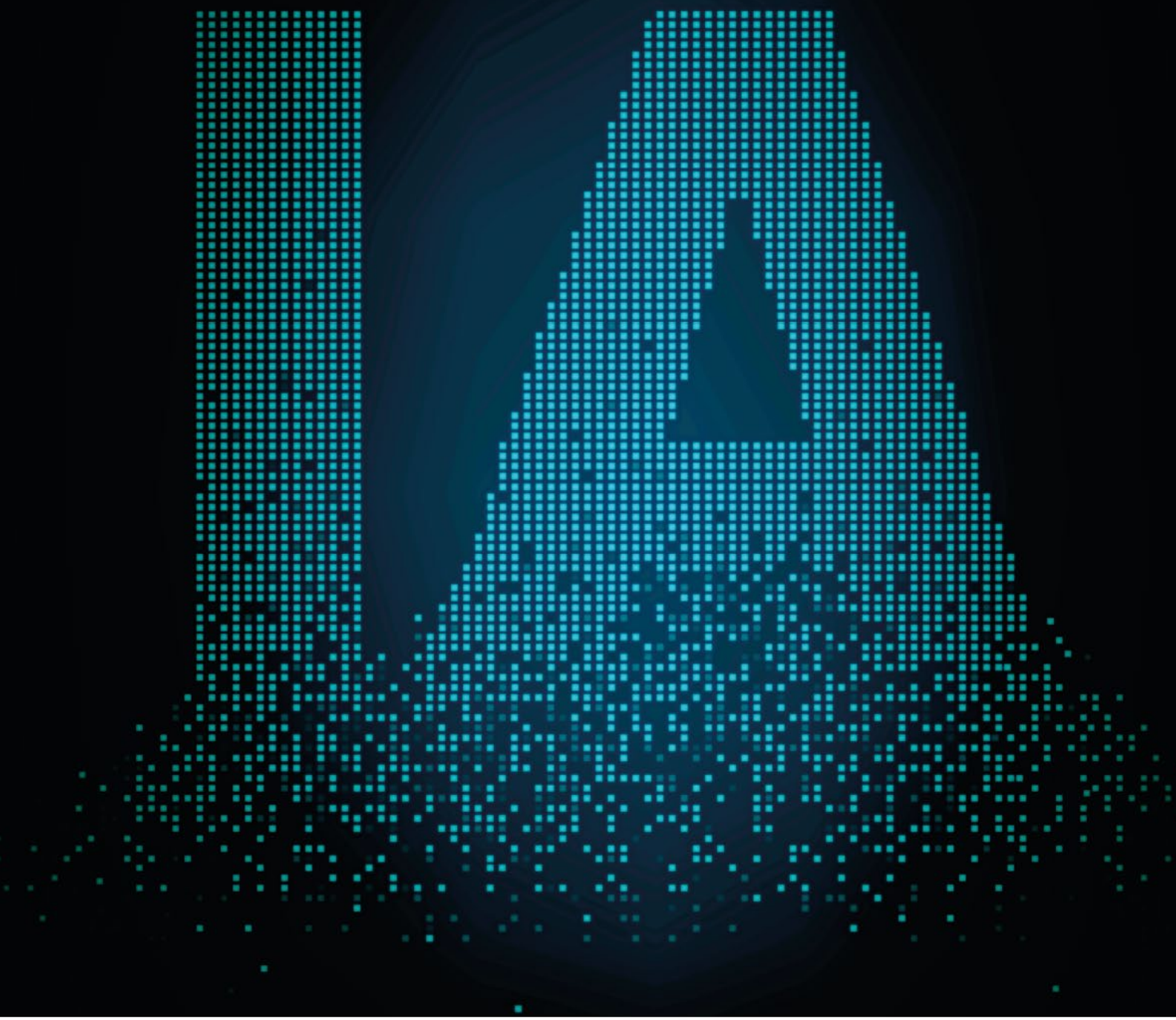
DIRETRIZES DE AUDITABILIDADE E CONFORMIDADE NO DESENVOLVIMENTO E TESTES DE SOLUÇÕES DE IA NO ÂMBITO DO LIAA-3R

2ª EDIÇÃO (REVISTA E ATUALIZADA)



LIAA-3R

iLabTRF3

laboratório de Inovação iJusplab

DIRETRIZES DE AUDITABILIDADE E CONFORMIDADE NO DESENVOLVIMENTO E TESTES DE SOLUÇÕES DE IA NO ÂMBITO DO LIAA-3R

2ª EDIÇÃO (REVISTA E ATUALIZADA)

LIAA-3R

iLabTRF3

laboratório de Inovação iJusplab

L123d Laboratório de Inteligência Artificial Aplicada da 3ª Região (LIIA-3R)
Diretrizes de auditabilidade e conformidade
no desenvolvimento e testes de soluções de IA no âmbito
do LIAA-3R / Grupo de Validação Ético-Jurídica (GVEJ)
do LIAA-3R, iLabTRF3, iJuspLab. -- 2. ed., rev. e
atual.-- São Paulo : LIIA-3R, 2022.
81 p.

DOI 10.5281/zenodo.6109232

1. Inteligência artificial. 2. Conformidade. 3. Auditabilidade.
4. Segurança da informação. 5. Conflito de interesses.
6. Aprendizado de máquina. 7. Aprovação. 8. Validação.
9. Registro. 10. Algoritmo. 11. Incidente. 12. Anonimização
de dados. 13. Desenvolvimento. 14. Lei geral de proteção
de dados (LGPD). 15. Impacto social. 16. BIBJF3RI.Lima,
Caio Moysés de. II.Lozada, Claudia de Oliveira. III.D'Eva,
Maíra Zau Serpa Spina. IV.Carvalho, Matheus Henrique
de Paiva. V.Mariano Júnior, Raul. VI.Monteiro, Renato
Arruda Rocha. VII.Laboratório de Inovação do TRF3
(iLabTRF3). VIII.Laboratório de Inovação da JFSP (iJuspLab).
IX. Título.

CDD: 004.678

CATALOGAÇÃO NA FONTE: Bibliotecária Silvana dos Santos Sant'Anna – CRB8-8/8353

Composição do Tribunal

Desembargadores Federais

MAIRAN GONÇALVES MAIA JÚNIOR - Presidente
PAULO OCTAVIO BAPTISTA PEREIRA
ANDRÉ NABARRETE NETO
MARLI MARQUES FERREIRA
NEWTON DE LUCCA
OTÁVIO PEIXOTO JÚNIOR
THEREZINHA ASTHOLPHI CAZERTA
NERY DA COSTA JÚNIOR
LUIS CARLOS HIROKI MUTA
CONSUELO YATSUDA MOROMIZATO YOSHIDA - Vice-Presidente
MARISA FERREIRA DOS SANTOS - Corregedora- Regional
LUÍS ANTONIO JOHONSOM DI SALVO
NELTON AGNALDO MORAES DOS SANTOS
SÉRGIO DO NASCIMENTO
ANDRÉ CUSTÓDIO NEKATSCHALOW
LUIZ DE LIMA STEFANINI
LUÍS PAULO COTRIM GUIMARÃES
ANTONIO CARLOS CEDENHO
JOSÉ MARCOS LUNARDELLI
DALDICE MARIA SANTANA DE ALMEIDA
FAUSTO MARTIN DE SANCTIS
PAULO GUSTAVO GUEDES FONTES
NINO OLIVEIRA TOLDO
MÔNICA AUTRAN MACHADO NOBRE
TORU YAMAMOTO
MARCELO MESQUITA SARAIVA
LUIZ ALBERTO DE SOUZA RIBEIRO
DAVID DINIZ DANTAS
MAURICIO YUKIKAZU KATO
GILBERTO RODRIGUES JORDAN
HÉLIO EGYDIO DE MATOS NOGUEIRA
PAULO SÉRGIO DOMINGUES
WILSON ZAUHY FILHO
NELSON DE FREITAS PORFIRIO JÚNIOR
VALDECI DOS SANTOS
CARLOS EDUARDO DELGADO
INÊS VIRGÍNIA PRADO SOARES
JOSÉ CARLOS FRANCISCO
LEILA PAIVA MORRISON

Coordenadora do iLabTRF3:
Desembargadora Federal Daldice Santana

Coordenador do iJuspLab:
Juiz Federal Caio Moysés de Lima

Coordenador do GVEJ:
Juiz Federal Raul Mariano Junior

Gerente Técnico do LIAA-3R:
Analista Judiciário Fábio Akahoshi Collado

Membros do GVEJ (em ordem alfabética):
Caio Moysés de Lima (iJuspLab)
Claudia de Oliveira Lozada (iJuspLab/UFAL)
Maíra Zau Serpa Spina D'Eva (iLabTRF3)
Matheus Henrique de Paiva Carvalho (iLabTRF3)
Raul Mariano Junior (iJuspLab)
Renato Arruda Rocha Monteiro (iLabTRF3)

Revisões e Alterações da 2ª Edição (em ordem alfabética):

Caio Moysés de Lima
Claudia de Oliveira Lozada
Fábio Akahoshi Collado
Maíra Zau Serpa Spina D'Eva
Matheus Henrique de Paiva Carvalho
Raul Mariano Junior
Renato Arruda Rocha Monteiro

Elaboração da 1ª Edição (em ordem alfabética):

Aki Ando Kojima
Caio Moysés de Lima
Claudia de Oliveira Lozada
Fábio Akahoshi Collado
Luciana Ortiz Tavares Costa Zanoni
Maíra Zau Serpa Spina D'Eva
Matheus Henrique de Paiva Carvalho
Natália Tavares Amato
Paulo Cezar Neves Junior
Raul Mariano Junior
Renata de Souza Plens
Renato Arruda Rocha Monteiro

Documentação e Apoio Operacional:

Cláudio Roberto Nóbrega Martins

Diagramação da 2ª Edição:

Wladimir Wagner Rodrigues

Diagramação da 1ª Edição:

Paulo César Polimeno

APRESENTAÇÃO

UM MANUAL COM A MARCA DOS LABORATÓRIOS DE INOVAÇÃO

Daldice Santana[*]
Caio Moysés de Lima[**]

Este manual que temos a satisfação de apresentar é fruto de intenso trabalho do Grupo de Validação Ético-Jurídica de Soluções de Inteligência Artificial (GVEJ) do Laboratório de Inteligência Artificial Aplicada da Justiça Federal da 3ª Região (LIAA-3R).

Seu texto traz as marcas típicas do trabalho dos laboratórios de inovação: colaboração, empatia e visão centrada no ser humano. Com efeito, a redação foi construída por muitas mãos, com visão multidisciplinar, buscando-se colocar, em primeiro plano, a perspectiva dos usuários internos e externos dos serviços da Justiça Federal, nunca os enxergando como coisas ou como números em tabelas estatísticas, mas sempre como pessoas dignas de respeito e consideração.

No processo de elaboração do texto, a equipe do GVEJ enfrentou diversos desafios: dificuldade de coordenar os trabalhos a distância em virtude da pandemia que se instalou logo no início de 2020; dificuldade de lidar com um tema novo e complexo,

[*] Desembargadora Federal no Tribunal Regional Federal da 3ª Região (TRF3), Coordenadora do Laboratório de Inteligência Artificial Aplicada da Justiça Federal da 3ª Região - LIAA-3R e do Laboratório de Inovação do TRF3 - iLabTRF3.
[**] Juiz Federal na Seção Judiciária de São Paulo, Coordenador do Laboratório de Inovação da Justiça Federal de São Paulo - iJuspLab.

em grande parte inexplorado; mudanças normativas e mudanças na composição da equipe ocorridas durante a construção do texto. Todos esses desafios foram superados e o resultado que o leitor tem em mãos é a prova de que os melhores trabalhos são realizados na adversidade.

O que motivou a elaboração deste manual foi a necessidade de realizar o trabalho de validação ético-jurídica dos projetos SINARA e SIGMA, primeiras soluções de inteligência artificial desenvolvidas na Justiça Federal da 3ª Região. Quando o GVEJ reuniu-se pela primeira vez, no primeiro semestre de 2020, ainda vigorava a Portaria do Presidente do Conselho Nacional de Justiça (CNJ) n. 25, de 19 de fevereiro de 2019, a qual definia, entre as premissas para a criação de modelos de inteligência artificial no Poder Judiciário, a sua "*validação jurídica e ética"*, assim entendida a possibilidade de auditar referidos modelos "*para análise dos resultados a partir de critérios éticos jurídicos*" (item 4-c do Anexo).

