvimento e validação do modelo, a equipe técnica identifique novas necessidades

de dados, devendo ser realizadas iterações do modelo para integrar as novas fontes de informação. O processo deve ser repetido até que o(s) modelo(s) gerado(s) seja(m) satisfatório(s) para a equipe técnica e para o diretor do projeto.

3. Uso e monitoramento: ao implementar a ferramenta de IA na população destinatária ou em um subconjunto dela, seu impacto deve ser medido, levando-se em consideração os objetivos do projeto. Também será necessário implementar um sistema de monitoramento periódico para corrigir os erros que podem surgir durante sua implementação e alocar os recursos necessários para garantir o uso da ferramenta ao longo do tempo.

4. Prestação de contas: sendo o diretor do projeto o responsável final pela implementação da ferramenta de IA, cabe-lhe documentar, juntamente com a equipe técnica, todo o processo de desenvolvimento a fim de justificar ao público e aos demais interessados as decisões tomadas durante todo o ciclo de vida da IA.

Nos passos 1 e 2 descritos acima, a equipe técnica será a principal responsável pelo modelo, pois ficará encarregada da modelagem dos dados. No entanto, caberá ao diretor do projeto supervisionar cada um dos passos e decisões da equipe técnica. No início de cada passo, será apresentado o papel do diretor do projeto.

No manual de González, Ortiz e Sánchez Ávalos (2020) mencionado neste texto, há dois materiais dirigidos à equipe técnica que serão incluídos a seguir para promover uma comunicação fluida entre a equipe e o diretor do projeto: o perfil de dados e o perfil do modelo de IA. Ambos servirão de apoio à documentação das decisões do projeto em nome da transparência e da prestação de contas.

1. Perfil de dados:
contém os principais achados dos bancos de dados a serem utilizados com a ferramenta. Baseia-se nas análises exploratórias dos dados realizadas pela equipe técnica sob a supervisão do diretor do projeto, com quem esses achados também serão debatidos.

2. Perfil do modelo:
descrição final de um modelo de IA. Contém aspectos cruciais para todo o ciclo de vida da IA.

Para ambas as ferramentas, as funções do diretor do projeto serão explicadas em cada um dos aspectos principais das etapas de execução.

Teoria

2.1 Coleta e processamento de dados

Papel do diretor do projeto
Durante a fase correspondente à coleta e processamento de dados, o diretor do projeto será responsável por:

- contextualizar os achados da equipe técnica na etapa exploratória dos dados;
- fornecer o histórico do processo de coleta de dados e o significado das variáveis, conectando a equipe técnica aos indivíduos da organização necessários para uma melhor compreensão dos dados;
- validar a formulação e utilidade da variável-alvo;
- tomar decisões sobre imputações de dados ou eliminação de variáveis, ponderando e registrando o que se obtém em contraste com o que se perde (*trade-offs*);
- propor melhorias nos processos de governança de dados do projeto que possam contribuir para o atingimento dos objetivos e facilitar a implementação do modelo; e
- ajustar a definição do problema, os objetivos e o plano de execução do projeto com base nos resultados desta etapa.

Neste caso, a equipe técnica deve, primeiro, realizar uma análise exploratória dos dados. Embora se espere que o diretor do projeto tenha um conhecimento básico dos bancos de dados e sua maturidade (conforme registrado no Item 1.5 da primeira parte, correspondente ao mapeamento de dados), é nesta fase que a equipe técnica começa a realizar uma análise detalhada e completa desses dados, usando ferramentas estatísticas convencionais.

Aqui, o papel do diretor do projeto é estabelecer uma conexão entre a equipe técnica e as áreas que geram ou gerenciam os dados, fornecendo informações sobre a metodologia de coleta, a frequência de atualização e até mesmo o dicionário de variáveis. Isso servirá para tornar a comunicação entre ambas as partes mais fluida, caso surjam dúvidas metodológicas ou conceituais.

Em segundo lugar, quando a análise exploratória for realizada, a equipe técnica poderá deparar-se com situações consideradas anômalas. Aqui, como o diretor do projeto é a pessoa que tem o conhecimento mais especializado sobre o problema de política pública a ser solucionado e que pode entrar em contato com especialistas do setor que tenham a experiência que o caso pede, ele deve ajudar a identificar possíveis achados nos dados que podem resultar de algum contexto setorial desconhecido pela equipe técnica.

Vários desafios relacionados à qualidade dos dados disponíveis podem surgir nesta fase do projeto de IA, alguns dos quais estão listados abaixo. Esses desafios também estão incluídos no manual de González, Ortiz e Sánchez Ávalos (2020) dirigido à equipe técnica[20], embora aqui sejam abordados da perspectiva do diretor do projeto, que, como responsável final pela ferramenta, deve resolvê-los.

Qualidade e relevância do banco de dados

20 https://publications.iadb.org/es/ia-responsable-manual-tecnico-ciclo-de-vida-de-la-inteligencia-artificial

Graças ao mapeamento de dados realizado na etapa de conceitualização e formulação, o diretor do projeto já conta com uma aproximação dos dados que lhe permite determinar se estão completos, se são relevantes e se possuem a granularidade necessária para a análise. No entanto, durante a exploração dos dados, a equipe técnica pode encontrar desafios adicionais relacionados à qualidade do banco de dados, como:

Estados indesejáveis ou subótimos nos dados coletados. Se os dados apresentarem vieses, a ferramenta treinada neles também os conterá, colocando sua implementação em risco. Para identificar esses vieses, o diretor do projeto solicitará à equipe técnica, com quem já terá debatido as implicações adversas de usar uma ferramenta enviesada, que realize uma análise exploratória dos dados como um todo, levando em consideração, com base no conceito definido previamente de justiça algorítmica, os subgrupos da população que podem vir a ser afetados.

As variáveis para as quais os dados estão disponíveis não são ideais. Pode ocorrer que não estejam disponíveis os dados necessários sobre a variável-alvo exigida pelo projeto, isto é, a que se busca explicar ou prever (Greene, 2003). Nesse caso, deve-se tomar a decisão de trocá-la por uma variável substituta (proxy) para a qual existam as informações necessárias.

Por exemplo, se for necessário medir a experiência de trabalho e esses dados não estiverem presentes no banco, mas a idade e os anos de estudo do indivíduo estiverem disponíveis, uma variável substituta poderá ser uma combinação das duas anteriores. Embora se procure chegar, assim, o mais perto possível da variável-alvo, não é exatamente igual e, consequentemente, seu uso pode afetar a qualidade das previsões. Nesse exemplo, a variável substituta não identifica períodos de inatividade ou desemprego, viagens prolongadas ao exterior ou mesmo licença parental, que é especialmente complexa para as mulheres (dependendo dos regulamentos nacionais referentes à licença de maternidade). Portanto, se a variável ideal não estiver disponível, o diretor do projeto deve informar a equipe técnica da necessidade de desenvolver uma variável substituta, se possível.

Por sua vez, a equipe técnica deve apresentar os resultados dessa variável, explicando claramente suas vantagens e desvantagens. Caberá ao diretor do projeto decidir qual será a variável substituta a ser utilizada ou se o problema deve ser reformulado de acordo com os dados disponíveis.

Informações incompletas sobre a população destinatária

Conforme visto no Item 1.5 deste documento, que trata do mapeamento de dados, pode haver atributos ausentes ou incompletos no banco de dados. Aqui, o principal desafio para o diretor do projeto será aprovar o tratamento a ser dado às referidas observações, com base nas recomendações da equipe técnica. Outras situações problemáticas das quais o diretor deve estar ciente estão resumidas abaixo:

Amostragem probabilística e naturais pertinentes. Se o banco de dados for alimentado por pesquisas ou entrevistas, o diretor do projeto deve saber que esses dados correspondem a um subgrupo da população. Existem dois possíveis extremos de amostragem relacionados à aleatoriedade da seleção da amostra[21]. Uma amostra não aleatória pode ter um viés de participação; portanto, o diretor do projeto deve deixar claro que ela pode não ser representativa da população em geral. Ao analisar o modo de amostragem, o diretor do projeto pode decidir

21 Para mais informações, consulte González, Ortiz e Sánchez Ávalos (2020, p. 18).

modificar a forma em que os dados são captados, o que colocaria o projeto em espera até que os dados satisfizessem os requisitos definidos.

Atributos ausentes ou incompletos. Por meio da análise exploratória realizada, podem ser detectadas situações adicionais de atributos ausentes ou incompletos. As razões para isso podem ser várias: problemas causados pelos responsáveis pela digitação dos dados; falhas do sistema; problemas de transcrição e/ou falta de informações sobre esse atributo para a pessoa em questão. Quando os atributos estiverem ausentes ou incompletos, o diretor do projeto deverá perguntar à equipe técnica sobre a natureza desses erros e se são aleatórios ou não. Dependendo de sua natureza e do número de erros, deve-se tomar a decisão de imputar dados ou eliminar observações dos dados, reduzindo-se o tamanho da amostra (ambos têm vantagens e desvantagens).

Comparação causal

O diretor do projeto deve deixar bem claro que os métodos tradicionais de aprendizado de máquina não estabelecem relações causais entre as variáveis utilizadas (Varian, 2014). Porém, também deve saber que existem técnicas para tentar estabelecer a causalidade em uma análise, o que deverá ser oportunamente comunicado à equipe técnica, para que decida qual é a ferramenta mais adequada para isso. Se uma análise de causalidade for exigida, os dados necessários deverão ser coletados de acordo com a metodologia de causalidade escolhida.

Ao analisar cuidadosamente os dados disponíveis e identificar suas limitações, o diretor do projeto e a equipe técnica podem tomar medidas futuras para melhorar as formas de coletar e armazenar as informações. Por exemplo, se não houver dados sobre a variável-alvo desejada, a forma de coletar as informações pode ser alterada para poder captá-los?

Atividade Uma vez finalizada esta seção, preencha, em conjunto com a equipe técnica, a parte correspondente à **coleta e processamento de dados tanto no perfil do modelo quanto no perfil de dados.**[22]

VER FICHA EM ANEXO 4

O perfil de dados e a seção de coleta e processamento de dados do perfil do modelo serão apresentados a seguir, destacando o papel do diretor do projeto no desenvolvimento dessas ferramentas, que deverão ser preenchidas pela equipe técnica[22].