Como a matéria era nova, o grupo entendeu por bem iniciar seus trabalhos com um levantamento do que já havia sido escrito, no mundo, a respeito do tema. Acabou reunindo um vasto material, do qual dois textos se destacaram em razão de seu pioneirismo, abrangência e sistematicidade: a "Carta Europeia sobre o Uso de Inteligência Artificial em Sistemas Judiciais e seu Ambiente" (*European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment*) – aprovada pela Comissão Europeia para a Eficiência da Justiça, órgão do Conselho Europeu, na 31ª sessão plenária realizada nos dias 3 e 4 de dezembro de 2018 em Estrasburgo – e as "Orientações Éticas para uma IA de Confiança", elaboradas pelo Grupo de Peritos de Alto Nível sobre a Inteligência Artificial (GPAN IA) nomeado pela Comissão Europeia.

A ideia inicial era criar uma lista de verificação para as equipes de projeto do LIAA-3R baseada na lista de verificação proposta nas Orientações do GPAN, mas logo se percebeu que isso não seria suficiente para alcançar toda a complexidade envolvida em projetos de inteligência artificial.

De fato, antes mesmo de elaborar uma lista de verificação, era preciso delimitar de forma clara a função dos laboratórios de inovação nos projetos de IA, por ser ela essencialmente distinta daquela desempenhada pelas equipes de Tecnologia da Informação. Sem compreender a atribuição específica dos laboratórios, não se poderia tampouco compreender o escopo dos projetos conduzidos no âmbito do LIAA-3R, nem, por conseguinte, até que ponto deveria avançar o trabalho de validação ético-jurídica conduzido pelo GVEJ. Além disso, muito mais do que propor perguntas para auxiliar o GVEJ na condução dos seus trabalhos, era preciso também transmitir às equipes do LIAA-3R orientações claras sobre como os projetos deveriam ser conduzidos, de modo que pudessem ser antecipados problemas e potenciais obstáculos, além de assegurar que cada projeto fosse apresentado ao GVEJ já devidamente documentado e evitar o desperdício de esforços em projetos com poucas chances de aprovação.

Diante dessas necessidades, concluiu-se pela elaboração de um manual, no qual todas essas questões pudessem ser abordadas de forma mais clara e com mais profundidade. Nesse meio tempo, foi editada pelo CNJ a Resolução n. 332, de 21 de agosto de 2020, que dispõe sobre "*a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências*". Felizmente, a referida resolução parecia inspirar-se nos mesmos documentos europeus selecionados pelo GVEJ, o que tornou relativamente fácil o trabalho de incorporar os seus preceitos ao texto do manual que já estava em processo de elaboração.

Dessa forma, em abril de 2021, depois de mais de um ano de intenso trabalho, o GVEJ finalmente publicou a primeira versão deste manual sob o título "Diretrizes de Auditabilidade e Conformidade no Desenvolvimento e Testes de Soluções de IA no âmbito do LIAA-3R", tendo sido ela a versão utilizada para guiar a validação dos projetos SINARA e SIGMA. Com a conclusão bem-sucedida desse trabalho, logo se mostrou evidente a necessidade de publicar a segunda edição do manual, revista e atualizada, a fim de incorporar o aprendizado obtido pelo GVEJ em sua atuação prática. Essa nova edição traz algumas alterações formais, como o acréscimo da palavra "Manual" no título, a inclusão de novos verbetes no glossário e algumas melhorias de redação. As principais modificações, no entanto, são de natureza substancial e podem ser resumidas da seguinte forma:

1º) na parte introdutória, foram acrescidas explanações a respeito da situação dos projetos do LIAA-3R e da atuação do GVEJ após a revogação da Portaria da Presidência do CNJ n. 25, de 19 de fevereiro de 2019;

2º) ainda na parte introdutória, foram incluídos esclarecimentos acerca do escopo de atuação do GVEJ;

3º) nas diretrizes específicas de conformidade, foram incluídos novos artefatos de documentação, em especial, os relatórios parciais de anotação de dados e das atividades de desenvolvimento de que tratam os Anexo IV e V e o relatório final sobre os *datasets* de que trata o Anexo VI;

4º) a lista de verificação destinada ao próprio GVEJ (Anexo VIII) sofreu completa reformulação, não apenas no tocante à sua redação, mas também quanto à ordem dos temas, a fim de tornar mais fácil a elaboração do parecer a ser emitido pelo GVEJ ao final dos trabalhos.

Este manual foi concebido inicialmente para uso interno, pelas equipes do LIAA-3R. Todavia, acreditamos que pode ser também de grande utilidade para o público em geral, pois se trata de trabalho pioneiro no Brasil que apresenta conteúdo de amplo interesse, já testado na prática nos trabalhos de validação dos projetos SINARA e SIGMA. Aliás, conforme mencionado na parte introdutória desta segunda edição, o manual é um documento "vivo" e "dinâmico", produzido para ser constantemente aprimorado.

Assim, o documento não apenas pode ser lido com proveito pelo público, como também pode beneficiar-se dos comentários, das sugestões e das críticas de quaisquer leitores interessados, a fim de tornar-se cada vez melhor. Todas as contribuições serão bem-vindas!

Desejamos a todos uma ótima leitura! Que este manual inspire a criação de novas soluções de inteligência artificial seguras e confiáveis.

Pela excelência de resultado obtido, cumprimentamos efusivamente a todos os que participaram dos projetos SINARA e SIGMA, em especial àqueles que contribuíram para a elaboração deste documento, fazendo-o na pessoa do atual coordenador do GVEJ, Juiz Federal Raul Mariano Junior, e do Gerente Técnico do LIAA-3R, Fábio Akahoshi Collado.

OS DESAFIOS DE UMA IA DE CONFIANÇA

Raul Mariano Junior[*]

No início do ano de 2020, diante da necessidade de criar condições para a elaboração e desenvolvimento de projetos de sistemas que envolviam IA – Inteligência artificial, tanto no Tribunal Regional da 3ª Região quanto na Seção Judiciária de São Paulo, a partir dos laboratórios de inovação então existentes, criou-se um grupo de voluntários, que seria integrado por recursos humanos de ambos os órgãos e que serviria para institucionalizar mais uma política inovadora e facilitar o intercâmbio institucional para o atingimento dos objetivos, constituído por magistrados, funcionários da área fim, técnicos de várias áreas administrativas e tecnológicas. Essa parceria foi batizada de LIAA - Laboratório de inteligência artificial aplicada da 3ª Região.

Contemporaneamente, com o lançamento do projeto Sinapses pelo CNJ, regulamentou-se alguns atributos desses sistemas, e, dentre eles, de que fossem confiáveis, permitindo que se auditasse o modelo (algoritmos e dados). Sendo a questão muito nova e não se dispondo de conhecimento institucional para certificar esse atributo, foi necessário que aquelas pessoas envolvidas com a inovação e com o LIAA se debruçassem sobre o tema, iniciando um grande ciclo de pesquisas, debates e estudos em grupo, para que se pudesse obter um primeiro nivelamento.

A ideia inicial era de que simplesmente se fizesse um *check list*, que pudesse orientar a equipe de projeto nesse escopo. Contudo, como efeito desse conhecimento adquirido, o grupo ganhou, igualmente, um alargamento da consciência sobre a complexidade multidisciplinar dessa demanda e percebeu que para atestar confiabilidade, tal qual acontece nas políticas de *compliance*, o juízo sobre atributos dos modelos e sistemas (dados, algoritmos, arquitetura, uso proposto e usos possíveis) não poderia recair sobre a mesma equipe de projeto e de desenvolvimento, a fim de não macular de dúvida a confiabilidade desse modelos e sistemas, tendo em vista os possíveis interesses ou vieses, mesmo inconscientes. Daí surgiu o GVEJ - Grupo de Validação Ética e Jurídica, equipe que tenho a honra de coordenar neste momento,

[*] Juiz Federal na Seção Judiciária de Campinas, Coordenador do Grupo de Validação Ético-Jurídica de Soluções de Inteligência Artificial (GVEJ) do LIAA-3R.

formada pelos membros do LIAA com atribuição específica de levar adiante a tarefa de validação ético-jurídica das soluções de IA criadas no âmbito do laboratório.

Nesse caminho, percebeu-se, ainda, que não bastava a elaboração de uma lista simplória de itens a serem observados, mas que também era necessário criar um documento básico, que trouxesse as informações mínimas sobre o que significa atestar a conformidade ético-jurídica de um modelo ou de um sistema de IA. Para isto, toda a equipe do LIAA e eventuais outros convidados debruçaram-se sobre textos científicos publicados, documentação disponível na internet sobre sistemas diversos que empregam a tecnologia, documentos oficiais e regulamentos internacionais e, em especial, sobre dois documentos produzidos na Europa, um no âmbito da própria União Europeia e outro no âmbito do Conselho Europeu, cujo conteúdo veio, posteriormente, a inspirar a redação da Resolução 332 do CNJ, normativo que hoje versa sobre o tema.

Com isto surge então este manual, cuja **primeira edição foi publicada em abril de 2021**. Este documento, contudo, foi inicialmente elaborado sem que tivesse ainda havido alguma aplicação prática de seu conteúdo.

Por outro lado, não havia também qualquer iniciativa correlata publicada em território nacional, e os modelos de IA conhecidos no âmbito do Poder Judiciário estavam, ou em fase de construção, ou já em funcionamento, sem que se tivesse notícia de que um trabalho sistematizado sobre sua confiabilidade tivesse sido realizado. Assim, a atividade de validação ético-jurídica do LIAA foi temporariamente suspensa até que o manual ficasse pronto.

Na atualidade, muitas ainda são as dúvidas da comunidade científica e jurídica quanto aos critérios de conformidade e confiabilidade dos modelos e sistemas de IA, e muito tem sido escrito sobre este tema. Questões sobre vieses cognitivos implícitos, influência indesejada, manipulação de decisões (inconscientes ou criminosas), fragilidades e vulnerabilidades, tanto decorrentes do projeto como da própria tecnologia atual têm sido levantadas e bradadas por muitos autores, quando se trata do uso geral de sistemas com inteligência artificial, e, em especial, quando se trata de seu uso nas atividades típicas de Estado.

Maior preocupação ainda se coloca quando se trata da atividade jurisdicional, a pedra de fechamento do regime legal e democrático desenhado pela Constituição de 1988. Levantaram-se dúvidas, com muita razão, sobre os trade-off implicados na utilização dessa tecnologia, apurando-se, discutindo-se e mesurando conveniências, inconveniências e os riscos de seu uso pelo Poder Judiciário.

A inteligência artificial, na atualidade, é uma realidade inescapável da qual o Judiciário, ainda iniciante na sua incorporação para desempenho de suas atividades institucionais ou administrativas, tem sido constantemente implicado ou provocado pela franca utilização dela pelos demais atores processuais, tais como a advocacia

pública e privada, as grandes empresas, ministério público e outros. Sistemas de consulta processual, de agendamentos, de distribuição e peticionamento, de mineração de dados ou de documentos, para citar apenas alguns usos mais utilizados, são comuns e fazem interface ou interagem com sistemas mais arcaicos e subdimensionados para esse uso massivo máquina-máquina. Os sistemas processuais atuais e de consulta, em geral disponíveis no Judiciário brasileiro, foram desenhados e implementados inicialmente para auxílio na tramitação de processos físicos, modernizados depois para autos eletrônicos, mas não implicaram ou expressaram qualquer mudança fundamental no sistema jurisdicional brasileiro do ponto de vista da automação com inteligência artificial.

O emprego dessa nova tecnologia, por sua vez, tem grande potencial disruptivo. Tem potencial para ameaçar ou simplesmente ignorar salvaguardas históricas do devido processo legal, cultivadas e alargadas durante muitos e muitos séculos. Como exemplo, um sistema desses poderia provocar manipulação ou interferência no julgador humano, acidental ou propositadamente, ou, ainda, o engessamento das decisões de um determinado órgão, vez que eventuais tarefas de apoio (elaboração de minutas, por exemplo) estariam sempre baseadas em registros do passado, já que o treinamento dos modelos de IA se dá a partir de dados preexistentes.

Nesse cenário de inovação e transformação, capitaneado agora por cientistas, engenheiros e técnicos da computação e das ciências exatas, o desafio de criar sistemas de IA confiáveis não poderia deixar de atrair o olhar e a preocupação da comunidade jurídica, em especial no âmbito do Judiciário. A concepção e a manutenção desses sistemas dentro dos limites da legalidade e, em especial, da ética é o grande foco deste documento.

Não é tarefa simples saber como se pode conceber, projetar, executar e manter sistemas computacionais jurídicos e judiciários na zona de conformidade ética e jurídica. Qualquer juízo sobre esse tema – utilização da inteligência artificial e automação – imprescinde de grande carga de conhecimentos das áreas de pesquisa envolvidas, inclusive no que se refere às ciências cognitivas, à psicologia, à filosofia, para além da computação e do direito. É uma área multidisciplinar por definição.

É nesse caldo de cultura que nasce este documento.

Produto das inquietações e do estudo pelos voluntários da equipe, preocupados inicialmente com a elaboração de um simples *check list* que pudesse orientar os trabalhos do GVEJ, acabou por tornar-se um documento didático, prático e científico, aberto à renovação e atualização, ao mesmo tempo em que está disponível para auxiliar outros órgãos que se deparem com as mesmas inquietações e necessidades.

Aprovado o texto da primeira edição em julho 2021, foi ele então levado a um teste real quando colocado em uso pelo GVEJ para validar os projetos Sinara e Sigma. Seguindo, então, seus princípios, recomendações e metodologia, o GVEJ se pôs a avaliar, concretamente, esses dois projetos, os quais receberam parecer favorável em dezembro de 2021.

Com a utilização prática do documento, percebeu-se a necessidade de algumas correções, melhorias e adições no texto original, tanto para deixá-lo mais claro, objetivo e conciso, com também para eliminar algumas incongruências e falhas percebidas pelo seu uso prático. Com isto, realizadas tais revisões, nasceu **esta segunda edição**. Mais robusta, mais clara e testada, pronta para ser novamente aplicada nos próximos projetos, dentro e fora da Justiça Federal da 3ª Região.

ÍNDICE

I - GLOSSÁRIO

Acurácia	Uma das métricas de avaliação do desempenho de um modelo de IA, pela qual se determina o grau de sucesso do modelo em retornar o *output* esperado considerando *datasets* de validação e testes.
ADEG	Assessoria de Desenvolvimento Integrado e Gestão Estratégica do TRF3.
AGES	Assessoria de Gestão de Sistemas de Informação da Presidência do TRF3.
Algoritmo	Um conjunto finito de instruções claras, precisas e suficientes para a resolução de um problema ou de uma classe de problemas computacionais. É a sequência de instruções fornecidas a um sistema computacional para transformar um *input* em um *output*[1]. O art. 3º, inciso I, da Resolução CNJ define o termo nessa mesma linha, como a designar uma "sequência finita de instruções executadas por um programa de computador, com o objetivo de processar informações para um fim específico".
Algoritmo de IA	Algoritmo utilizado para a criação de modelos de IA.

[1] ALPAYDIN, 2016, p. 16.

Anonimização	Nos termos do art. 5º, inciso XI, da LGPD, é a "utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo".
Anotar um *dataset*	Atividade exercida por especialistas da área de negócios, de forma padronizada e sistematizada, consistente em "enriquecer" um *dataset* com informações adicionais necessárias para treinar um algoritmo de aprendizagem de máquina. A anotação pode ser, por exemplo, uma "etiqueta" ou um "rótulo" associado ao dado anotado.
API	*Application Programming Interface*, conjunto de regras e padrões definidos por um sistema computacional para que outro sistema possa ter acesso a suas funcionalidades.
Aprendizado de máquina (AM)	Tradução do termo inglês "*machine learning*". Um "ramo da IA que estuda as formas de os computadores melhorarem sua performance em uma tarefa (aprenderem) por meio da experiência. Dividem-se as formas em que pode ocorrer esse aprendizado em: supervisionado - quando a base de dados usada para treinamento recebe 'anotações' de um especialista; não supervisionado - quando cabe ao sistema encontrar padrões em dados não anotados; e por reforço - quando acontece pela interação, maximizando sinais de bom desempenho"[2]. Segundo ALPAYDIN, o objetivo da aprendizagem de máquina é "criar um programa que se ajuste ao conjunto de dados"[3]. Nesse contexto, um algoritmo de AM é um "molde geral [*general template*] com parâmetros modificáveis, de modo que, ao atribuir valores

[2] PEIXOTO & SILVA, 2019, p. 104. Além do aprendizado supervisionado, não supervisionado e por reforço, Faceli et al. (2021, p. 4) mencionam também o aprendizado semi-supervisionado e o aprendizado ativo.
[3] *Idem*, p. 24.

	diferentes a esses parâmetros, o programa pode fazer coisas diferentes. O algoritmo de aprendizagem ajusta os parâmetros do molde, criando o que se chama de 'modelo' [*model*], mediante a otimização do critério de performance definido em relação aos dados"[4]. Veja também "Aprendizado supervisionado" e "Aprendizado não supervisionado".
Aprendizado supervisionado	Tipo de aprendizado de máquina que requer *datasets* anotados por especialistas humanos para treinamento, validação e testes dos algoritmos de IA. O aprendizado supervisionado está relacionado a tarefas preditivas, nas quais "algoritmos de AM são aplicados a conjuntos de dados de treinamento rotulados para induzir um modelo preditivo capaz de predizer, para um novo objeto representado pelos valores de seus atributos preditivos [características dos objetos], o valor de seu atributo alvo [rótulo]"[5]. Entre as técnicas de aprendizado supervisionado, incluem-se a classificação e a regressão. Veja também "Aprendizado de máquina (AM)" e "Aprendizado não supervisionado".
Aprendizado não supervisionado	Aprendizado de máquina relacionado a tarefas descritivas, nas quais, "ao invés de predizer um valor, algoritmos de AM extraem padrões dos valores preditivos de um conjunto de dados"[6]. Entre as técnicas de aprendizado não supervisionado, incluem-se o agrupamento, a associação e a sumarização. Veja também "Aprendizado de Máquina (AM)" e "Aprendizado Supervisionado".
Auditabilidade	Nos termos das Orientações GPAN, "auditabilidade refere-se à capacidade de um sistema de IA se sujeitar à avaliação dos seus algoritmos, dados e processos de concepção. [...] Tal não implica necessariamente que as informações sobre os modelos de negócios e a

[4] *Idem*, p. 24-25.
[5] FACELI et al., 2021, p. 2-3.
[6] FACELI et al., 2021, p. 2-3.

	propriedade intelectual relacionadas com o sistema de IA tenham de estar sempre publicamente disponíveis"[7].
Carta CEPEJ	Documento intitulado "European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment", aprovado pela CEPEJ na 31ª sessão plenária realizada nos dias 3 e 4 de dezembro de 2018 em Estrasburgo.
CEPEJ	Comissão Europeia para a Eficiência da Justiça, órgão do Conselho Europeu.
CGPDP-3R	Comitê Gestor de Proteção de Dados Pessoais da Justiça Federal da 3ª Região, instituído pela Resolução nº 385, de 20 de outubro de 2020, da Presidência do TRF3, responsável pela avaliação dos mecanismos de tratamento e proteção dos dados existentes com vistas ao cumprimento da LGPD.
CJF	Conselho da Justiça Federal.
CNJ	Conselho Nacional de Justiça.
Comissão Local de Resposta a Incidentes	Comissão Local de Resposta a Incidentes de Segurança da Informação da Justiça Federal da 3ª Região.
Comissão Local de Segurança da Informação	Comissão Local de Segurança da Informação da Justiça Federal da 3ª Região.
CORE	Corregedoria-Regional da Justiça Federal da 3ª Região.
Dado anonimizado	Nos termos do art. 5º, inciso III, da LGPD, é o "dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento".

[7] Orientações GPAN, p. 47, § 148.

Dado pessoal	Nos termos do art. 5º, inciso I, da LGPD, é a "informação relacionada a pessoa natural identificada ou identificável".
Dado pessoal sensível	Nos termos do art. 5º, inciso II, da LGPD, é o "dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural".
Dados estruturados	Dados organizados em formato adequado para leitura por um sistema computacional, de modo a viabilizar a execução de uma determinada tarefa, independentemente do meio de armazenamento. Os dados estruturados podem estar inseridos em bancos de dados, planilhas de cálculo, arquivos de texto, arquivos binários etc.
Dataset	Conjunto de dados, estruturados ou não, utilizados no desenvolvimento, validação e testes de um modelo de IA.
Edital SINAPSES	Edital nº 2/2019, expedido pelo Coordenador do Centro de Inteligência Artificial Aplicada ao PJe do CNJ, com base na Portaria SINAPSES, disponibilizado no DJe de 26 de abril de 2019. Tendo sido a Portaria SINAPSES revogada pela Resolução nº 395/2021 do CNJ, o Edital SINAPSES também deixou de vigorar. Apesar disso, sua menção neste documento permanece necessária para fins de registro histórico.
Enviesamento	As Orientações GPAN definem o enviesamento como "uma tendência parcial a favor ou contra uma pessoa, um objeto ou uma posição. Os enviesamentos podem surgir de muitas formas nos sistemas de IA. Por exemplo, nos sistemas de IA baseados em dados,

	como os produzidos por via da aprendizagem automática, o enviesamento na recolha de dados e na fase de treino pode levar um sistema de IA que apresenta enviesamentos. Na IA baseada na lógica, como os sistemas baseados em regras, podem surgir enviesamentos devido à forma como um engenheiro do conhecimento entenda as regras aplicáveis num determinado contexto. Também podem surgir enviesamentos devido à aprendizagem em linha e à adaptação através da interação. Podem ainda surgir através da personalização, que visa apresentar aos utilizadores recomendações ou fluxos de informações adaptadas aos seus gostos. Não estão necessariamente relacionados com preconceitos humanos ou uma recolha de dados baseada no ser humano. Podem ser suscitados, por exemplo, pelos contextos limitados em que um sistema é utilizado, não havendo nesse caso oportunidades de generalização para outros contextos. O enviesamento pode ser bom ou mau, intencional ou não intencional. Em alguns casos, o enviesamento pode causar resultados discriminatórios e/ou injustos, designados no presente documento por enviesamentos injustos”[8].
Equipe de anotadores	Membros da equipe de projeto designados para as atividades de anotação e curadoria.
Equipe de desenvolvedores	Membros da equipe de projeto designados para as atividades atribuídas pelo Anexo da Portaria SINAPSES ao Gerente Técnico e aos cientistas de dados, cientistas de IA, engenheiros de IA e analistas desenvolvedores full-stack.
Equipe de documentação	Membros da equipe de projeto designados para atividades de documentação.

[8] Orientações GPAN, p. 47-48, § 149.

Equipe de pré-processamento	Membros da equipe de projeto designados para atividades de pré-processamento e *data augmentation* relacionadas à formação dos *datasets*.
Equipe de validação	Membros da equipe de projeto designados para atividades de validação ético-jurídica das soluções de IA.
Equipe de projeto	Pessoas designadas para conduzir um determinado projeto no âmbito do LIAA-3R.
Explicabilidade	Para BARREDO ARRIETA et al., uma solução de IA é explicável quando, "fornece detalhes ou razões para tornar seu funcionamento claro ou fácil de compreender", levando em conta o público-alvo[9]. Segundo as Orientações GPAN, "a explicabilidade diz respeito à capacidade de explicar tanto os processos técnicos de um sistema de IA como as decisões humanas com eles relacionadas (p. ex., os domínios de aplicação de um sistema de IA)"[10].
GPAN IA	Grupo de peritos de alto nível sobre a inteligência artificial selecionados pela Comissão Europeia[11].
GVEJ	Grupo de Validação Ética e Jurídica do LIAA-3R, composto por integrantes do laboratório destacados para elaborar o presente documento e avaliar, com base nas diretrizes aqui definidas, os projetos desenvolvidos no âmbito do laboratório sob o ponto de vista ético-jurídico, com ou sem a participação de outros servidores, magistrados e terceiros convidados para integrarem a equipe de validação específica de cada projeto.