PERFIL DO MODELO	
Coleta e processamento de dados	
3. Dados de treinamen-to23	Conjunto de dados usado e rotulagem
	Etapas de pré-processamento ou preparação dos dados
	Possíveis vieses e deficiên-cias, dependendo do caso de uso (2)

Papel do Diretor do Projeto

Revisar, junto com a equipe técnica, os resultados da análise exploratória dos dados. Explicar, com base no conhecimento especializado, as anomalias encontradas (se houver) e os resultados da análise, e contextualizar os principais achados.

Solicitar à equipe técnica uma análise exploratória dos dados por subgrupos populacionais. Adotar medidas de mitigação de vieses e equidade algorítmica.

PERFIL DE DADOS
Descrição geral dos dados e motivação
Nome do conjunto de dados usado
Que instituição criou o banco de dados?
Com que finalidade a instituição criou o banco de dados utilizado?

Papel do Diretor do Projeto

Conectar a equipe técnica aos responsáveis pelos dados (internos e externos) para obter informações detalhadas sobre os bancos de dados, o método de coleta e as possíveis deficiências nas informações.

22 Para mais informações sobre ambos os perfis (dados e modelo), consulte González, Ortiz e Sánchez Ávalos (2020).

23 Essa numeração corresponde à ferramenta de perfil do modelo do manual de González, Ortiz e Sánchez Ávalos (2020).

Que mecanismos ou procedimentos foram usados para coletar os dados (por exemplo, pesquisa domiciliar, sensor, software, API)?	Debater com a equipe jurídica as restrições de uso dos dados, considerando-se a forma em que foram captados. Por exemplo: é necessário obter consentimento informado? São informações públicas ou privadas?
Número de indivíduos cujos dados foram coletados	Dependendo do número, debater com a equipe técnica se a amostra pode ser representativa. Se não, como isso afetará o treinamento da ferramenta?
Frequência de captação (semanal, mensal, diária) ou quantidade média de observações por indivíduo. O conjunto de dados será atualizado (por exemplo, adicionando-se novas instâncias ou removendo-se outras)?	
Controles essenciais	
Obter a documentação para cada variável do conjunto de dados. Breve descrição, incluindo nome e tipo, o que representa, como seu valor é medido etc.	
Realizar uma análise exploratória dos dados. Calcular as estatísticas descritivas, identificando a porcentagem de valores ausentes, e determinar a distribuição de cada variável dentro do banco de dados.	Estudar os resultados da análise exploratória dos dados. Certificar-se de que a maturidade do banco de dados seja adequada para solucionar o problema de política pública.
Analisar a cobertura espacial e temporal dos dados.	
Analisar a cobertura de grupos protegidos (sexo, raça, idade etc.).	Comunicar à equipe técnica quais grupos serão considerados protegidos.
Descrever as dimensões relevantes nas quais a amostra de dados pode ser diferente da população; em particular, os vieses de seleção não medidos. Utilizar a literatura relacionada ao tema e informações de especialistas.	Ajudar a determinar as razões por trás das diferenças encontradas com base no conhecimento especializado. Compartilhar essas informações com a equipe técnica.
Identificar “estados indesejáveis” nos dados, como vieses e desigualdades prejudiciais a subgrupos, ou qualquer outro padrão que seja considerado subótimo ou indesejável do ponto de vista da política social.	Explicar, con base en conocimiento experto, aquellos sesgos de los datos que puedan surgir de un estado indeseado de recolección. Definir lo que se considera un estado subóptimo. Discutir con el equipo técnico las posibles medidas de mitigación de sesgos en los grupos protegidos.
Faltam valores? Em caso afirmativo, explicar por que essas informações não estão disponíveis (isso inclui informações removidas intencionalmente). Determinar se os dados ausentes estão associados à variável a ser prevista.	Ajudar a determinar os motivos dos valores ausentes (por exemplo, uma circunstância adversa que não permitiu a devida coleta de dados) por meio do conhecimento especializado.

Teoria

2.2 Desenvolvimento e validação do modelo

Papel do diretor do projeto

Durante a fase correspondente ao desenvolvimento e validação do modelo, o diretor do projeto será responsável por:

- avaliar o modelo com métricas que levem em consideração, além da eficiência e da eficácia, a equidade;
- decidir o modelo a ser implementado, considerando os recursos humanos e financeiros disponíveis na instituição; e
- entender as limitações do modelo e ajustar o plano de execução do projeto com base nos resultados desta etapa.

Esta seção explicará alguns conceitos básicos da fase de treinamento de um modelo, tarefa que cabe à equipe técnica, embora seja o diretor do projeto quem decidirá em última instância qual modelo implementar.[24]

Como um modelo é desenvolvido?
No Item 1.6 deste manual (Análise/ferramenta), o diretor do projeto já deve ter informado a equipe técnica sobre o tipo de ferramenta ou análise baseada em IA que será necessário para solucionar o problema. Embora algumas ferramentas específicas de IA já tenham sido introduzidas nesta fase, a técnica de modelagem precisa dependerá da natureza do problema a ser solucionado e dos dados. Aqui, a equipe técnica precisa testar diferentes opções que, posteriormente, serão apresentadas ao diretor do projeto e à equipe multidisciplinar.

No caso de um modelo de aprendizado supervisionado, quando alimentado com informações sobre as variáveis, os dados devem ser divididos em pelo menos duas grandes seções (idealmente, três):

- ***Dados de treinamento:*** usados para treinar um modelo.
- ***Dados de validação:*** usados para validar o modelo com base no treinamento anterior.
- ***Dados de teste:*** ficam ocultos até depois da seleção do modelo e são usados para confirmar os resultados.

A equipe técnica ficará encarregada de dividir os dados de acordo com as informações disponíveis, mas caberá ao diretor do projeto decidir quais dados (não todos) serão utilizados na fase de treinamento. Além disso, o diretor deve informar a equipe técnica sobre possíveis situações anômalas nos dados decorrentes de determinadas contingências (situações extremas em âmbito nacional, alteração na definição de uma variável etc.), para que as medidas cabíveis sejam tomadas quando os dados forem usados.

Todos os modelos serão treinados com os dados designados, que serão refinados com base nos dados de validação. Por fim, seu desempenho será avaliado com os dados de teste.
É importante que o diretor do projeto tenha em mente que os modelos são simplificações da

24 Para mais informações sobre modelagem e validação, consulte González, Ortiz e Sánchez Ávalos (2020).

realidade. Portanto, eles não serão 100% precisos, nem se espera que sejam. O objetivo é que o modelo de IA tenha um nível de ajuste suficiente e possa prever, por exemplo, resultados futuros, para os quais será inevitável uma certa porcentagem de erros no processo de validação e teste.

A seguir, será descrita a matriz de confusão, uma ferramenta que ajuda a validar o desempenho de um modelo de IA durante seu desenvolvimento, bem como a função do diretor do projeto e da equipe técnica em sua implementação.

- ***Matriz de confusão:*** quando as variáveis de destino são categóricas (sim/não), as métricas de ajuste podem ser obtidas em termos de falsos positivos e falsos negativos, que são obtidos com base no valor previsto em contraste com os dados reais. Por exemplo, para uma classificação binária, temos a seguinte matriz de confusão:

		Real	
		Positivo	**Negativo**
Previsão	**Positivo**	**Verdadeiro positivo**	**Falso positivo**
	Negativo	**Falso negativo**	**Verdadeiro negativo**

A partir desses erros e acertos na previsão, é possível elaborar diferentes métricas de ajuste de um modelo,[25] tarefa que caberá à equipe técnica. Por sua vez, o diretor do projeto deve ficar atento aos níveis de erro de classificação no modelo a ser implementado, de acordo com a definição do problema. Os falsos positivos ocorrem quando o modelo estabelece, por exemplo, que uma pessoa é beneficiária de um programa, quando, na realidade, não deveria ser. Por sua vez, os falsos negativos ocorrem quando, segundo o modelo, uma pessoa não deveria receber benefícios do programa, quando, de fato, recebe. Além das métricas desenvolvidas pela equipe técnica em função da natureza do problema, o diretor do projeto deverá decidir qual tipo de erro o projeto suporta mais. Por exemplo, no caso anterior de concessão de benefícios, levando-se em consideração que os recursos são escassos, o fato de uma pessoa não receber um benefício ao qual tem direito (falso negativo) pode ser um erro mais grave que conceder benefícios a quem não precisa deles (principalmente se for possível fazer uma análise mais profunda dos limites para ser um beneficiário e descobrir que a pessoa está realmente muito perto do limite).

Porém, o fato de os erros produzidos pelo modelo serem aceitáveis também dependerá dos recursos da instituição. Assim, o diretor do projeto, além de estar ciente de sua ocorrência, também deve realizar uma análise de custo-benefício. Prosseguindo com o exemplo anterior, se o orçamento institucional for muito limitado, poderão ser impostos limites à concessão de benefícios quando forem indevidos (falso positivo), o que precisará ser equilibrado com a não concessão quando forem devidos (falso negativo). A realização dessa análise de custo-benefício é fundamental antes de implementar o projeto.

Observe que todos esses erros são cometidos “no laboratório”; por isso, é necessário ter em mente que outros desafios podem surgir durante a implementação. É possível que a ferramenta seja muito difícil de executar ou não muito bem aceita pelos usuários. Nesse caso, pode-se retornar ao Item 1.6 para analisar novamente a necessidade do projeto e ao início do Item 2.2 para avaliar outra ferramenta baseada em IA cuja implementação tenha uma maior probabilidade de sucesso.

25 Consulte González, Ortiz e Sánchez Ávalos, 2020, p. 16.).

Equidade e desempenho diferencial de preditores

Levando-se em consideração o disposto acima, o diretor do projeto deve estudar detalhadamente os vieses e desigualdades que podem surgir no modelo para tomar a decisão, seja para corrigi-los, seja para implementar políticas e medidas de mitigação.