[9] BARREDO ARRIETA et al., p. 85.
[10] Orientações GPAN, p. 22, § 77.
[11] Informações sobre o GPAN AI estão disponíveis em: <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>. Acesso em: 13 set. 2020.

IA

"Inteligência Artificial", área de estudos multidisciplinar[12] que busca pesquisar e desenvolver, por meio de sistemas computacionais, formas de resolver problemas que tradicionalmente eram considerados solucionáveis apenas pela mente humana. São exemplos de subáreas da IA: visão computacional (*computer vision*), reconhecimento e criação de discurso (*speech recognition and synthesis*), processamento de linguagem natural (*natural language processing - NLP*), representação do conhecimento (*knowledge representation*), raciocínio (*reasoning*) e planejamento (*planning*)[13]. Segundo as Orientações GPAN, "os sistemas de inteligência artificial (IA) são sistemas de *software* (e eventualmente também de *hardware*) concebidos por seres humanos, que, tendo recebido um objetivo complexo, atuam na dimensão física ou digital percecionando [sic] o seu ambiente mediante a aquisição de dados, interpretando os dados estruturados ou não estruturados recolhidos, raciocinando sobre o conhecimento ou processando as informações resultantes desses dados e decidindo as melhores ações a adotar para atingir o objetivo estabelecido. Os sistemas de IA podem utilizar regras simbólicas ou aprender um modelo numérico, bem como adaptar o seu comportamento mediante uma análise do modo como o ambiente foi afetado pelas suas ações anteriores. Enquanto disciplina científica, a IA inclui diversas abordagens e técnicas, tais como a aprendizagem automática (de que a aprendizagem profunda e a aprendizagem por reforço são exemplos específicos), o raciocínio automático (que inclui o planeamento, a programação, a representação do conhecimento e o raciocínio, a pesquisa e a otimização) e a robótica (que inclui o controlo, a perceção, os sensores e atuadores, bem como a

[12] Segundo RUSSEL & NORVIG (2013, p. 7 e seguintes), contribuíram para a IA a Filosofia, a Matemática, a Economia, a Neurociência, a Psicologia, a Engenharia de Computadores, a Cibernética (Teoria do Controle) e a Linguística.

[13] PEIXOTO & SILVA, 2019, p. 33.

	integração de todas as outras técnicas em sistemas ciberfísicos)"[14].
iJuspLab	Laboratório de Inovação da JFSP.
iLabTRF3	Laboratório de Inovação do TRF3.
iNovaTRF3	Grupo formado por servidores lotados nas diversas áreas do TRF3, buscando-se a maior representatividade, prestigiando-se a diversidade de formações e respeitando-se a pluralidade de ideias, criado com o objetivo de desenvolver atividades voltadas para a gestão da inovação, a gestão estratégica, a rede de governança integrada, colaborativa e participativa, a gestão da comunicação, a gestão por resultados e a gestão de dados.
Input	Dado ou conjunto de dados fornecido a um programa de computador para processamento.
JFSP	Justiça Federal de 1º Grau em São Paulo.
LGPD	Lei Geral de Proteção de Dados (Lei nº 13.709, de 14 de agosto de 2018).
LIAA-3R	Laboratório de Inteligência Artificial Aplicada da 3ª Região criado pela Portaria Instituidora.
Modelo de IA	Algoritmo ou conjunto de algoritmos de aprendizagem de máquina já calibrado ou parametrizado para a resolução de um certo tipo de problema computacional. Nos termos do art. 3º, inciso II, da Resolução CNJ, trata-se de um "conjunto de dados e algoritmos computacionais, concebidos a partir de modelos matemáticos, cujo objetivo é oferecer resultados inteligentes, associados ou comparáveis a determinados aspectos do pensamento, do saber ou da atividade humana".

[14] Orientações GPAN, p. 47, §§ 143 e 144.

MVP	Produto Mínimo Viável (*Minimum Viable Product*). Segundo o autor que disseminou o termo, Eric Ries, trata-se do "produto que tem apenas os recursos necessários (e não mais) para satisfazer as necessidades dos *early adopters* [...]"[15]. O termo "*early adopters*" designa os primeiros usuários de uma solução, que acreditam no projeto e estão dispostos a contribuir para o seu sucesso dando seu *feedback* aos desenvolvedores[16].
NUIT	Núcleo de Inovação Tecnológica da JFSP.
Orientações GPAN	Versão oficial em língua portuguesa do documento intitulado "Orientações Éticas para uma IA de Confiança" elaborado pelo GPAN IA (cf. UNIÃO EUROPEIA, 2019).
Output	Dado ou conjunto de dados retornado por um programa de computador como resultado de sua execução.
PGP3R	Portal de Gestão de Projetos da Justiça Federal da 3ª Região.
PJe	Processo Judicial Eletrônico do Poder Judiciário, instituído pela Resolução CNJ nº 185, de 18 de dezembro de 2013.
Portaria CNJ	Portaria CNJ nº 271, de 4 de dezembro de 2020, que regulamenta o uso de Inteligência Artificial no âmbito do Poder Judiciário.
Portaria Instituidora	Portaria Conjunta nº 1, de 30 de janeiro de 2020, editada pela Presidência do TRF3 e pela Diretoria do Foro da JFSP para instituir parceria permanente entre os laboratórios de inovação do Tribunal Regional Federal da 3ª Região (iLabTRF3) e da Seção Judiciária de São Paulo (iJuspLab) para a pesquisa e o

[15] RIES, 2009.
[16] RIES, 2012, pos. 173.

	desenvolvimento de modelos de inteligência artificial, mediante a criação do Laboratório de Inteligência Artificial Aplicada da 3ª Região - LIAA-3R.
Portaria SINAPSES	Portaria nº 25, de 19 de fevereiro de 2019, editada pelo Presidente do CNJ para instituir o Laboratório de Inovação para o Processo Judicial em meio Eletrônico – Inova PJe e o Centro de Inteligência Artificial aplicada ao PJe. Apesar de ter sido a portaria revogada pela Resolução nº 395/2021 do CNJ, a sua menção neste documento permanece necessária para fins de registro histórico.
Precisão	Índice que compõe a acurácia e mede, dentre as previsões do algoritmo, quantas estão corretas.
Programa de computador	Conjunto de um ou mais algoritmos expressos em instruções executáveis por um sistema computacional ou que possam ser traduzidas em instruções executáveis por um sistema computacional.
Protótipo	Segundo o site Usability.gov, protótipo é o "rascunho de um produto que permite explorar suas ideias e mostrar a intenção por trás de um recurso ou o conceito geral de design para os usuários antes de investir tempo e dinheiro no desenvolvimento. Um protótipo pode ser qualquer coisa, desde desenhos em papel (baixa fidelidade) a algo que permite clicar em algumas partes do conteúdo para um site em pleno funcionamento (alta fidelidade)"[17].
Rastreabilidade	Segundo as Orientações GPAN, "a rastreabilidade de um sistema de IA refere-se à capacidade de acompanhar os dados do sistema e os processos de desenvolvimento e implantação do mesmo, normalmente por meio de uma identificação registada [sic] documentada"[18]. Nos termos do § 76

[17] Disponível em: <https://www.usability.gov/how-to-and-tools/methods/prototyping.html>. Acesso em 13 set. 2020.
[18] Orientações GPAN, p. 49, § 158.

	do mesmo documento, "os conjuntos de dados e os processos que produzem a decisão do sistema de IA, incluindo os processos de recolha e etiquetagem dos dados, bem como os algoritmos utilizados, devem ser documentados da melhor forma possível para permitir a rastreabilidade e um aumento da transparência. Isto também se aplica às decisões tomadas pelo sistema de IA. Deste modo, é possível identificar os motivos por que uma decisão de IA foi errada, o que, por sua vez, poderá ajudar a evitar erros futuros. A rastreabilidade facilita, assim, a auditabilidade e a explicabilidade"[19].
Recall	Índice que compõe a acurácia e calcula, dentre as situações que o algoritmo deveria prever como positivas, quantas foram previstas.
Reprodutibilidade	Segundo as Orientações GPAN, "a reprodutibilidade descreve se uma experiência de IA apresenta o mesmo comportamento quando repetida nas mesmas condições"[20].
Resolução CNJ	Resolução CNJ nº 332, de 21 de agosto de 2020, que dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências.
SEI	Sistema Eletrônico de Informações utilizado na Justiça Federal para tramitação de processos administrativos.
SETI	Secretaria de Tecnologia da Informação do TRF3.
SINAPSES	Plataforma de desenvolvimento e disponibilização em larga escala de modelos de IA desenvolvida e mantida pelo CNJ, na qual os modelos de IA são disponibilizados e consumidos sob a forma de APIs. Segundo o art. 3º, inciso III, da Resolução CNJ, a plataforma é uma "solução computacional, mantida

[19] *Idem*, p. 21-22.
[20] *Idem*, p. 48, § 150.

	pelo Conselho Nacional de Justiça, com o objetivo de armazenar, testar, treinar, distribuir e auditar modelos de Inteligência Artificial".
Sistema ou solução de IA	Solução computacional que se utiliza de um ou mais modelos de IA para a realização das tarefas para as quais foi concebida.
Tratamento de dados	Nos termos do art. 5º, inciso X, da LGPD, é "toda operação realizada com dados pessoais, como as que se referem a coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração".
TI	Tecnologia da Informação.
Usuário	Para os fins da Resolução CNJ, entende-se por usuário qualquer "pessoa que utiliza o sistema inteligente e que tem direito ao seu controle, conforme sua posição endógena ou exógena ao Poder Judiciário, pode ser um usuário interno ou um usuário externo" (art. 3º, inciso IV).
Usuário externo	Para fins da Resolução CNJ, é qualquer "pessoa que, mesmo sem ser membro, servidor ou colaborador do Poder Judiciário, utiliza ou mantém qualquer espécie de contato com o sistema inteligente, notadamente jurisdicionados, advogados, defensores públicos, procuradores, membros do Ministério Público, peritos, assistentes técnicos, entre outros" (art. 3º, inciso VI).
Usuário interno	Para fins da Resolução CNJ, é qualquer "membro, servidor ou colaborador do Poder Judiciário que desenvolva ou utilize o sistema inteligente" (art. 3º, inciso V).

II - INTRODUÇÃO

1) Objetivo deste Documento

O Laboratório de Inteligência Artificial Aplicada da Justiça Federal da 3ª Região - LIAA-3R foi criado para "incentivar a pesquisa e o desenvolvimento de modelos de inteligência artificial que contribuam para o aprimoramento dos serviços judiciais e administrativos no âmbito da 3ª Região". Do ponto de vista organizacional, o LIAA-3R não é um novo órgão, mas uma "parceria permanente" entre os dois laboratórios de inovação já existentes: o iLabTRF3 e o iJuspLab[21]. Assim, o LIAA-3R se vale, no desempenho de suas atividades, de todos os recursos materiais e humanos pertencentes aos dois laboratórios de inovação da Justiça Federal da 3ª Região[22].

A parceria foi concebida não somente para ampliar a atuação conjunta dos dois laboratórios de inovação, mas sobretudo para oferecer à área de TI o aporte de recursos humanos, materiais e metodológicos adicionais necessários para o desenvolvimento de modelos de IA. Os benefícios resultantes não são meramente quantitativos, pois os laboratórios trazem para os projetos de IA os seguintes aspectos qualitativos que as equipes técnicas de TI dificilmente poderiam obter sozinhas:

[21] "Art. 1º Instituir parceria permanente entre o Laboratório de Inovação do Tribunal Regional Federal da 3º Região - iLabTRF3 e o Laboratório de Inovação da Justiça Federal de São Paulo - iJuspLab, este último por intermédio de seu Centro de Estudos e Pesquisas em Inteligência Artificial e Jurimetria, para o fim de incentivar a pesquisa e o desenvolvimento de modelos de inteligência artificial que contribuam para o aprimoramento dos serviços judiciais e administrativos no âmbito da 3ª Região."