Conforme discutido na seção anterior, podem aparecer vieses nos dados, que, então, poderão ser mitigados. No entanto, com as métricas de erro, pode ocorrer que um modelo favoreça um subgrupo da população. Por exemplo, suponhamos que o modelo de detecção de retinopatia diabética seja mais preciso para homens que para mulheres, um viés que não necessariamente advém dos dados, mas do próprio modelo. O diretor do projeto deve estar ciente desse viés, informando-o ao público e mitigando-o com medidas de políticas públicas. Uma solução possível seria enviar os exames de fundo de olho correspondentes a mulheres a um oftalmologista para análise posterior, enquanto o modelo estiver sendo recalibrado. Isso serviria para garantir uma maior precisão dos resultados desses testes em mulheres e alimentar o modelo com essas análises.

Também pode ocorrer que o objetivo de um projeto seja, na verdade, favorecer um subgrupo específico da população. Tomemos como exemplo o caso de um subsídio que historicamente foi atribuído aos homens, mas que agora visa a beneficiar as mulheres em pelo menos 50%. Aqui, o objetivo da ferramenta de previsão (sensibilidade, precisão e exatidão, entre outros) não corresponde ao objetivo do projeto, ou seja, atingir 50% de mulheres beneficiárias. Por isso, o diretor do projeto é obrigado a informar efetivamente a equipe técnica sobre o objetivo do projeto (que pode ser diferente da própria modelagem) e identificar claramente os subgrupos da população que se pretende beneficiar.

No Item 1.7 deste manual, apontou-se a importância de o diretor do projeto estar ciente das implicações éticas e jurídicas do que foi mencionado acima, mas também do próprio modelo de IA. Agora, com informações mais detalhadas sobre o modelo a ser implantado, o diretor deve retornar a esse item e verificar se outros riscos éticos, vieses e discriminações não surgiram em decorrência do treinamento. No Item 2.1, que trata da coleta e processamento de dados, os vieses que podem surgir já foram debatidos, mas, conforme indicado aqui, o modelo pode originar outros vieses sem que os dados sejam necessariamente enviesados.

Será a equipe técnica que informará à equipe multidisciplinar qual modelo deve ser implementado, cabendo ao diretor do projeto a decisão final, levando-se em conta a natureza do problema, os recursos financeiros e humanos disponíveis na instituição, e os riscos e vieses que a implementação da ferramenta pode gerar.

Atividade No fim desta fase, recomenda-se revisar a seção correspondente ao **desenvolvimento do modelo da ferramenta no perfil do modelo**, em conjunto com a equipe técnica e conforme as funções identificadas.

VER FICHA EM ANEXO 4

A seção de desenvolvimento do modelo será apresentada a seguir, destacando o papel do diretor do projeto na formulação dessa ferramenta, que deve ser preenchida pela equipe técnica. Esse perfil é detalhado em González, Ortiz e Sánchez Ávalos (2020).

PERFIL DO MODELO	
Desenvolvimento do modelo	
4. Modelagem	Algoritmos usados no treinamento, parâmetros adotados ou restrições
5. Métricas de desempenho	Métricas técnicas usadas para selecionar e avaliar modelos
	Análise de custo-benefício do modelo para o caso de uso em questão
	Definição de grupos protegidos e medidas de equidade selecionadas
6. Dados de validação	Conjuntos de dados usados e sua rotulagem
	Etapas de pré-processamento
	Avaliação da adaptação dos dados de validação de acordo com o caso de uso
	Possíveis vieses e deficiências, dependendo do caso de uso
7. Resumo da análise quantitativa	Erro de validação relatado
	Resumo da análise de custo-benefício
	Relatório sobre medidas de equidade para grupos protegidos

Papel do Diretor do Projeto

Realizar a análise de custo-benefício da implementação da ferramenta, perguntando-se se o problema de política pública pode ser solucionado sem o uso de IA.

Definir medidas de equidade nos resultados (subgrupos da população para os quais se pretende garantir a equidade).

Estabelecer, junto com a equipe técnica, limites aceitáveis para os vieses com base no conhecimento especializado. Desenvolver medidas de mitigação com base em políticas públicas, se necessário.

Decidir, junto com a equipe técnica, qual modelo será implementado.

Teoria

2.3 Uso e monitoramento

Papel do diretor do projeto

Durante a fase correspondente ao uso e monitoramento do modelo, o diretor do projeto será responsável por:

- garantir a capacitação dos servidores públicos que interagirão com o modelo de IA, para que a ferramenta seja sustentável ao longo do tempo;
- elaborar um manual do usuário voltado para os funcionários que interagirão com o modelo;
- estabelecer mecanismos de feedback para as pessoas que interagirão com o modelo;
- projetar e implementar processos administrativos simples para corrigir os erros presentes no modelo que afetam os usuários;
- estabelecer sistemas automatizados, ou pelo menos periódicos, de monitoramento do modelo;
- manter um registro dos resultados do modelo, levando em conta as restrições de acesso e segurança;
- implementar as melhorias necessárias do modelo e do processo com base nos achados do monitoramento e da avaliação; e
- alocar os recursos necessários para manter a ferramenta ao longo do tempo.

Uma vez escolhido o modelo mais adequado na etapa anterior, é hora de colocá-lo em prática. Nesta fase, é importante que o diretor do projeto crie um piloto da ferramenta que permita monitorar sua utilização em condições limitadas, antes da implantação em toda a população ou nas entidades envolvidas.

Um dos aspectos cruciais a serem considerados é o treinamento dos servidores públicos que interagirão com o modelo. Isso deve incluir informações sobre seus principais aspectos, objetivo, forma de funcionamento e como seus sistemas de feedback funcionarão. Se as pessoas certas forem incluídas na etapa de conceitualização e formulação, as equipes relacionadas já terão informações sobre o projeto; portanto, sua participação e posterior treinamento ajudará a mitigar o risco de a ferramenta acabar nunca sendo usada. Pode ser necessário criar uma equipe especial de treinamento dentro da instituição. Sugere-se, também, a elaboração de um manual do usuário interno para os funcionários que interagirão com o modelo. Esse manual deve ser uma versão resumida das ferramentas apresentadas aqui. Sugere-se que ele contenha, no mínimo:

- a definição do problema de política pública;
- a definição dos objetivos;
- a definição e justificação da população protegida;
- o regulamento referente ao uso e armazenamento de dados (juntamente com as considerações jurídicas pertinentes a cada país);
- uma explicação simples sobre o funcionamento da ferramenta e sobre as variáveis utilizadas em seu treinamento;

- uma seção de perguntas frequentes, formuladas em linguagem clara, que podem ser feitas pelas pessoas afetadas pela implementação da ferramenta;
- informações de contato para tirar dúvidas e solicitar informações adicionais.

Avaliação da eficácia
Esta etapa contempla a formulação da avaliação do modelo não mais no laboratório (conforme explicado no Item 2.2, relativo ao desenvolvimento e validação), mas no contexto em que será implantado, a fim de medir sua real eficácia durante a implementação do projeto. O objetivo aqui é poder atribuir causalidade à implementação da ferramenta de IA, para a qual existem várias técnicas de avaliação de impacto, inclusive experimentos naturais, experimentos controlados e randomizados, e quase-experimentos (Shadish, Cook e Campbell, 2002). Se possível, recomenda-se privilegiar os métodos de experimentos randomizados e controlados, dado seu alto nível de validade.

Para realizar um experimento randomizado e controlado, é necessário obter o consentimento informado dos participantes e tomar as devidas precauções contra os riscos éticos que podem implicar questões como a experimentação em seres humanos, privar beneficiários em potencial de uma medida por fazerem parte de um grupo de controle e atribuições aleatórias versus por necessidade, bem como outras questões jurídicas (Shadish, Cook e Campbell, 2002). Embora os ensaios controlados e randomizados sejam um dos métodos preferidos para medir o impacto, a ferramenta utilizada no projeto dependerá dos dados disponíveis e das restrições éticas e jurídicas de cada país. Será o diretor do projeto quem decidirá e comunicará à equipe técnica como a eficácia da ferramenta será medida no mundo real.

O sucesso da ferramenta e, portanto, do projeto, será determinado com base no atingimento dos objetivos. Nesse sentido, o propósito de uma avaliação de impacto **não será medir o grau de correspondência do modelo com a realidade**, mas o atingimento dos objetivos. Uma medição adequada do cumprimento das metas estabelecidas exigirá uma linha de referência clara sobre o desempenho do processo atual (sem uma ferramenta de IA), que será definida na etapa de análise exploratória dos dados. Se a ferramenta não atingir o impacto esperado em relação ao estado anterior, a implementação deve ser revisada para verificar se houve falhas, ou retornar à fase de desenvolvimento do modelo para produzir um novo modelo ou reformular o projeto a partir da etapa de conceitualização e formulação.

Degradação do desempenho
O diretor do projeto deve estar ciente de que os resultados de um modelo podem degradar ao longo do tempo por vários motivos:

- a relação entre as variáveis de entrada e saída pode mudar;
- a forma de coletar e armazenar os dados pode mudar;
- é possível que os sistemas de feedback usuário-ferramenta sejam fechados, o que significa que as interações não darão origem à incorporação de novos dados provenientes de situações emergentes que enriquecem o modelo.

Para mitigar esses riscos, será necessário que o diretor do projeto monitore o comportamento das variáveis de entrada e saída, e que, junto com a equipe técnica, atualize as premissas do modelo. Podem ser estabelecidos sistemas automáticos ou com certa periodicidade para monitorar os resultados da ferramenta e o comportamento das métricas de erro e equidade estabelecidas previamente. A periodicidade dependerá da natureza do problema.

Da mesma forma, **deve haver um sistema de feedback da experiência de uso da ferramenta entre os usuários, os beneficiários finais, a equipe técnica e o diretor do projeto**, que será exclusivamente responsável por ele. O objetivo da implantação de um sistema de feedback é dispor de informações em tempo hábil sobre as dificuldades e erros detectados por quem interage com o sistema a fim de corrigi-los o mais rápido possível, seja voltando-se à etapa anterior ou por meio de políticas públicas corretivas.

Nesta fase, é fundamental que todos os recursos necessários (humanos e financeiros) sejam alocados para retreinar o modelo e capacitar as pessoas da instituição que interagirão com ele. Ao aplicar uma solução baseada em IA, é importante que ela seja sustentável ao longo do tempo e se ajuste de acordo com novos dados, tecnologias emergentes e/ou novas interfaces de usuário.