[22] "Art. 3º O LIAA-3R não disporá de recursos próprios, mas utilizará os recursos materiais, humanos, tecnológicos e metodológicos disponíveis no iLabTRF3 e no iJuspLab, observando, para tanto, as regras vigentes em cada laboratório quanto ao uso dos referidos recursos.
[...]"

1º) Ao formarem equipes multidisciplinares[23] e utilizarem metodologias que colocam em primeiro plano a perspectiva dos usuários dos serviços[24], os laboratórios contribuem para a "legitimação do Poder Judiciário perante o cidadão, que em tempos de ampla transparência de dados, exige a prestação de um serviço público de melhor qualidade"[25], facilitam "o entendimento completo dos problemas complexos"[26] e reduzem o risco de enviesamento e discriminação injustos.

2º) Por sua vocação experimental, que se opõe "aos costumes burocráticos" e que tem, entre seus atributos, "o da liberdade para o erro"[27], os laboratórios permitem antecipar problemas e reduzir custos no desenvolvimento das soluções, tanto no aspecto financeiro quanto em horas de trabalho humano.

3º) Os laboratórios de inovação facilitam o trabalho colaborativo e a integração com outros indivíduos, órgãos e entidades internos ou externos (empresas, universidades, outros órgãos públicos), assim como a prospecção de soluções inovadoras, tendo em vista que operam em rede, mantendo contato frequente com diversos outros agentes de inovação no Brasil e no exterior[28].

---

[23] "Além disso, a inovação construída sob olhares multidisciplinares, a partir da perspectiva de que todos temos talentos, se traduz em melhores soluções, não pensadas nas perspectivas solitárias. Nesse sentido, juízes, servidores e demais atores envolvidos com o serviço, juntos em ambientes horizontais, favorecem a construção de inovações que aprimoram o serviço." (ZANONI, 2019, p. 49) "Por isso optamos por criar times multidisciplinares,nos quais cada integrante possui background, vivência e opiniões diferentes, tornando o projeto mais completo e rico em informações e perspectivas. Além do time de inovação que lidera o projeto, os usuários, envolvidos e interessados também são convidados a cocriar com a equipe, pois são eles quem mais podem contribuir de acordo com suas experiências e contato com o problema." (DOURADO, 2019, p. 81) "Segundo as diretrizes do projeto, ao utilizar métodos multidisciplinares, as atividades do laboratório integram magistrados, servidores, cidadãos e demais stakeholders na colaboração entre essas diferentes visões dos mesmos problemas, para a eliminação da hierarquia na construção coletiva de novas formas e modelos para a prestação de serviços pelo Poder Judiciário." (*idem*, p. 219)

[24] "O olhar aprofundando para o usuário do serviço, e o pensar o serviço público a partir de sua perspectiva, legitima a atuação do poder público, cujo desiderato no Estado Democrático de Direito é o de entregar um serviço público que atenda às suas expectativas. Os processos de empatia que compõem as técnicas de inovação, revelam necessidades e expectativas ocultas ou que nunca foram sentidas. Este exercício de empatia constitui processo difícil, uma vez que o juiz e o servidor público estão habituados a desenvolver seus projetos dentro dos gabinetes na ótica de quem presta o serviço, e da forma, não raras vezes, que melhor atenda às suas possibilidades. Este exercício, portanto, inverte a forma de pensar e construir o serviço, certamente, proporcionando aprovação do serviço pelo usuário." (ZANONI, 2019, p. 49) "Se o direito material está se adaptando a essas novas demandas oriundas do uso de automação, de inteligência artificial e da análise de big data no dia a dia da sociedade; se a automação e as técnicas de ciências de dados estão nos auxiliando a otimizar a produção de bens e serviços e a resolver os nossos problemas cotidianos; é certo que o uso das técnicas mais avançadas de design também estão, por sua vez, ajudando a colocar o ser humano como o centro de todo esse movimento." (COELHO, 2019, p. 217)

[25] ZANONI, 2019, p. 47.

[26] GREGÓRIO, 2019, p. 62.

[27] *Idem*, p. 76.

[28] "A inovação trabalha em rede, o compartilhamento com outras instituições na mesma vibração permite a conexão com o espírito da mudança e a troca de experiências. A prospecção com o que está

Assim, a atuação do LIAA-3R é complementar à dos órgãos técnicos do TRF3[29], especialmente a SETI, não representando conflito, interferência ou retrabalho, visto que os experimentos realizados no âmbito no laboratório favorecem a descoberta de novas soluções, melhor ajustadas às necessidades reais dos usuários, e ajudam a identificar riscos e a antecipar problemas. Os benefícios desse modelo de atuação complementar foram comprovados na prática, pelo sucesso do projeto SIGMA, vencedor da 18ª edição do Prêmio Innovare, na categoria CNJ/Tecnologia[30]. Essa solução, inicialmente concebida e desenvolvida pelo Gerente Técnico do LIAA-3R, Fábio Akahoshi Collado, e depois aperfeiçoada no LIAA-3R pelo desenvolvimento do modelo de IA denominado SINARA e implementada pela SETI no PJe para uso pelos gabinetes de 1º e 2º graus de jurisdição, consiste em um assistente de ranqueamento de modelos de decisões judiciais para auxiliar os magistrados e seus assessores a identificarem quais deles melhor se ajustam a cada caso analisado, poupando o tempo outrora despendido por seres humanos na realização dessa tarefa.

Os projetos SINARA/SIGMA foram os primeiros submetidos à avaliação do GVEJ. A sua tramitação interna e os artefatos elaborados para documentá-los podem servir de referência para a validação ético-jurídica de outros projetos de IA no futuro[31].

Desde a sua origem, o LIAA-3R foi orientado a desenvolver seus projetos para

---

acontecendo no mundo contribui para a espiral de inovação que deve existir na instituição. Inovar hoje pode significar atraso amanhã, considerando o avanço exponencial da tecnologia e seus reflexos, como alteração dos formatos de organização e prestação de serviços. Por isso, canais de abertura de inovação constante contribuem para a frequente assimilação institucional da rápida e inevitável transformação das organizações decorrentes da revolução digital. Mas nada é mais importante para a cultura da inovação do que a construção coletiva e gestão compartilhada. A perspectiva de que todos os servidores contribuem para os avanços do serviço público faz com que as propostas de mudanças estejam conectadas com as necessidades e possibilidades, com forte percentual de sucesso. Quando construímos a solução em conjunto, projetamos na sua concretização, afastando o personalismo. Além disso, a inovação construída sob olhares multidisciplinares, a partir da perspectiva de que todos temos talentos, se traduz em melhores soluções, não pensadas nas perspectivas solitárias. Nesse sentido, juízes, servidores e demais atores envolvidos com o serviço, juntos em ambientes horizontais, favorecem a construção de inovações que aprimoram o serviço." (ZANONI, 2019, p. 49)

[29] "Art. 3º [...]O LIAA-3R não disporá de recursos próprios, mas utilizará os recursos materiais, humanos, tecnológicos e metodológicos disponíveis no iLabTRF3 e no iJuspLab, observando, para tanto, as regras vigentes em cada laboratório quanto ao uso dos referidos recursos.
[...]
§ 2º As atividades do LIAA-3R deverão ser desempenhadas de modo a não interferir com outras iniciativas das áreas técnicas do Tribunal Regional Federal da 3.a Região.
§ 3º A criação de mais de um modelo de inteligência artificial, por equipes diferentes, para a solução de um mesmo tipo de problema, não significará, por si, a existência de conflito, interferência ou retrabalho.
[...]"

[30] TRIBUNAL REGIONAL FEDERAL DA 3ª REGIÃO. Projeto Sigma, do TRF3, Ganha Prêmio Innovare 2021. Disponível em: <http://web.trf3.jus.br/noticias-intranet/Noticiar/ExibirNoticia/412508-projeto-sigma-do-trf3-ganha-premio-innovare-2021>. Acesso em: 6 jan. 2022.

[31] O parecer do GVEJ referente aos dois projetos citados pode ser encontrado no documento SEI nº 8354929, expediente nº 0295888-14.2021.4.03.8000.

a plataforma SINAPSES, antecipando-se, desse modo, ao que hoje preceitua o art. 10 da Resolução CNJ[32]. Além disso, o laboratório deve estar sempre alinhado aos objetivos estratégicos do TRF3 e às diretrizes definidas pelo CNJ para o desenvolvimento de modelos de IA no âmbito do Poder Judiciário, nos termos dos §§ 2º e 3º do art. 4º da Portaria Instituidora. Por essas razões, os §§ 3º e 4º do art. 4º da Portaria Instituidora conferem autonomia ao LIAA-3R para submeter seus projetos diretamente ao CNJ/SINAPSES, independentemente de nova autorização específica[33].

Na época em que o LIAA-3R foi criado, a inscrição de projetos no CNJ ainda era regida pela Portaria nº 25, de 19 de fevereiro de 2019, da Presidência daquele órgão ("Portaria SINAPSES"), e pelo Edital nº 2, de 26 de abril de 2019 ("Edital SINAPSES"). Esses normativos exigiam que os projetos fossem apresentados como projetos de pesquisa, a serem avaliados segundo os critérios definidos no item 4.5 do edital[34]. Exigia-se, ademais, que a equipe contasse com certos papéis específicos, a saber, um coordenador, um gestor técnico, cientistas de dados, cientistas de IA, engenheiros de IA, analistas desenvolvedores "full-stack" e curadoria (item 5 do anexo à portaria[35]).

---

[32] "Art. 10. Os órgãos do Poder Judiciário envolvidos em projeto de Inteligência Artificial deverão:
I – informar ao Conselho Nacional de Justiça a pesquisa, o desenvolvimento, a implantação ou o uso da Inteligência Artificial, bem como os respectivos objetivos e os resultados que se pretende alcançar;
II – promover esforços para atuação em modelo comunitário, com vedação a desenvolvimento paralelo quando a iniciativa possuir objetivos e resultados alcançados idênticos a modelo de Inteligência Artificial já existente ou com projeto em andamento;
III – depositar o modelo de Inteligência Artificial no Sinapses."

[33] "Art. 4º Os projetos do LIAA-3R serão documentados em expedientes eletrônicos específicos e serão submetidos a aprovação interna e registrados para acompanhamento segundo os procedimentos e as boas práticas definidos no âmbito da 3ª Região.
[...]
§ 3º Com o intuito de obter acesso à Plataforma SINAPSES e a outros recursos de desenvolvimento que vierem a ser disponibilizados pelo Conselho Nacional de Justiça - CNJ, o LIAA-3R poderá submeter ao referido órgão os projetos que já tenham obtido aprovação interna no âmbito da 3ª Região, independentemente de nova autorização específica.
§ 4º Na hipótese do parágrafo anterior, caberá à própria equipe do LIAA-3R estruturar os projetos a serem submetidos ao CNJ e providenciar toda a documentação exigida pelo referido órgão, nos termos do Edital nº 2, de 26 de abril de 2019, da Portaria nº 25, de 19 de fevereiro de 2019, e de quaisquer outros atos normativos que vieram a modificá-los ou a sucedê-los.
[...]"

[34] "4.5. O projeto de pesquisa receberá uma nota de 0 (zero) a 100 (cem) pontos, distribuídos conforme a avaliação dos seguintes itens:
4.5.1. capacidade de formular o projeto com clareza, coesão e concisão (0-20 pontos);
4.5.2. coerência entre tema, problema, objetivo geral e objetivos específicos (0-20 pontos);
4.5.3. alinhamento do projeto aos Macrodesafios do Poder Judiciário (0-20 pontos);
4.5.4. potencial de impacto da pesquisa para o Poder Judiciário, em especial no contexto do processo judicial em meio eletrônico (0-20 pontos);
4.5.5. escalabilidade do projeto em relação aos diversos tipos de processos judiciais e segmentos de justiça (0-20 pontos)."