Actividad

No fim desta fase, proceda à revisão, em conjunto com a equipe técnica, da seção de **uso e monitoramento da ferramenta no perfil do modelo**, de acordo com as funções identificadas.

VER FICHA EM ANEXO 4

A seção de uso e monitoramento do perfil do modelo será apresentada a seguir, com destaque ao papel do diretor do projeto no desenvolvimento dessa ferramenta, que deve ser preenchida pela equipe técnica. Esse perfil é detalhado em González, Ortiz e Sánchez Ávalos (2020).

PERFIL DO MODELO	
Uso e monitoramento	
8. Recomendações de monitoramento	Estratégia de monitoramento e melhora em produção
	Estratégias de monitoramento humano das previsões (se aplicável)

Papel do Diretor do Projeto

Definir, em conjunto com a equipe técnica, um programa-piloto ou ensaio randomizado antes de implantar o modelo para toda a população.

Realizar o monitoramento periódico da ferramenta com conjunto com a equipe técnica e os funcionários que interagem com ela.

Teoría

2.4 Prestação de contas

Papel do diretor do projeto

Na fase de prestação de contas, o diretor do projeto será responsável por:

- criar mecanismos de comunicação permanentes e atualizados sobre o funcionamento do sistema de IA, onde deverá utilizar uma linguagem clara;
- prestar contas periódicas sobre o monitoramento e os impactos do sistema; e
- estabelecer um sistema de resposta adequado para lidar com solicitações individuais de resultados da aplicação do sistema de IA.

Vale reiterar que os princípios básicos da administração pública incluem a transparência e o direito de acesso à informação por parte dos cidadãos, que abrangem, também, os projetos de inteligência artificial executados pelo Estado.

Interpretabilidade e explicação das previsões

Entende-se por interpretabilidade ou explicabilidade a possibilidade de o público em geral entender como uma ferramenta de IA funciona (Miller, 2019). A seguir, alguns dos benefícios advindos disso:

- conhecimento de como o problema será solucionado;
- obtenção de licença social;
- detecção de vieses nos algoritmos, e
- depuração e aprimoramento do modelo.

O diretor do projeto é responsável por criar mecanismos de comunicação permanentes e periódicos sobre o funcionamento da ferramenta de IA (problemas, impactos e ajustes ao longo do tempo) para todas as partes interessadas.

A utilização de ferramentas automáticas de tomada de decisão pode ser estranha e confusa para os cidadãos, o que pode gerar resistência à sua implementação, principalmente porque os algoritmos muitas vezes são vistos como "caixas-pretas". Cabe reiterar aqui que a explicabilidade e a transparência algorítmica de um modelo, bem como as informações que podem ser fornecidas sobre seu funcionamento, dependerão de seu nível de opacidade (Burrell, 2016 e Buenadicha et al., 2020), conforme discutido no Item 1.7, que trata das considerações éticas, jurídicas e de governança.

Possíveis vieses e depuração
O diretor do projeto será responsável por fornecer aos interessados informações sobre as limitações e vieses da ferramenta.

Já se observou como o modelo e/ou os dados podem ser enviesados, o que exige que o diretor do projeto implemente medidas para garantir a equidade, caso os vieses não tenham sido corrigidos nas etapas anteriores à execução.

Além disso, para obter e manter a licença social, é importante que o diretor do projeto fique atento aos riscos e considerações éticas decorrentes da implementação do modelo, que os comunique e que indique claramente como pretende mitigá-los.

Explicação das previsões individuais
Aqui, o diretor do projeto deve desenvolver protocolos de resposta para os casos particulares em que as pessoas solicitam informações adicionais sobre o funcionamento e aplicação de um sistema de IA. Por exemplo, se uma pessoa não for considerada como tendo direito a um auxílio do governo, ela precisa poder solicitar informações sobre o motivo dessa decisão automática. Isso exigirá que haja pessoas encarregadas de revisar casos particulares e dar uma resposta satisfatória aos receios dos afetados.

Por sua vez, a equipe técnica deve avaliar a melhor forma de explicar essas previsões individuais.

Rastreabilidade
Julga-se muito importante que o diretor do projeto supervisione a documentação detalhada de todas as decisões tomadas ao longo das diferentes etapas de execução da ferramenta e que essas informações estejam à disposição de todos os interessados, independentemente do nível de opacidade.

O objetivo da rastreabilidade das decisões é mitigar os riscos identificados aqui.

Actividad no fim desta fase, proceda ao preenchimento, em conjunto com a equipe técnica, da seção de **prestação de contas do perfil do modelo**, de acordo com as funções identificadas.

VER FICHA EM ANEXO 4

A seção de prestação de contas do perfil do modelo será apresentada a seguir, destacando o papel do diretor do projeto no desenvolvimento dessa ferramenta, que deve ser preenchida pela equipe técnica. Esse perfil é detalhado em González, Ortiz e Sánchez Ávalos (2020).

PERFIL DO MODELO	
Prestação de contas	
9. (Opcional) Explicabilidade das previsões	Estratégia para explicar previsões específicas (se necessário)
	Estratégia para entender a importância de diferentes atributos
10. Outras considerações éticas, recomendações e advertências	

Papel do Diretor do Projeto

Criar uma política de resposta ou revisão para atender a possíveis solicitações dos cidadãos sobre o resultado da ferramenta.

Debater com a equipe jurídica os riscos que podem surgir na implementação do modelo (proteção de dados pessoais, vieses ou outros) e desenvolver medidas de mitigação e explicabilidade, se necessário.

O QUE ACONTECEU COM O DART NA ETAPA DE EXECUÇÃO?

Coleta e processamento de dados

Inicialmente, os dados utilizados para o modelo incluíam 550 imagens de fundo de olho do Centro de Referência em Saúde (CRS) Cordillera Oriente da comuna de Peñalolén, em Santiago, Chile.

A equipe técnica precisou pré-processar as imagens obtidas, comprimindo-as para padronizar seu tamanho à luz de trabalhos internacionais anteriores na mesma área.

No momento do desenvolvimento do modelo, verificou-se que, nas imagens em que o contraste era insuficiente, a previsão da ferramenta era afetada, pois era mais difícil realizar a segmentação dos vasos sanguíneos na retina, o primeiro passo fundamental da detecção da retinopatia diabética (Arenas, 2012).

O modelo continuou a ser retreinado à medida que mais exames de fundo de olho foram sendo disponibilizados, e as mais recentes técnicas de IA foram incorporadas. Em um estudo observacional publicado em 2021, foram utilizadas 1.123 imagens de fundo de olho (Arenas et al., 2021).

Desenvolvimento e validação do modelo

Na etapa inicial de desenvolvimento, os dados foram divididos em duas partes iguais (275 imagens cada) para criar os conjuntos de treinamento e teste.

O conhecimento especializado de oftalmologistas da Universidade do Chile foi usado para rotular as lesões oculares presentes nas imagens. A detecção da retinopatia diabética consiste em quatro etapas: segmentação dos vasos sanguíneos na retina, localização do disco óptico, detecção de lesões brilhantes e detecção de lesões vermelhas.

Ao analisar os custos do modelo, constatou-se que o investimento inicial implicou um custo extra para o Ministério da Saúde devido à aquisição da licença de utilização do modelo de IA desenvolvido pelo DART. No entanto, o benefício começou a materializar-se quando se passou da ação corretiva para a preventiva, com o efeito que isso tem na saúde das pessoas com diabetes. Em um horizonte de 10 anos, estima-se que os recursos monetários destinados ao tratamento da retinopatia diabética serão reduzidos em aproximadamente 50% (Shokiche, 2013).

Uso e monitoramento

Por meio do Laboratório do Governo do Chile[26], firmou-se uma parceria com o município de Recoleta, o Instituto de Saúde Pública e o Ministério da Saúde para realizar um piloto naquela cidade durante o ano de 2016.

Para utilizar a ferramenta de detecção, criou-se uma interface web que permite que os oftalmologistas interajam com os resultados. Nessa plataforma, os dados pessoais do paciente são

26 O *Laboratorio de Gobierno* é a agência do Estado chileno que desde 2015 se dedica à cocriação de soluções para problemas públicos prioritários e à capacitação para inovar em instituições públicas, a fim de melhorar os serviços prestados aos cidadãos.

inseridos junto com o exame de fundo de olho. Em seguida, os exames são enviados para análise. O sistema automático estuda as imagens e devolve-as à plataforma junto com os resultados. Por fim, os oftalmologistas entram na plataforma e obtêm informações sobre casos prováveis e descartados, para que possam concentrar seus esforços e tempo em casos prováveis e dar um diagnóstico definitivo (Arenas et al., 2015).

O uso da ferramenta exigiu treinamento do pessoal dos centros de saúde, dos tecnólogos médicos que realizam o exame de fundo de olho e que carregam as imagens da ferramenta na nuvem, e dos oftalmologistas dos centros de saúde que recebem os resultados da triagem. Desde 2018, videoconferências síncronas vêm sendo realizadas com grupos de usuários em todo o Chile. Atualmente, tutoriais interativos estão sendo desenvolvidos.

O Ministério da Saúde do Chile adotou a tecnologia como padrão em seus centros de atendimento; atualmente, o DART é usado em mais de 130 deles.

Existe uma relação constante entre os centros de saúde e as autoridades do Ministério da Saúde, que, em 2021, realizou pesquisas para apurar o nível de satisfação dos beneficiários.

Quanto ao funcionamento da própria ferramenta, conduziu-se um estudo observacional para avaliar as métricas de ajuste. Isso foi feito com base em uma amostra de 1.123 imagens de cinco centros de atenção primária do Chile. Treze casos de falsos negativos foram detectados, enquanto a área sob a curva ROC (AUC) foi de 0,915. As medidas de desempenho superaram as exigidas pelo Ministério da Saúde chileno (Arenas et al., 2021).

Prestação de contas

O DART começou como um projeto de tese de José Tomás Arenas, atual CEO da Teledx. Por isso, todas as informações sobre os dados utilizados na primeira etapa (antes de sua implementação) e modelagem inicial encontram-se no repositório da Universidade do Chile.

Após sua implementação no sistema de saúde chileno, vários documentos foram publicados em revistas científicas internacionais, com participação em palestras internacionais, nas quais foram compartilhadas porções significativas das informações em termos de dados, processos e até funcionamento do algoritmo. Embora o modelo subjacente à ferramenta seja extremamente complexo, o valor de sua aplicação em diferentes centros de saúde tem sido demonstrado ao reduzir o tempo de espera para obter os resultados dos exames e os custos associados ao tratamento da doença. Porém, devem ser estudadas detalhadamente as possíveis discriminações do modelo, que podem advir, por exemplo, das condições de realização do exame de fundo de olho, pois podem afetar sua leitura. Isso pode ser mais prevalente em algumas localidades que em outras.

Como o sistema está atualmente implementado sob os auspícios do Ministério da Saúde do Chile, seu funcionamento está sujeito às leis de transparência e proteção de dados pessoais vigentes no país.

CONCLUSÕES

Este manual descreve as etapas necessárias para a correta implementação de um projeto de IA do ponto de vista do responsável pela tomada de decisão referente às políticas públicas e direção do projeto, considerando-se os desafios que surgem durante o ciclo de vida da IA.

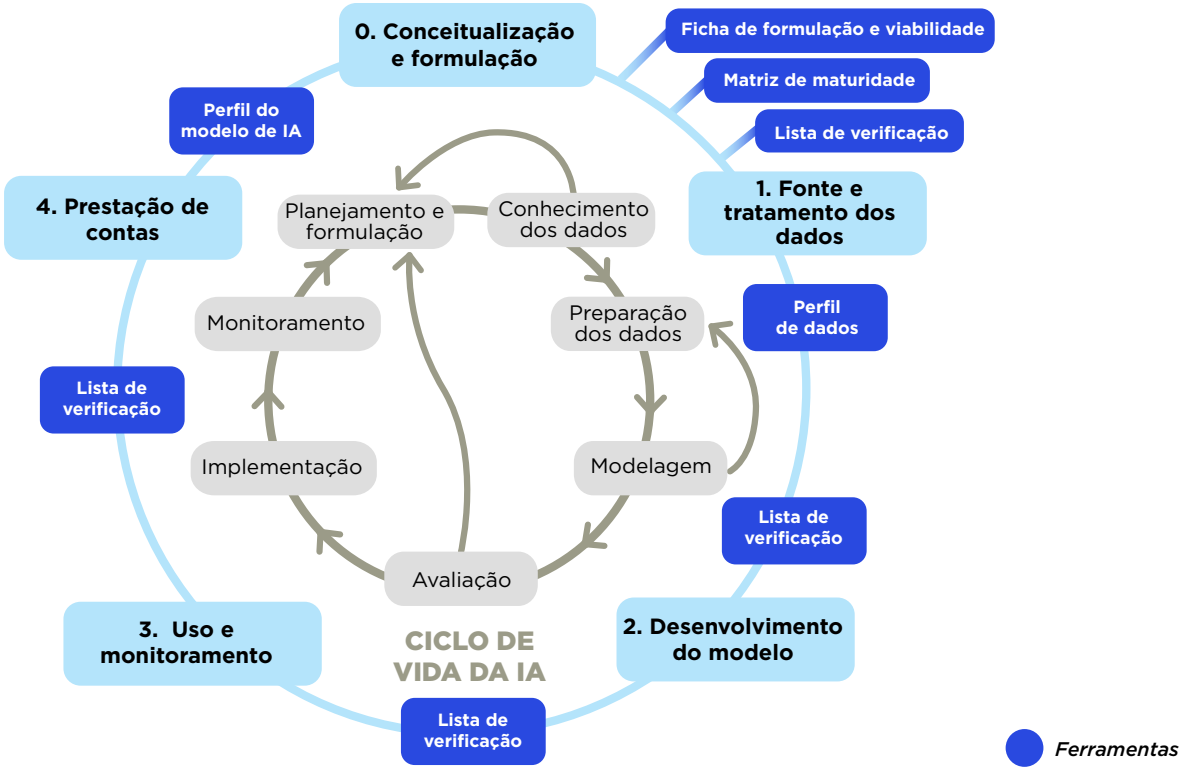
Um projeto de tomada de decisão e/ou suporte à decisão baseado em IA está intimamente ligado ao ciclo de políticas públicas, na medida em que deve resolver problemas reais e significativos que afetam a sociedade.

Os projetos muitas vezes começam com o interesse em querer usar os dados disponíveis na instituição ou uma nova tecnologia. No entanto, cabe reiterar que, para um projeto ter valor social, ele deve surgir da necessidade de solucionar um problema de política pública real.

Embora a ferramenta de IA seja desenvolvida pela equipe técnica, caberá ao diretor do projeto tomar as decisões e liderar sua conceitualização e formulação, a comunicação com a equipe técnica, a análise de viabilidade da ferramenta e sua implementação.

Vale ressaltar aqui a importância do fato de que a criação de um sistema de suporte à decisão e/ou tomada de decisão é resultado de um processo iterativo. É comum que as equipes refaçam as etapas aqui mencionadas, redefinindo até mesmo a conceitualização e formulação do projeto, à medida que novos desafios forem surgindo. Essa etapa é crucial porque obriga o diretor do projeto a fazer perguntas-chave que maximizam as chances de sucesso do projeto. Uma vez definida sua viabilidade, inicia-se a etapa de execução com o processamento dos dados. A Figura 6 contém as ferramentas que este manual oferece para o diretor do projeto, de acordo com o ciclo de vida da IA.

Figura 6. Ciclo de vida da IA e ferramentas do manual



Desde o início, o diretor do projeto precisará ter clareza sobre os riscos em potencial da ferramenta de IA em termos de licença social, implementação e questões éticas. Também precisará se certificar de que conta com os recursos humanos e/ou financeiros para mantê-la e atualizá-la periodicamente. A ferramenta estará em constante ajuste, conforme o contexto for mudando ou novos dados forem sendo disponibilizados. Da mesma forma, é necessário que os funcionários entendam seus benefícios, adotem-na e incorporem-na às suas tarefas para os fins propostos.

Este manual apresenta o ciclo de vida da IA dentro do ciclo de vida das políticas públicas, que não termina no fim do ciclo da IA. Assim, é necessário avaliar a implementação da política não só do ponto de vista da ferramenta, mas também de sua eficácia em termos de resolução do problema identificado na sociedade.

Esperamos que este manual seja útil para os tomadores de decisão que estão pensando em implementar um sistema baseado em IA. Este manual deve ser utilizado em conjunto com o manual técnico de IA responsável (González, Ortiz e Sánchez Ávalos, 2020), que especifica as tarefas e desafios da equipe técnica.

REFERÊNCIAS

Arenas Cavalli, J. T. A. (2012). Diseño y desarrollo de un sistema para la detección automática de retinopatía diabética en imágenes digitales. Disponível em http://repositorio.uchile.cl/handle/2250/104406

Arenas-Cavalli, J. T., Abarca, I., Rojas-Contreras, M., Bernuy, F. e Donoso, R. (2021). Clinical validation of an artificial intelligence-based diabetic retinopathy screening tool for a national health system. Eye, 1-8.

Arenas-Cavalli, J. T., Ríos, S. A., Pola, M. e Donoso, R. (2015). A web-based platform for automated diabetic retinopathy screening. Procedia Computer Science, 60, 557-563.

Bojalil, P. Vela-Treviño, C. (2019). Despuntan las reformas en materia de protección de datos en América Latina. Conhecimento aberto. Disponível em https://blogs.iadb.org/conocimiento-abierto/es/proteccion-de-datos-gdpr-america-latina/. Acessado em 29 de março de 2021.

Buenadicha, C., Galdon, G., Hermosilla, M. P., Loewe, D. e Pombo, C. (2019). La gestión ética de los datos. Por qué importa y cómo hacer un uso justo de los datos en un mundo digital. Disponível em https://publications.iadb.org/publications/spanish/document/La_Gesti%C3%B3n_%C3%89tica_de_los_Datos.pdf.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, 3(1), 2053951715622512.

Cabrol, M., González, N., Pombo, C. e Sánchez, R. (2020). Adopción ética y responsable de la Inteligencia Artificial en América Latina y el Caribe. Disponível em https://publications.iadb.org/publications/spanish/document/fAIr_LAC_Adopci%C3%B3n_%C3%A9tica_y_responsable_de_la_inteligencia_artificial_en_Am%C3%A9rica_Latina_y_el_Caribe_es.pdf

Dastin, J. (2018). Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women. Disponível em https://www.reuters.com/article/us-amazon-com-jobs-automationinsight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-womenidUSKCN1MK08G Acessado em 19 de janeiro de 2021.

Data Futures Partnership (2017). A Path to Social Licence: Guidelines for Trusted Data Use.

Gebru, T., Morgenstern, J., Vecchione, B., Wortman, J., Wallach, H., Daumé, H., & Crawford, K. (2018). Datasheets for Datasets. Obtido em https://arxiv.org/pdf/1803.09010.pdf

Giorgi, S. (2017). How to improve the evaluation of complex systems to better inform policymaking Learning from evaluating Defra's Reward & Recognition Fund. Fellowship Report, julho. Disponível em https://www.cecan.ac.uk/sites/default/files/2018-01/Guidance%20Report%20_%20RRF%20Fellowship%20Final.pdf

Gómez, C., del Pozo, C., Martínez, C. e Martín del Campo, A. (2020). La inteligencia artificial al servicio del bien social en América Latina y el Caribe: Panorámica regional e instantáneas de doce países. Disponível em https://publications.iadb.org/publications/spanish/document/La-inteligencia-artificial--al-servicio-del-bien-social-en-America-Latina-y-el-Caribe-Panor%C3%A1mica-regional-e-instant%-

C3%A1neas-de-doce-paises.pdf

González, F., Ortiz, T. e Sánchez Ávalos, R. (2020). IA Responsable: Manual técnico: Ciclo de vida de la inteligencia artificial. Disponível em https://publications.iadb.org/publications/spanish/document/IA-Responsable-Manual-tecnico-Ciclo-de-vida-de-la-inteligencia-artificial.pdf

Greene, W. H. (2003). Econometric Analysis. India: Pearson Education.