[35] "5. Atores
Participar das pesquisas do Centro de IA exigirá perfis determinados, os quais independerão, em um primeiro momento, de conhecimento técnico específico. Além disso, um mesmo participante poderá

Além disso, os projetos deveriam observar certas premissas definidas no item 4 do Anexo à Portaria nº 25/2019, dentre as quais incluía-se a "validação jurídica e ética dos modelos", assim definida no item 4-c (não grifado no original):

> "Os modelos de IA que forem utilizados na tomada de decisões ou produção de artefatos deverão ser passíveis de auditoria para análise dos resultados a partir de critérios éticos jurídicos. O processo de auditoria será definido pelo CNJ."

Com a revogação da Portaria SINAPSES pela Resolução nº 395, de 7 de junho de 2021, do CNJ, as exigências acima mencionadas deixaram de subsistir. Desde então, não há mais necessidade de formatar os projetos de IA como projetos de pesquisa ou

---

reunir diversos perfis ou, ainda, o tribunal poderá não possuir pessoas para todos os perfis. Os atores (perfis) desejados para as equipes observarão as regras abaixo:

a) Coordenador. Caberá ao coordenador recepcionar as demandas, solicitações e relatórios do Gestor Técnico quanto às atividades desempenhadas pela equipe do tribunal. Será de sua responsabilidade alinhar, em conjunto com os Coordenadores dos demais tribunais parceiros, a manutenção das ações necessárias ao andamento do projeto em acordo com as premissas adotadas com a comunidade colaborativa para avanço das pesquisas elencadas no projeto. Ele atua como ponte entre a equipe que está alocada no laboratório e o tribunal de origem.Só será constituído um Coordenador responsável pelo tribunal partícipe quando da existência de um Gestor Técnico. Este papel não demanda atividade presencial.

b) Gestor Técnico. Caberá ao Gestor Técnico acompanhar, gerenciar e administrar a execução das atividades e pesquisas desenvolvidas pelos analistas de seu tribunal. As demandas e pesquisas desempenhadas pela equipe serão geridas por esse papel, que será responsável por garantir o alinhamento e a integração com as pesquisas desenvolvidas por outros tribunais. Só será constituído um Gestor Técnico atuando pelo tribunal partícipe quando da existência de ao menos 01 (um) representante, em sua equipe, dos seguintes papéis: Cientista de Dados, Cientista de Inteligência Artificial, Engenheiro de Inteligência Artificial, Curador. Na falta de um desses papéis, a equipe poderá ser complementada com um ou mais Analistas Desenvolvedor Full Stack. Quando não houver o atendimento aos critérios acima citados, a equipe, na quantidade que estiver provida, será incorporada a do próprio Centro de IA ou de outro tribunal.

c) Cientistas de Dados. Responsável por realizar coletas de grandes massas de dados e, em uma segunda etapa, transformá-los em um formato mais prático. Para tal, utilizará técnicas de extração de dados com variações de linguagens de programação como R, Python, entre outras.

d) Cientista de Inteligência Artificial. Responsável pela pesquisa de subáreas da IA, tais como análise semântica, processamento de linguagem natural, Deep Learning, Machine Learning, Visão Computacional. Entre suas funções, está a responsabilidade de entender o negócio e alinhar as melhores técnicas para criação dos modelos de IA aplicáveis a cada caso. É o líder da pesquisa e desenvolvimento em sua equipe.

e) Engenheiro de Inteligência Artificial. Responsável pelo desenvolvimento e aplicação de softwares destinados ao uso de modelos de IA. Deve possuir conhecimento avançado em umas linguagens de programação usualmente aplicáveis, tais como Python ou Java.

f) Analista Desenvolvedor Full-Stack. Responsável por atuar em várias áreas. Para esse ator, são necessários conhecimentos avançados nas principais linguagens e ferramentas utilizadas nos sistemas Sinapses e PJe. Seu trabalho consiste em desenvolver componentes de software nessas tecnologias e integrá-los ao Sinapses ou ao PJe.

g) Curadoria. Responsável por efetuar o treinamento supervisionado do modelo de IA e arbitrar divergências entre os resultados apresentados por esse e a escolha do usuário, quando aplicável. Para o treinamento supervisionado é desejável um nível satisfatório de conhecimento jurídico, para que possa operar a atividade com melhor precisão. Quando se tratar de arbitragem o conhecimento jurídico deve ser pleno."

de contemplar determinados papéis mínimos nas equipes de projeto. Todavia, entre 2019 e 2021, outros preceitos normativos passaram também a regular os projetos de IA desenvolvidos no âmbito do Poder Judiciário, impondo exigências de ordem diferente, porém muito mais amplas. Merecem destaque as seguintes:

a) a Lei Geral de Proteção de Dados Pessoais – LGPD (Lei nº 13.709, de 14 de agosto de 2018), que define as regras sobre o tratamento de dados pessoais e entrou totalmente em vigor em 1º de agosto de 2021[36];

b) a Resolução nº 332, de 21 de agosto de 2020, do CNJ ("Resolução CNJ"), que "dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências"; e

c) a Portaria nº 271, de 4 de dezembro de 2020, da Presidência do CNJ (doravante "Portaria CNJ"), que "regulamenta o uso de Inteligência Artificial no âmbito do Poder Judiciário".

Com essas novas regras, tornou-se obrigatório o depósito, na plataforma SINAPSES, de todos os modelos de IA produzidos no âmbito do Poder Judiciário (cf. art. 10, inciso III, da Resolução CNJ). Além disso, tais modelos devem agora observar os preceitos da LGPD quanto ao tratamento de dados pessoais, assim como os preceitos ético-jurídicos previstos na Resolução CNJ, muitos dos quais inspirados no "European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment" da Comissão Europeia para a Eficiência da Justiça (órgão do Conselho Europeu) e nas "Orientações Éticas para uma IA de Confiança" elaborado pelo Grupo de peritos de alto nível sobre a inteligência artificial selecionados pela Comissão Europeia.

Assim, não obstante a revogação da Portaria nº 25/2019 e do Edital nº 2/2019, a validação ético-jurídica continua a ser atividade relevante nos projetos de IA realizados no âmbito do Poder Judiciário.

Nesse contexto, o LIAA-3R destacou alguns de seus integrantes para compor um grupo de validação ético-jurídica (o GVEJ), cujo propósito é não apenas realizar as atividades de validação em si, como também definir as diretrizes e dar as orientações necessárias para que as equipes do LIAA-3R possam observar desde o início dos projetos os preceitos que serão depois avaliados na etapa da validação ético-jurídica.

Este documento consolida as diretrizes e orientações definidas pelo GVEJ até o momento e contém, ainda, instruções destinadas ao próprio GVEJ sobre o modo de

[36] Nos termos do art. 65 da referida lei, os seus arts. 55-A, 55-B, 55-C, 55-D, 55-E, 55-F, 55-G, 55-H, 55-I, 55-J, 55-K, 55-L, 58-A e 58-B entraram em vigor no dia 28 de dezembro de 2018; os demais artigos, salvo pelos arts. 52, 53 e 54, entraram em vigor 24 meses após a sua publicação, ou seja, em 15 de agosto de 2020; e os arts. 52, 53 e 54 somente passaram a vigorar a partir de 1º de agosto de 2021.

conduzir os seus procedimentos de validação. A primeira edição, elaborada no primeiro semestre de 2021, ainda estava moldada pelos preceitos da Portaria SINAPSES. A presente edição já contempla as inovações normativas, além de agregar a experiência e o conhecimento adquiridos pelos membros do GVEJ no procedimento de validação dos projetos SINARA/SIGMA, concluído no final de 2021.

2) Escopo da Atuação do GVEJ

Sobre o escopo da atuação do GVEJ, o primeiro ponto a ser ressaltado diz respeito à própria natureza do trabalho de validação ético-jurídica, que tem dois propósitos apenas: assegurar a auditabilidade das soluções IA e verificar a conformidade com as regras em vigor, segundo o escopo e a finalidade de cada projeto.

Isso significa, em primeiro lugar, que as atividades do GVEJ não são atividades de auditoria propriamente dita, pois visam apenas a assegurar a auditabilidade dos projetos de IA, para o caso de se mostrar necessária uma auditoria no futuro. Por conseguinte, o GVEJ não precisa necessariamente empreender pesquisa aprofundada sobre todos os assuntos ou verificar cada uma das informações prestadas pelas equipes de projeto, mas deve zelar para que todas as informações necessárias estejam disponíveis caso uma auditoria deseje empreender tais pesquisas ou verificações.

Em segundo lugar, a análise de conformidade não se dá em abstrato ou hipoteticamente, mas deve conectar-se ao escopo e à finalidade específicos de cada projeto. Assim, não cabe ao GVEJ conjecturar os riscos ou os problemas que poderiam resultar de eventuais desvios no uso das soluções de IA em relação à finalidade para as quais foram concebidas, até porque isso tornaria o seu trabalho impossível. Cabe ao GVEJ, no entanto, zelar para que o escopo e a finalidade dos projetos fiquem devidamente documentados, a fim de que os eventuais usuários das soluções de IA possam compreender com clareza os casos de uso recomendados e os casos de uso não recomendados ou não abrangidos pelo projeto.

Essas são as restrições de escopo que decorrem da natureza das atividades do GVEJ. Há outras restrições ainda, decorrentes da natureza dos laboratórios de inovação em geral e do escopo de atuação do LIAA-3R em especial.

Conforme já mencionado, o LIAA-3R não concorre com as atividades desempenhadas pelos órgãos técnicos do TRF3, como a SETI ou a AGES, mas atua de modo complementar a esses órgãos técnicos, aportando aos processos de desenvolvimento de soluções de IA métodos de experimentação, com equipes multidisciplinares e metodologias que colocam em primeiro plano a perspectiva do usuário.

Além disso, assim como outros laboratórios de inovação, o LIAA-3R não tem agenda própria, funciona como um espaço de criação à disposição de outros órgãos administrativos e dos usuários dos serviços, a fim de que estes possam desenvolver e propor à administração as suas próprias soluções.

Por conseguinte, não compete ao laboratório colocar em produção soluções de IA ou implementá-las dentro do parque tecnológico da Justiça Federal. Sua atividade está adstrita à criação de "protótipos ou produtos de viabilidade mínima"[37] e à colaboração "em projetos de inteligência artificial desenvolvidos e mantidos por terceiros", segundo preceituam os incisos I e III do art. 3º da Portaria Instituidora.

O laboratório também não é um escritório de projetos. Não lhe cabe definir procedimentos gerais para aprovação de novas ideias, nem cuidar de sua posterior implantação. Por essa razão é que, nos termos do art. 4º, *caput* e §§ 1º e 2º, da Portaria Instituidora, o LIAA-3R deve submeter seus projetos à aprovação interna e registrá-los para "acompanhamento segundo os procedimentos e as boas práticas definidos no âmbito da 3ª Região". Deve, por isso, contar com o apoio operacional da ADEG, que zelará também para que os projetos do laboratório "estejam alinhados com os objetivos estratégicos da Justiça Federal da 3ª Região, observem as regras legais e infralegais de segurança da informação e não conflitem com outros projetos conduzidos pelas áreas técnicas".