Harini Suresh , J. V. (2019). A Framework for Understanding Unintended Consequences of Machine Learning. MIT. Obtido em https://arxiv.org/pdf/1901.10002.pdf

Heaven, W.D. (2020). Google's medical AI was super accurate in a lab. Real life was a different story. Disponível em https://www.technologyreview.com/2020/04/27/1000658/google-medical-ai-accurate-lab-real-life-clinic-covid-diabetes-retina-disease/ Acessado em 29 de janeiro de 2021.

Hewitt, T. (2019). Canada is a leader in AI research – but that's not enough. Disponível em https://www.theglobeandmail.com/opinion/article-canada-is-a-leader-in-ai-and-we-must-lead-in-research--on-how-we-want/ Acessado em 19 de janeiro de 2021.

Hojman Cano, I. (2014). El mercado de las especialidades médicas de anestesiología y oftalmología en Chile. Disponível em http://repositorio.uchile.cl/handle/2250/130678

Miller, T. (2019). Explanation in Artificial Intelligence: Insights from the Social Sciences. Artif. Intell., 267, págs. 1-38.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Gebru, T. (2019). Model Cards for Model Reporting. Disponível em https://arxiv.org/abs/1810.03993

Morrison, J. (2014). The Social License. How to keep your organization legitimate. Palgrave Macmillan, ISBN: 9781137370716.

OECD. (2019). Artificial Intelligence in Society. Paris: OECD Publishing.

OCDE (no prelo). OECD Framework for the Classification of AI Systems. Paris: OECD Publishing.

Oliver, D. (1994). Law, politics and public accountability. The search for a new equilibrium. Public Law, 238-238.

Red Iberoamericana de Protección de datos personales (2017). Estándares de protección de datos personales. Disponível em https://www.redipd.org/es/documentos/estandares-iberoamericanos

Sangüesa, R. (2018). Inteligencia artificial y transparencia algorítmica: "It's complicated". BiD: textos universitaris de biblioteconomia i documentació, No. 41, dezembro. <http://bid.ub.edu/es/41/sanguesa.htm>. Disponível em http://bid.ub.edu/es/41/sanguesa.htm#:~:text=La%20transparencia%20de%20datos%20y,vital%20de%20quien%20reclama%20esta Acessado em 29 de janeiro de 2021.

Sapunar, J. (2016). Epidemiología de la diabetes mellitus en Chile. Revista Médica Clínica Las Condes, 27(2), 146-151.

Shadish, W. R., Cook, T. D., e Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin. ISBN: 0395615569. Disponível em https://www.alnap.org/system/files/content/resource/files/main/147.pdf

Shokiche Vega, D.A. (2013). Estudio del problema de salud pública asociado a la patología oftalmológica retinopatía diabética en Chile y dimensionamiento del potencial impacto de detección basado en tecnología para abordar este problema.

Varian, Hal R. (2014). Big Data: New Tricks for Econometrics. Journal of Economic Perspectives, 28 (2): 3-28.

ANEXOS

Anexo 1. Ficha de formulação e viabilidade do projeto

Anexo 2. Matriz de maturidade dos dados

Anexo 3. Lista de verificação para o diretor do projeto

Anexo 4. Perfil de dados

Anexo 5. Perfil do modelo

Anexo 1. Ficha de formulação e viabilidade do projeto

1 Nome do projeto:

2 Nome da organização:

3 Definição do problema:

Descreva a(s) população(ões) afetada(s). Quem é afetado pelo problema? (Certos tipos de pessoa, organização, bairro, ambiente).

Quantas pessoas/organizações/locais/etc. são afetados e em que medida? (Por exemplo, tempo médio de espera por uma cirurgia, número de alunos que abandonam a escola, custos da evasão fiscal etc.).

Por que resolver esse problema é uma prioridade para a instituição?

Você conhece algum caso de uso de IA semelhante que tenha sido implementado antes? Qual?

4 Análise de pré-viabilidade

Está dentro da nossa alçada adotar medidas para solucionar o problema? Será necessário firmar parcerias com outros órgãos públicos? Os recursos humanos e financeiros necessários para realizar o projeto estão disponíveis?

Há dados relevantes (suficientes para mudar a forma atual de responder ao problema)? É possível ter acesso a esses dados?

Quais são os riscos do projeto (éticos, licença social, implementação etc.)?

5 Definição dos objetivos

Os objetivos geralmente são expressos em termos de melhoria, maximização, aumento ou diminuição, mitigação e/ou redução de um resultado. O objetivo deve ser mensurável, o que requer o estabelecimento de uma métrica ou indicador que reflita o progresso. Atingir o objetivo deve ajudar a resolver o problema. A solução técnica (por exemplo, um modelo preditivo) não é o objetivo.

Restrições típicas estão relacionadas a orçamento, falta de capital humano, restrições jurídicas, vontade política e licença social.

Lembre-se de que, quando há objetivos concorrentes, pode ser necessário sacrificar algo para obter um benefício.

	Objetivo	**Limitações**
1		
2		
3		

6 Descrição das ações:

As ações são atividades que as instituições realizam ou podem realizar a respeito de um problema específico, como os programas específicos que executam de acordo com sua missão para com a sociedade ou seus processos operacionais habituais (contratações, atendimento a usuários, pagamento de salários etc.).

Essas ações podem ser aprimoradas quando a instituição dispõe das informações geradas pelo projeto de ciência de dados. Também devem ter uma conexão com os resultados gerados pelo sistema de IA e ajudar a atingir os objetivos definidos (seção anterior).

	Preencher		
	Ação 1	**Ação 2**	**Ação 3**
Ação Por exemplo: oficina de sexualidade responsável para alunos de 13 anos/ distribuição de métodos contraceptivos na enfermaria da escola. Observação: cada ação deve ser explicada em uma caixa separada.			
Quem executa a ação? Por exemplo, a equipe de psicologia e psicopedagogia de cada estabelecimento e/ ou a equipe de enfermagem da escola.			
Para quem ou o que a ação está sendo executada? Por exemplo, alunos dos estabelecimentos que estão atualmente no 7º ano e/ou o corpo discente em geral.			
Com que frequência é tomada a decisão de realizar a ação? Por exemplo, anualmente/mensalmente.			
Quais canais estão sendo utilizados ou podem ser utilizados para realizar a ação? Por exemplo, de forma presencial.			
Outras informações úteis sobre a ação			

7 Mapeamento de dados

Para que a instituição atinja seu objetivo, precisa haver uma conexão entre os dados e as ações que ela pretende apoiar. Projetos típicos de IA usam dados administrativos como fonte primária, aprimorando-a com outras fontes de dados de domínio público (censo, outros dados abertos etc.). A parceria com o setor privado ou organizações sem fins lucrativos pode ajudar a obter dados ausentes internamente.

Que dados estão disponíveis internamente?

	Preencher		
	Fonte de dados 1	**Fonte de dados 2**	**Fonte de dados 3**
Nome Por exemplo, sistema de alta hospitalar.			
O que contém? Descreva os atributos da forma mais detalhada possível (por exemplo, registros de internação e alta hospitalar em todo o país, com dados sociodemográficos do paciente, diagnóstico, dias de internação, tipo de plano de saúde, informações do médico).			
Qual é o nível de granularidade? Por exemplo, transação, pessoa, organização, localização.			
Com que frequência as informações são coletadas e/ou atualizadas depois de captadas? Por exemplo, em tempo real, diariamente, semanalmente, mensalmente, anualmente, ocasionalmente.			
Há identificadores exclusivos e confiáveis que podem ser vinculados a outras fontes de dados? Por exemplo, RUN, SSN, DNI, dependendo do país.			
Quem é o responsável pelos dados? Por exemplo, o departamento de registros hospitalares.			
Como são armazenados? Por exemplo, em um banco de dados, PDF, Excel, SPSS.			
Comentários adicionais			

Que dados podem ser obtidos de fontes externas, privadas ou públicas?

	Preencher		
	Fonte de dados 1	**Fonte de dados 2**	**Fonte de dados 3**
Nome Por exemplo, registro da qualidade do ar.			
O que contém? Descreva os atributos da forma mais detalhada possível (por exemplo, concentração de poluentes no ar, como material particulado de vários tamanhos).			
Qual é o nível de granularidade? Por exemplo, estação de monitoramento geolocalizada por hora.			
Com que frequência as informações são coletadas e/ou atualizadas depois de captadas? Por exemplo, diariamente.			
Há identificadores exclusivos e confiáveis que podem ser vinculados a outras fontes de dados? Por exemplo, código da estação de monitoramento.			
Quem é o responsável pelos dados? Ministério do Meio Ambiente.			
São necessários acordos legais para a troca e/ou acesso às informações?			
Como são armazenados? Por exemplo, banco de dados que pode ser baixado por meio de uma API em um portal de dados aberto.			
Comentários adicionais			

Em um mundo ideal, existem dados adicionais relevantes relacionados a esse problema que você gostaria de obter? (Pesquisas, CFTV, registros telefônicos, DNA, faixa de frequência ou granularidade para dados atualmente disponíveis etc.)

8 Análise/Ferramenta

Projetos típicos de IA incluem uma combinação de várias análises, dependendo das necessidades e particularidades de cada projeto. As análises são ferramentas, não o objetivo do projeto.

Escolha a(s) análise(s) adequada(s) para o problema certo.

- As análises ou ferramentas escolhidas devem melhorar as ações atuais em resposta ao problema.
- Você deve testar a(s) análise(s) e o processo de validação deve corresponder ao seu objetivo.

	Preencher		
	Análise/ ferramenta 1	**Análise/ ferramenta 2**	**Análise/ ferramenta 3**
Tipo de análise/ ferramenta Por exemplo, descrição, previsão, detecção, mudança de comportamento.			
Finalidade da análise Por exemplo, compreender o comportamento histórico das pessoas; estimar o risco de doença de um paciente; identificar ações que reduziriam a sobrepesca.			
Para que tipos de ação as informações geradas a partir dessa análise serão utilizadas? Por exemplo, inspeção de embarcações de pesca industrial e artesanal.			
Como a análise será validada com base nos dados existentes? Por exemplo, usando-se dados históricos, conduzindo-se um estudo randomizado e controlado etc.			