Dessas considerações decorrem algumas consequências importantes:

1) As diretrizes aqui definidas dizem respeito principalmente à auditabilidade

[37] Sobre a definição de protótipos e MVPs, veja o glossário. Ao tratar do emprego de protótipos na construção de sites da Internet (o que se aplica ao desenvolvimento de soluções computacionais em geral), diz a página Usability.gov que "é muito mais barato alterar um produto no início do processo de desenvolvimento do que fazer alterações depois de desenvolver o site. Portanto, você deve considerar a construção de protótipos no início do processo. A prototipagem permite que você reúna feedback dos usuários enquanto você ainda está planejando e projetando seu site." (Disponível em: <https://www.usability.gov/how-to-and-tools/methods/prototyping.html>. Acesso em 13 set. 2020. Sobre o uso de MVPs, Eric Ries explica que a sua utilidade está em permitir "uma volta completa do ciclo construir-medir-aprender, com o mínimo de esforço e o menor tempo de desenvolvimento" (Ries, 2012, pos. 1380). Segundo mesmo autor, "ao contrário do desenvolvimento de produto tradicional, que, em geral, envolve um período de incubação longo e ponderado e aspira à perfeição do produto, o objetivo do MVP é começar o processo de aprendizagem, não terminá-lo. Diferentemente de um protótipo ou teste de conceito, um MVP é projetado não só para responder a perguntas técnicas ou de design do produto. Seu objetivo é testar hipóteses fundamentais do negócio" (*idem*, pos. 1695). Não significa, contudo, que os MVPs sejam pouco complexos ou devam distanciar-se consideravelmente do produto final. Na verdade, não há um parâmetro definido a priori para determinar se um produto é ou não um MVP, já que "os produtos mínimos viáveis variam em complexidade, desde testes muito simples (pouco mais do que um anúncio) até protótipos iniciais reais, incluindo problemas e recursos ausentes. Uma decisão exata sobre a complexidade que um MVP precisa ter, não pode ser tomada por meio de fórmulas. É necessário julgamento" (*idem*, pos. 1736).

dos modelos de IA criados no âmbito do LIAA-3R. Não faz parte do escopo deste documento definir como deveria ser conduzida eventual auditoria nem a quem caberia realizá-la. Trata-se, em outras palavras, de documento voltado aos próprios integrantes do laboratório, exclusivamente para orientação dos trabalhos internos, sem qualquer intuito de orientar atividades de controle, de prestação de contas ou de responsabilização.

2) Não obstante, é bastante útil e necessário que os membros de equipes de projeto saibam de antemão o que deles se espera, de modo a tornar mais fácil a observância das regras vigentes e a compreensão das exigências de eventual auditoria por órgão de controle. Este documento busca atender, em parte, a essa necessidade, fornecendo algumas diretrizes práticas de verificação de conformidade, sem pretender, contudo, esgotar o assunto.

3) Tendo em vista que o LIAA-3R não desenvolverá os produtos finais, mas tão somente protótipos e MVPs, alguns dos requisitos de conformidade previstos na Resolução CNJ e Portaria CNJ, na Carta CEPEJ e nas Orientações GPAN estarão necessariamente fora do escopo da avaliação preliminar empreendida pelo GVEJ no âmbito do laboratório, especialmente no que se refere aos aspectos de desenvolvimento e implantação do produto final. Alguns exemplos são as questões relativas à segurança do sistema e dos ambientes de homologação e produção e aos mecanismos para medir e reduzir os impactos ambientais e sociais. Nada obsta, é claro, que o laboratório, com intuito de colaboração, especialmente no que se refere à gestão de riscos, antecipe aos órgãos responsáveis pela implantação eventuais problemas dessa ordem e sugira desde logo algumas medidas para minimizá-los. Tais questões sempre estarão, contudo, para além do seu controle.

4) Uma vez que o LIAA-3R não exerce atividade de controle, os seus integrantes não podem definir regras de natureza cogente, nem mesmo no âmbito do próprio laboratório. Estão, ao contrário, sujeitos às regras definidas pelos órgãos de administração e controle da Justiça Federal da 3ª Região e do Poder Judiciário como um todo, especialmente CNJ e CJF. Assim, este documento não pode servir senão como um conjunto de diretrizes não vinculativas, a serem interpretadas como recomendações e conselhos, nunca como regras de natureza cogente. Além disso, caberá aos responsáveis pelos projetos desenvolvidos no âmbito do laboratório zelar para que esses projetos recebam as aprovações necessárias perante os órgãos internos e externos de controle e cuidar de atender às exigências desses órgãos. Por conseguinte, nenhuma das diretrizes contidas neste documento pode ser interpretada de modo a alterar as determinações dos órgãos competentes, seja para limitá-las ou ampliá-las. O seu intuito é, ao contrário, auxiliar o GVEJ e os demais integrantes das equipes de projeto a cumprirem tais determinações da melhor forma possível. Logo, eventuais omissões aqui contidas não

isentam os membros das equipes de projeto da obrigação de conhecer as regras aplicáveis e de cumpri-las integralmente.

5) Sendo este documento um conjunto de diretrizes não vinculativas, ele foi concebido como um documento "vivo" e "dinâmico", a ser constantemente aprimorado a partir da experiência acumulada dos membros do LIAA-3R e dos comentários, sugestões e críticas dos leitores interessados. Em vista disso, o GVEJ receberá de bom grado quaisquer informações sobre erros ou omissões que ajudem a tornar este documento cada vez melhor.

3) Estrutura e Organização deste Documento

As diretrizes de auditabilidade e conformidade formuladas no presente documento estão organizadas em três partes.

Na primeira delas, intitulada "Diretrizes Gerais", procurou-se seguir a divisão de matérias e os preceitos da Resolução CNJ nº 332, de 21 de agosto de 2020 ("Resolução CNJ"), da Portaria do CNJ nº 271/2020 ("Portaria CNJ"), assim como do documento intitulado *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment*, aprovado na 31ª sessão plenária da CEPEJ, realizada em Estrasburgo, nos dias 3 e 4 de dezembro de 2018 ("Carta CEPEJ").

A Resolução CNJ é fortemente inspirada na Carta CEPEJ. A divisão das matérias delineada em seus Capítulos II a VII baseia-se de modo evidente nos cinco princípios éticos do documento europeu, a saber: (i) respeito pelos direitos fundamentais; (ii) não discriminação; (iii) qualidade e segurança; (iv) transparência, imparcialidade e equidade; e (v) controle do usuário. Além disso, a Resolução CNJ vale-se, por vezes, de conceitos e fórmulas traduzidos diretamente da Carta CEPEJ[38]. Assim, a leitura conjunta dos dois normativos mostra-se de grande utilidade para a correta compreensão do sentido e do alcance das disposições da Resolução CNJ.

Para analisar as regras descritas nessa parte, recorremos com frequência às Orientações Éticas para uma IA de Confiança elaboradas pelo GPAN IA ("Orientações GPAN"). Embora não se trate de documento oficial[39], ele fornece alguns subsídios

[38] A título de exemplo, compare-se os arts. 13 e 14 da Resolução CNJ com o texto do terceiro princípio ético da Carta CEPEJ (UNIÃO EUROPEIA, 2018, p. 11).

[39] As Orientações GPAN trazem as seguintes advertências na contracapa: (i) "os membros do GPAN IA nomeados no presente documento apoiam o quadro geral para uma inteligência artificial de confiança proposto nas presentes orientações, embora não concordem necessariamente com todas as afirmações contidas no documento"; e (ii) "embora o pessoal dos serviços da Comissão tenha facilitado a elaboração das orientações, as opiniões expressas no presente documento refletem o parecer do GPAN IA e não podem, em caso algum, ser consideradas como uma posição oficial da Comissão Europeia".

valiosos para a compreensão da Resolução CNJ e da Carta CEPEJ.

A finalidade precípua das Orientações GPAN é "promover uma IA de confiança", o que inclui não apenas os aspectos tecnológicos das soluções de IA, como também os processos de trabalho subjacentes ao desenvolvimento, implantação e uso das referidas soluções[40]. Segundo o referido documento, uma IA de confiança deve conter três componentes ao longo de todo o ciclo de vida do sistema: "a) deve ser **Legal**, cumprindo toda a legislação e regulamentação aplicáveis; b) deve ser **Ética**, garantindo a observância de princípios e valores éticos; c) deve ser **Sólida**, tanto do ponto de vista técnico como do ponto de vista social, uma vez que, mesmo com boas intenções, os sistemas de IA podem causar danos não intencionais"[41]. A atenção do GPAN IA recai especialmente sobre os dois últimos componentes[42].

Em cada subitem dessa primeira parte, após a reprodução dos preceitos pertinentes da Resolução CNJ, da Carta CEPEJ e da LGPD, procuramos explicar de forma sucinta como esses preceitos se aplicam às atividades desenvolvidas no âmbito do laboratório. Para compreensão dessas explicações, deve-se ter em conta as limitações de escopo expostas no item 2.

Na segunda parte, intitulada "Diretrizes Específicas", abordam-se algumas das regras aplicáveis especificamente à Justiça Federal da 3ª Região e a seus laboratórios de inovação.

A terceira e última parte cuida de algumas diretrizes relacionadas à observância da LGPD.

---

[40] "A confiança no desenvolvimento, na implantação e na utilização dos sistemas de IA diz respeito não só às propriedades inerentes à tecnologia, mas também às qualidades dos sistemas sociotécnicos que envolvem aplicações de IA. À semelhança das questões de (perda de) confiança na segurança da aviação, da energia nuclear ou dos alimentos, não são apenas as componentes do sistema de IA, mas o próprio sistema no seu contexto global, que podem, ou não, gerar confiança. Por conseguinte, os esforços para promover uma IA de confiança não só devem visar a fiabilidade do próprio sistema de IA, mas exigem uma abordagem holística e sistémica que abranja a fiabilidade de todos os intervenientes e processos que fazem parte do contexto sociotécnico do sistema ao longo do seu ciclo de vida." (UNIÃO EUROPEIA, 2019, p. 6) "A solidez de um sistema de IA abrange tanto a sua solidez técnica (adequada num determinado contexto, como o domínio de aplicação ou a fase do ciclo de vida) como a sua solidez do ponto de vista social (assegurando que o sistema de IA tem devidamente em conta o contexto e o ambiente em que o sistema opera). Este aspeto é crucial para garantir que, mesmo com boas intenções, não se podem produzir danos não intencionais. A solidez é a terceira das três componentes necessárias para alcançar uma IA de confiança." (*idem*, p. 49, § 156)

[41] *Idem*, p. 2.

[42] "As presentes orientações estabelecem um quadro para alcançar uma IA de confiança. O quadro não se ocupa explicitamente da primeira componente da IA de confiança (a IA legal). Ao invés, procura dar indicações sobre a forma de promover e assegurar uma IA ética e sólida (segunda e terceira componentes)." (*ibidem*)

III - DIRETRIZES GERAIS DE CONFORMIDADE

1) Respeito aos Direitos Fundamentais

1.1) Resolução CNJ

"Art. 4º No desenvolvimento, na implantação e no uso da Inteligência Artificial, os tribunais observarão sua compatibilidade com os Direitos Fundamentais, especialmente aqueles previstos na Constituição ou em tratados de que a República Federativa do Brasil seja parte.

Art. 5º A utilização de modelos de Inteligência Artificial deve buscar garantir a segurança jurídica e colaborar para que o Poder Judiciário respeite a igualdade de tratamento aos casos absolutamente iguais.

Art. 6º Quando o desenvolvimento e treinamento de modelos de Inteligência exigir a utilização de dados, as amostras devem ser representativas e observar as cautelas necessárias quanto aos dados pessoais sensíveis e ao segredo de justiça.

Parágrafo único. Para fins desta Resolução, são dados pessoais sensíveis aqueles assim considerados pela Lei nº 13.709/2018, e seus atos regulamentares."

1.2) Carta CEPEJ

"***1) Principle of respect for fundamental rights****: ensure that the design and implementation of artificial intelligence tools and services are compatible with fundamental rights.*

- *The processing of judicial decisions and data must serve clear purposes, in full compliance with the fundamental rights guaranteed by the European Convention on Human Rights (ECHR) and the Convention on the Protection of Personal Data (Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, ETS No. 108 as amended by the CETS amending protocol No. 223).*

- *When artificial intelligence tools are used to resolve a dispute or as a tool to assist in judicial decision-making or to give guidance to the public, it is essential to ensure that they do not undermine the guarantees of the right of access to the judge and the right to a fair trial (equality of arms and respect for the adversarial process).*
- *They should also be used with due respect for the principles of the rule of law and judges' independence in their decision-making process.*
- *Preference should therefore be given to ethical-by-design or human-rights-by-design approaches. This means that right from the design and learning phases, rules prohibiting direct or indirect violations of the fundamental values protected by the conventions are fully integrated.*"

1.3) Aplicação no Âmbito do Laboratório

O respeito aos direitos fundamentais no desenvolvimento de soluções de IA envolve o *design* do projeto (concepção, finalidade, escopo, uso esperado e tecnologias a serem utilizadas) e a seleção e uso dos *datasets*.