9 Considerações éticas e jurídicas

Proporcionalidade	**Você acha que um sistema de ciência de dados/IA é o meio certo para solucionar o problema? Por quê? Já avaliou alternativas?**	
	Que impactos negativos o seu projeto pode ter? *Estude casos de uso semelhantes identificados na seção "Definição do problema".*	
Licença social	**Você acha que a população aceitará o uso dos dados propostos para solucionar o problema? Por quê??**	
	Se a população do projeto tomar ciência do uso da IA para os propósitos declarados, ela o aprovará? Por quê?	
	Você identificou a justificativa ou base jurídica para trabalhar com esses dados?	
	Você identificou os regulamentos que podem afetar o projeto?	
	São necessários mecanismos (de acesso, eliminação ou retificação, por exemplo) para garantir a qualidade dos dados pessoais?	
Transparência	**Quais partes interessadas devem estar cientes do projeto?** *As partes interessadas geralmente incluem formuladores de políticas, trabalhadores da linha de frente, organizações da sociedade civil, órgãos públicos, pessoas que serão afetadas pelas ações etc. Cite organizações e/ou tipos específicos de pessoa.*	
	Você considerou um mecanismo para que as partes interessadas entrem em contato com a instituição para obter informações sobre o projeto?	
	Será necessário explicar os mecanismos de tomada de decisão ou as análises a serem implementadas? Por quê?	

Discrimi-nação/ equi-dade	**Que desigualdades estruturais existem no processo ou ambiente no qual o projeto está inserido?**	
	Existem grupos específicos (vulneráveis) para os quais se busca garantir a equidade nos resultados ou a proteção de seus direitos? *Por exemplo, por gênero, idade, localização, classe social, escolaridade, origem urbana ou rural, etnia.*	
	Que vieses você acha que os dados podem ter?	
Prestação de contas	**Caso seja apresentada uma solicitação de informações sobre o projeto, quem é responsável por responder?**	
	Quem é responsável por quaisquer erros cometidos pelo sistema?	
	Existe uma previsão de mecanismos de monitoramento, controle e avaliação? Como e com que frequência eles serão documentados?	
	Foram planejados mecanismos de treinamento para que a equipe responsável compreenda as responsabilidades e obrigações jurídicas e éticas que o projeto envolve?	

10 Formação da equipe

Geralmente, os projetos de inteligência artificial requerem a articulação de vários profissionais de uma mesma entidade pública, mas também de outras organizações relacionadas. Eles exigem a participação dos responsáveis pelos dados, pela infraestrutura de TI e pelo problema ou processo em questão, além de especialistas em análise de dados, área jurídica e comunicação. Adicione as linhas necessárias na tabela a seguir.

Organização/ departamento	**Descrição da participação desejada**	**Nome/função da contraparte**
Departamento de TI	Fornecer a infraestrutura de dados	Chefe do Departamento de TI
Órgão de estatística	Fornecer os dados populacionais	Chefe do Departamento de Estatística

Esta ficha foi desenvolvida originalmente pelo Center for Data Science and Public Policy da Universidade de Chicago. Para mais informações sobre os nossos programas e trabalho, acesse http://datasciencepublicpolicy.org ou escreva a info@datascienceforsocialgood.org.

Esta versão da folha de trabalho foi atualizada por meio de uma colaboração entre o GobLab UAI, Carnegie Mellon University e o Instituto Tecnológico de Monterrey.

O GobLab UAI é o laboratório de inovação da Escola de Governança da Universidade Adolfo Ibáñez. Sua missão é contribuir para a inovação em políticas públicas em prol da sociedade. Ele colabora com órgãos públicos, organizações da sociedade civil e pesquisadores para criar políticas públicas mais eficazes, eficientes e equitativas por meio da ciência de dados. Para mais informações, acesse https://goblab.uai.cl ou envie um e-mail a goblab@uai.cl.

Atribuição ShareAlike (CC BY-SA)

Anexo 2. Matriz de maturidade dos dados

Matriz de maturidade dos dados[27]

Categoria	Áreas	Deficiente	Essencial	Intermediário	Avançado
Como as informações são armazenadas	**Acesso**	Acessados apenas a partir do aplicativo que coleta os dados.	Os dados podem ser acessados, mas com um software especial e específico.	Os dados estão em formatos acessíveis, como CSV, JSON, XML ou um banco de dados acessível remotamente.	Os dados estão em formatos acessíveis e podem ser acessados por meio de uma API.
	Armazenamento	Papel.	PDF ou imagens.	Arquivos de texto.	Bancos de dados.
	Integração	Os dados ficam apenas no sistema onde as informações são coletadas.	Os dados são ocasionalmente exportados e integrados de forma ad hoc.	Data center único com atualização automática.	Os dados internos e externos são integrados ao banco de dados.
Que informações são coletadas	**Relevância**	Os dados são irrelevantes para o problema. Por exemplo, você pode querer ver a probabilidade de conclusão de cursos superiores, mas não tem dados sobre quem se formou.	Alguns dos dados são relevantes, mas são insuficientes porque faltam componentes-chave. Por exemplo, voltando ao caso anterior, só se conta com a presença nas aulas.	Há informações úteis e relevantes, mas não estão completas. Por exemplo, no caso ao lado, há dados demográficos e notas anuais, mas não há informações de apoio extracurricular.	Todas as informações relevantes existem e são suficientes para resolver o problema sem exigir transformações consideráveis.
	Qualidade	Linhas ausentes (observações).	Colunas ausentes (variáveis-chave de certas observações).	Todos os dados estão presentes; há apenas pequenos erros, como, por exemplo, de digitação.	Não há problemas de falta de dados ou erros de digitação; as bases estão limpas.
	Frequência	Apenas uma vez.	Anual.	Frequente.	Tempo real.
	Granularidade	Acrescentado na camada municipal.	Na camada do quarteirão ou código postal.	Na camada individual (pessoa ou endereço).	Detalhes na camada de eventos/marcos.
	História	Não há um histórico; dados antigos são apagados.	Os dados históricos são salvos, mas as atualizações são substituídas.	As informações históricas são salvas, ocupando uma marca de tempo.	Todos os dados são mantidos e relacionados aos dados anteriores com um modelo integrado.
Outros	Privacidade	Não há uma política de privacidade.	A política não permite o uso de nenhum dado.	A aprovação ad hoc é necessária para usar os dados.	O acesso aos dados é definido e sua privacidade é controlada para preservar a privacidade dos indivíduos.
	Documentação	Não há uma documentação digital nem metadados. Os códigos das variáveis não são documentados.	Define-se um dicionário de dados que explica as variáveis e categorias.	Existe um dicionário de dados com disponibilidade de metadados.	Existe um dicionário de dados, metadados e, adicionalmente, as suposições, vieses e dados que não estão sendo obtidos.

27 Atribuição ShareAlike (CC BY-SA). Adaptada do Data Maturity Framework da U. de Chicago http://dsapp.uchicago.edu/resources/datamaturity/

Categoria	Áreas	Deficiente	Essencial	Intermediário	Avançado
Como as informações são armazenadas	Acesso				
	Armazenamento				
	Integração				
Que informações são coletadas	Relevância				
	Qualidade				
	Frequência				
	Granularidade				
	História				
Outros	Privacidade				
	Documentação			.	

Anexo 3. Lista de verificação para o diretor do projeto

A lista de verificação a seguir contém as principais tarefas e funções do diretor de projeto (responsável pela tomada de decisão) ao longo do ciclo de vida da IA.

Conceitualização e formulação de políticas públicas

- [ ] Definir claramente o problema de política pública que se busca resolver, identificando e quantificando os grupos de pessoas afetadas, e determinando seu impacto no orçamento. A definição do problema deve ser facilmente compreendida por uma pessoa de fora da instituição.
- [ ] Entrar em contato com as pessoas da sua instituição encarregadas de abordar o problema para entender como isso está sendo feito atualmente. Que informações elas podem fornecer sobre como aprimorar o sistema de resposta?
- [ ] Pesquisar como outros órgãos (nacionais ou estrangeiros) com um problema semelhante têm implementado soluções baseadas em IA. O ideal seria contatá-los para saber mais sobre os desafios e dificuldades com os quais se depararam ao longo do caminho.
- [ ] Conversar com a diretoria da sua instituição sobre a prioridade de solucionar o problema e obter adesão ao projeto no mais alto escalão. Se você fizer parte da diretoria, documente como essa prioridade é refletida nos planos estratégicos da instituição.
- [ ] Confirmar se a lista de verificação da equipe técnica referente a esta fase foi concluída.

Ciclo de vida

Coleta e processamento de dados

- [ ] Contextualizar os achados da equipe técnica na etapa exploratória de dados.
- [ ] Fornecer o histórico do processo de coleta de dados e o significado das variáveis, conectando a equipe técnica aos indivíduos da organização necessários para uma melhor compreensão dos dados.
- [ ] Validar a formulação e utilidade da variável-alvo.
- [ ] Tomar decisões sobre imputações de dados ou eliminação de variáveis, ponderando e registrando o que se obtém em contraste com o que se perde (trade-offs).
- [ ] Propor melhorias nos processos de governança de dados do projeto que possam contribuir para o atingimento dos objetivos e facilitar a implementação do modelo.
- [ ] Ajustar a definição do problema, os objetivos e o plano de execução do projeto com base nos resultados desta etapa.
- [ ] Confirmar se a lista de verificação da equipe técnica referente a esta fase foi concluída.

Desenvolvimento de modelo

- [ ] Avaliar o modelo com métricas que levem em consideração, além da eficiência e da eficácia, a equidade.
- [ ] Decidir o modelo a ser implementado, considerando os recursos humanos e financeiros disponíveis na instituição.
- [ ] Entender as limitações do modelo e ajustar o plano de execução do projeto com base nos resultados desta etapa.
- [ ] Confirmar se a lista de verificação da equipe técnica referente a esta fase foi concluída.

Uso e monitoramento

- [ ] Garantir a capacitação dos servidores públicos que interagirão com o modelo de IA, para que a ferramenta seja sustentável ao longo do tempo.
- [ ] Elaborar um manual do usuário voltado para os funcionários que interagirão com o modelo.
- [ ] Estabelecer mecanismos de feedback para as pessoas que interagirão com o modelo.
- [ ] Projetar e implementar processos administrativos simples para corrigir os erros presentes no modelo que afetam os usuários.
- [ ] Estabelecer sistemas automatizados, ou pelo menos periódicos, de monitoramento do modelo.
- [ ] Manter um registro dos resultados do modelo, levando em conta as restrições de acesso e segurança.
- [ ] Implementar as melhorias necessárias do modelo e do processo com base nos achados do monitoramento e da avaliação.
- [ ] Alocar os recursos necessários para manter a ferramenta ao longo do tempo.
- [ ] Confirmar se a lista de verificação da equipe técnica referente a esta fase foi concluída.

Prestação de contas

- [ ] Criar mecanismos de comunicação permanentes e atualizados sobre o funcionamento do sistema de IA, onde deverá utilizar uma linguagem clara.
- [ ] Prestar contas periódicas sobre o monitoramento e impactos do sistema.
- [ ] Estabelecer um sistema de resposta adequado para lidar com solicitações individuais de resultados da aplicação do sistema de IA.
- [ ] Confirmar se a lista de verificação da equipe técnica referente a esta fase foi concluída.

Anexo 4. Perfil de dados

O perfil de dados é uma análise exploratória que fornece informações para avaliar a qualidade, completude, temporalidade, consistência e possíveis vieses do conjunto de dados que será usado para treinar um modelo de aprendizado de máquina (Gebru et al., 2018). Para que a equipe técnica gere um perfil de dados adequado, o diretor do projeto deve garantir o acesso aos conjuntos de dados (se forem internos) ou coordenar sua obtenção com organizações externas, juntamente com a documentação de cada variável dentro deles. Da mesma forma, poderá colaborar na identificação de dimensões importantes para as quais a amostra de dados pode ser diferente da população em geral, bem como na identificação de estados indesejados, que podem incluir vieses, desigualdades prejudiciais para determinados subgrupos e/ou qualquer outro padrão considerado subótimo ou indesejável do ponto de vista da política social. Por fim, pode ajudar a identificar os motivos da falta de informações e decidir, junto com a equipe técnica, como preencher essa lacuna.

	Papel do Diretor do Projeto
Descrição geral dos dados e motivação	Conectar a equipe técnica aos responsáveis pelos dados (internos e externos) para obter informações detalhadas sobre os bancos de dados, o método de coleta e as possíveis deficiências nas informações.
Nome do conjunto de dados usado	
Que instituição criou o banco de dados?	
Com que finalidade?	
Que mecanismos ou procedimentos foram usados para coletar os dados (por exemplo, pesquisa domiciliar, sensor, software, API)?	Debater com a equipe jurídica as restrições de uso dos dados, considerando-se a forma em que foram captados. Por exemplo: é necessário obter consentimento informado? São informações públicas ou privadas?
Tamanho da amostra	Dependendo do número, debater com a equipe técnica se a amostra pode ser representativa. Se não, como isso afetará o treinamento da ferramenta?
Frequência de captação (semanal, mensal, diária) ou quantidade média de observações por indivíduo. O conjunto de dados será atualizado (por exemplo, adicionando-se novas instâncias ou removendo-se outras)?	

Papel do Diretor do Projeto

Controles essenciais	
Obter a documentação para cada variável do conjunto de dados. Breve descrição, incluindo nome e tipo, o que representa, como seu valor é medido etc.	
Realizar uma análise exploratória dos dados. Calcular as estatísticas descritivas, identificando a porcentagem de valores ausentes, e determinar a distribuição de cada variável dentro do banco de dados.	Estudar os resultados de uma análise exploratória dos dados. Certifique-se de que o vencimento do banco de dados é adequado para resolver o problema de política pública.
Analisar a cobertura espacial e temporal dos dados.	
Analisar a cobertura de grupos protegidos (sexo, raça, idade etc.).	Comunicar à equipe técnica quais grupos serão considerados protegidos.
Descrever as dimensões relevantes nas quais a amostra de dados pode ser diferente da população; em particular, os vieses de seleção não medidos. Utilizar a literatura relacionada ao tema e informações de especialistas.	Explicar à equipe técnica, com base no conhecimento especializado, as diferenças encontradas.
Identificar possíveis "estados indesejáveis" nos dados, como vieses e desigualdades prejudiciais a subgrupos, ou qualquer outro padrão que seja considerado subótimo ou indesejável do ponto de vista da política social.	Explicar, com base no conhecimento especializado, os vieses dos dados que podem surgir de um estado indesejado de coleta. Definir o que é considerado um estado subótimo. Debater com a equipe técnica as possíveis medidas para mitigar vieses nos grupos protegidos.
Faltam valores? Em caso afirmativo, explicar por que essas informações não estão disponíveis (isso inclui informações removidas intencionalmente). Identificar os motivos da ausência desses dados e determinar se os dados ausentes estão associados à variável a ser prevista.	Explicar, com base no conhecimento especializado, os valores ausentes (por exemplo, uma circunstância adversa que não permitiu o levantamento correto dos dados).

Anexo 5. Perfil do modelo

O perfil do modelo é uma ferramenta que sintetiza as principais características de um sistema de tomada de decisão e/ou suporte à decisão baseado em aprendizado de máquina, destacando suas principais premissas e características, bem como as medidas de mitigação implementadas (Mitchell et al., 2019). O diretor do projeto poderá comprovar que a conceitualização e formulação de políticas públicas, bem como os casos de uso, correspondem ao problema que está sendo abordado, sendo necessário, para isso, verificar o histórico, a população destinatária, o horizonte de tempo e os atores e sistemas envolvidos. Também é necessário garantir que os casos de uso se encaixem nas situações em que se espera que o sistema de suporte à decisão seja usado, levando-se em conta inclusive os usos não considerados e as advertências relacionadas. Por fim, é necessário validar os grupos protegidos e possíveis vieses. O perfil do modelo deve ser preenchido em conjunto pelo diretor do projeto e pela equipe técnica.

Conceitualização e formulação de políticas públicas		**Papel do Diretor do Projeto**
		Responsável por toda a fase de conceitualização e formulação do projeto
1. Informações básicas	Pessoas que desenvolveram o modelo, data, versão, tipo	
2. Casos de uso	Antecedentes	Definir corretamente o problema de política pública que a instituição busca resolver.
	População destinatária e horizonte de tempo das previsões	Identificar corretamente a população afetada. Definir de forma concisa e clara os objetivos que devem ser
	Atores e componentes que interagirão com os resultados	Entrar em contato com as pessoas que atualmente respondem à situação para determinar como elas procederam e como o processo pode ser aprimorado com uma ferramenta de IA. Levar em consideração as particularidades da ação que podem comprometer
	Casos de uso considerados durante o desenvolvimento	Junto com a equipe técnica, identificar os diretores de projetos que já resolveram ou estão tentando resolver problemas semelhantes com IA, e entrar em contato com eles para familiarizar-se com suas experiências.
	Usos não considerados e advertências relacionadas	Um conjunto com a equipe técnica, identificar casos específicos de utilização incorreta da ferramenta de IA, se houver, e elaborar medidas de mitigação para prevenir sua ocorrência.
	Definição de grupos protegidos	Definir, com base no conhecimento especializado, quais grupos serão considerados protegidos e para quais se deseja garantir a equidade nos res

Coleta e processamento de dados		
3. Dados de treinamento	Conjunto de dados usado e sua rotulagem	
	Etapas de pré-processamento ou preparação de dados	Revisar, junto com a equipe técnica, os resultados da análise exploratória dos dados; explicar as anomalias encontradas (se houver) e os resultados da análise com base no conhecimento especializado; e contextualizar os principais achados.
	Possíveis vieses e deficiências, dependendo do caso de uso (consulte o Item 2 desta tabela)	Solicitar à equipe técnica uma análise exploratória dos dados por subgrupos da população. Adotar medidas de mitigação de vieses e equidade algorítmica.
Desenvolvimento do modelo		
4. Modelagem	Algoritmos usados para treinar, parâmetros adotados ou restrições	
5. Métricas de desempenho	Métricas técnicas usadas para selecionar e avaliar modelos	
	Análise de custo-benefício do modelo para o caso de uso em questão	Realizar a análise de custo-benefício da implementação da ferramenta e determinar se o problema de política pública pode ser solucionado sem o uso de IA.
	Definição de grupos protegidos e medidas de equidade selecionadas	Definir medidas de equidade nos resultados (subgrupos da população para os quais se pretende garantir a equidade).
6. Dados de validação	Conjuntos de dados usados e sua rotulagem	
	Etapas de pré-processamento	
	Avaliação da adaptação dos dados de validação de acordo com o caso de uso (consulte o Item 2 desta tabela)	
	Possíveis vieses e deficiências, dependendo do caso de uso (consulte o Item 2 desta tabela)	Estabelecer, junto com a equipe técnica, limites aceitáveis para os vieses com base no conhecimento especializado. Desenvolver medidas de mitigação com base em políticas públicas, se necessário.

7. Resumo da análise quantitativa	Erro de validação relatado	
	Resumo da análise de custo-benefício	
	Relatório sobre medidas de equidade para grupos protegidos	Decidir, junto com a equipe técnica, qual modelo será implementado.
Uso e monitoramento		
8. Recomendações de monitoramento	Estratégia de monitoramento e melhora em produção	Definir, em conjunto com a equipe técnica, um programa-piloto ou ensaio randomizado antes de implantar o modelo para toda a população.
	Estratégias de monitoramento humano das previsões (se aplicável)	Realizar o monitoramento periódico da ferramenta com conjunto com a equipe técnica e os funcionários que interagem com ela.
Prestação de contas		
9. (Opcional) Explicabilidade das previsões	Estratégia para explicar previsões específicas (se necessário)	Criar uma política de resposta ou revisão para atender a possíveis solicitações dos cidadãos sobre o resultado da ferramenta.
	Estratégia para entender a importância de diferentes atributos	
10. Outras considerações éticas, recomendações e advertências		Debater com a equipe jurídica os riscos que podem surgir na implementação do modelo (proteção de dados pessoais, vieses ou outros) e desenvolver medidas de miti

BID

Melhorando vidas