No que diz respeito ao *design*, uma vez que o LIAA-3R deve sempre atuar no interesse do Poder Judiciário, serão admitidos no laboratório somente os projetos que visem à promoção direta ou indireta dos direitos fundamentais. Assim, por exemplo, devem ser descartados projetos que impeçam ou dificultem o acesso do jurisdicionado ou do advogado ao juiz da causa; que impeçam ou dificultem que o juiz da causa decida com independência; que desrespeitem o cidadão, buscando tratá-lo como "coisa" a ser examinada, triada, classificada, arregimentada, condicionada ou manipulada; que desprezem ou ameacem a integridade física ou mental dos seres humanos, o seu sentido de identidade pessoal e cultural e a satisfação das suas necessidades essenciais; que desprezem ou ameacem a autonomia individual (o direito de cada indivíduo de controlar a própria vida e decidir por si mesmo, sem coerção indevida ou manipulação); que discriminem injustamente certos indivíduos ou grupos; que impeçam ou ameacem as liberdades de expressão, de crença, de empresa, de reunião ou de associação; que sejam contraditórios com os valores do Estado Democrático de Direito etc[43].

Quanto à seleção dos *datasets*, a fim de evitar resultados injustamente tendenciosos ou enviesados, as equipes de projeto devem cuidar para que as amostras sejam representativas em número e diversidade, segundo os critérios definidos pela ciência estatística, e não excluam os grupos potencialmente vulneráveis, como trabalhadores, mulheres, pessoas com deficiência, minorias étnicas, crianças, consumidores ou outras pessoas em risco de exclusão, salvo quando essa discriminação for necessária pela própria natureza do projeto e não conflite com os direitos

[43] Cf. Orientações GPAN, p. 13-14, §§ 41-45.

fundamentais desses grupos potencialmente vulneráveis, mas antes os promova.

Havendo dados sigilosos ou dados pessoais sensíveis ou de crianças e adolescentes nos *datasets*, a equipe de projeto deve solicitar autorização interna específica do CGPDP-3R antes de divulgar os seus resultados ou disponibilizar os dados ou a solução de IA em ambientes externos ao laboratório, à SETI ou ao órgão detentor dos dados. Não sendo concedida a autorização, o projeto será encerrado e os datasets terão o destino que lhes for dado pelo CGPDP-3R. No silêncio, deverão ser mantidos em ambiente de acesso restrito e não poderão mais ser utilizados no projeto no âmbito do qual foram criados ou em quaisquer outros sem a autorização do CGPDP-3R.

Os *datasets* devem ser armazenados e utilizados sempre em conformidade com as orientações da coordenação do LIAA-3R, ouvida a SETI, e as normas de segurança da informação em vigor na Justiça Federal da 3ª Região.

2) Não Discriminação

2.1) Resolução CNJ

> "Art. 7º As decisões judiciais apoiadas em ferramentas de Inteligência Artificial devem preservar a igualdade, a não discriminação, a pluralidade e a solidariedade, auxiliando no julgamento justo, com criação de condições que visem eliminar ou minimizar a opressão, a marginalização do ser humano e os erros de julgamento decorrentes de preconceitos.
>
> § 1º Antes de ser colocado em produção, o modelo de Inteligência Artificial deverá ser homologado de forma a identificar se preconceitos ou generalizações influenciaram seu desenvolvimento, acarretando tendências discriminatórias no seu funcionamento.
>
> § 2º Verificado viés discriminatório de qualquer natureza ou incompatibilidade do modelo de Inteligência Artificial com os princípios previstos nesta Resolução, deverão ser adotadas medidas corretivas.
>
> § 3º A impossibilidade de eliminação do viés discriminatório do modelo de Inteligência Artificial implicará na descontinuidade de sua utilização, com o consequente registro de seu projeto e as razões que levaram a tal decisão."

2.2) Carta CEPEJ

> "***2) Principle of non-discrimination****: specifically prevent the development or intensification of any discrimination between individuals or groups of individuals.*
>
> *Given the ability of these processing methods to reveal existing discrimination, through grouping or classifying data relating to individuals or groups of individuals, public and private stakeholders must ensure that the methods do not reproduce or aggravate such discrimination and that they do not lead to deterministic analyses or uses.*
>
> *Particular care must be taken in both the development and deployment phases, especially when*

the processing is directly or indirectly based on "sensitive" data. This could include alleged racial or ethnic origin, socio-economic background, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data, health-related data or data concerning sexual life or sexual orientation. When such discrimination has been identified, consideration must be given to corrective measures to limit or, if possible, neutralise these risks and as well as to awareness-raising among stakeholders.

However, the use of machine learning and multidisciplinary scientific analyses to combat such discrimination should be encouraged."

2.3) Aplicação no Âmbito do Laboratório

Os modelos de IA criados para apoiar a decisão judicial devem ser testados para identificação de eventuais preconceitos ou generalizações injustas, especialmente se envolverem dados pessoais sensíveis.

Sendo identificados riscos dessa ordem, deve-se buscar a realização de testes que permitam identificar eventual (i) distribuição desigual de benefícios ou custos; (ii) enviesamentos injustos; ou (iii) discriminação e estigmatização contra pessoas e grupos. Esses testes podem ser realizados no âmbito do próprio laboratório ou postergados para a fase de implementação, hipótese em que a equipe técnica de TI responsável deverá ser comunicada dos riscos identificados. Em qualquer caso, os resultados dos testes e/ou a informação sobre a necessidade de realizá-los devem ser registrados na documentação do projeto.

Os desenvolvedores podem realizar os testes que entenderem mais adequados para cada projeto, considerando sempre a sua finalidade e escopo. Devem procurar, contudo, utilizar metodologias que favoreçam a ocorrência de resultados inesperados, de modo a antecipar problemas e ajudar na identificação de novos riscos[44]. O GVEJ

[44] Cf. Orientações GPAN, p. 26-27:
"100) Devido à natureza não determinística e dependente dos contextos dos sistemas de IA, os testes tradicionais não são suficientes. As falhas dos conceitos e representações utilizados pelo sistema podem manifestar-se apenas quando um programa é aplicado a dados suficientemente realistas. Por conseguinte, para verificar e validar o tratamento dos dados, a estabilidade, a solidez e o funcionamento do modelo subjacente devem ser cuidadosamente monitorizados, dentro de limites bem compreendidos e previsíveis, tanto durante a fase de treino como durante a implantação. Tem de ser garantido que o resultado do processo de planejamento é coerente com os dados de entrada e que as decisões são tomadas de modo a permitir a validação do processo subjacente.
101) Os testes e a validação do sistema devem ser realizados o mais cedo possível, garantindo que o sistema se comporte da forma prevista ao longo de todo o seu ciclo de vida e, em especial, após a implantação. Devem incluir todas as componentes de um sistema de IA, incluindo os dados, os modelos pré-treinados, os ambientes e o comportamento do sistema em geral, e devem ser concebidos e executados por um grupo de pessoas o mais diversificado possível. Devem desenvolver-se múltiplos critérios para analisar as categorias testadas segundo diferentes perspectivas. Poderá ponderar-se a realização de testes antagônicos por «red teams» fiáveis e diversificadas, que tentem deliberadamente «penetrar» no sistema para encontrar vulnerabilidades, e a oferta de «bug bounties» que incentivam pessoas estranhas ao sistema a deletarem e comunicarem de forma responsável os erros e fragilidades do mesmo. Por último, deve assegurar-se que os seus resultados ou ações são coerentes com os

levará em conta os resultados dos testes eventualmente realizados e poderá solicitar novos testes, se entender necessário.

Para evitar que os modelos de IA apresentem enviesamento injusto, a equipe de projeto deve guardar certos cuidados metodológicos ao longo de todo o processo de desenvolvimento, dentre os quais:

a) na formação dos *datasets*, cuidar para que os dados sejam representativos, não apresentem desvios históricos inadvertidos ou lacunas e para que eventual enviesamento discriminatório identificado nessa fase seja prontamente eliminado, sempre que possível[45];

b) nos processos de supervisão, analisar e abordar de forma clara e transparente a finalidade, os condicionantes, os requisitos e as "decisões" (predições, saídas, outputs) do sistema[46];

c) na formação da equipe, buscar assegurar, tanto quanto possível, a diversidade e a multidisciplinaridade[47];

d) respeitar as normas de acessibilidade e os princípios de concepção universal, de modo a que os modelos de IA sejam acessíveis à maior variedade possível de usuários em termos de idade, sexo, raça, origem social etc.[48];

e) procurar envolver no projeto usuários internos e externos, efetivos e potenciais[49].

3) Publicidade e Transparência

3.1) Resolução CNJ

> "Art. 8º Para os efeitos da presente Resolução, transparência consiste em:
>
> I – divulgação responsável, considerando a sensibilidade própria dos dados judiciais;
>
> II – indicação dos objetivos e resultados pretendidos pelo uso do modelo de Inteligência Artificial;
>
> III – documentação dos riscos identificados e indicação dos instrumentos de segurança da

resultados dos processos precedentes, comparando-os com as políticas previamente definidas para garantir que não são violadas."

[45] *Idem*, p. 22, § 80.
[46] *Ibidem*.
[47] *Ibidem*.
[48] *Idem*, p. 23, § 81.
[49] *Idem*, p. 23, § 82.

informação e controle para seu enfrentamento;

IV – possibilidade de identificação do motivo em caso de dano causado pela ferramenta de Inteligência Artificial;

V – apresentação dos mecanismos de auditoria e certificação de boas práticas;

VI – fornecimento de explicação satisfatória e passível de auditoria por autoridade humana quanto a qualquer proposta de decisão apresentada pelo modelo de Inteligência Artificial, especialmente quando essa for de natureza judicial.

[...]

Art. 19. Os sistemas computacionais que utilizem modelos de Inteligência Artificial como ferramenta auxiliar para a elaboração de decisão judicial observarão, como critério preponderante para definir a técnica utilizada, a explicação dos passos que conduziram ao resultado.

Parágrafo único. Os sistemas computacionais com atuação indicada no caput deste artigo deverão permitir a supervisão do magistrado competente."

3.2) Portaria CNJ

"Art. 3º A pesquisa e desenvolvimento em matéria de inteligência artificial observará:

I – economicidade, evitando-se a realização de pesquisas e projetos com conteúdo semelhante em diferentes órgãos, sem colaboração e interação;

II – promoção da interoperabilidade tecnológica dos sistemas processuais eletrônicos do Poder Judiciário;

III – adoção de tecnologias, padrões e formatos abertos e livres;

IV – acesso à informação;

V – transparência;

VI – capacitação humana e sua preparação para a reestruturação dos fluxos processuais e de trabalho, à medida que a inteligência artificial é implantada;

VII – foco na celeridade processual; e

VIII – estabelecimento de mecanismos de governança colaborativa e democrática, com a participação do Poder Judiciário, daqueles que exercem funções essenciais à justiça, da comunidade acadêmica e da sociedade civil.

Art. 4º O uso de inteligência artificial no âmbito do Poder Judiciário se dará em plataforma comum, acessível por todos, que incentive a colaboração, a transparência, o aprimoramento e a divulgação dos projetos.

Parágrafo único. A plataforma de inteligência artificial do Poder Judiciário Nacional é o Sinapses, disponibilizada pelo CNJ em parceria com o Tribunal de Justiça do Estado de Rondônia."

[...]

Art. 12. Os modelos de inteligência artificial utilizados para auxiliar a atuação do Poder Judiciário na apresentação de análises, de sugestões ou de conteúdo devem adotar medidas que possibilitem o rastreamento e a auditoria das predições realizadas no fluxo de sua

aplicação.

Parágrafo único. A plataforma Sinapses provê o registro automatizado do processo de aprendizagem e consultas para cumprimento das disposições supracitadas. Os modelos devem constar da plataforma e registrar sua API em modo 'REGISTRAR PREDIÇÃO'.

Art. 13. Os sistemas judiciais que fizerem uso dos modelos de inteligência artificial devem retornar para a API registrada na plataforma a informação de eventual discordância quanto ao uso das predições, de forma que se assegure a auditoria e a melhoria dos modelos de inteligência artificial."

3.3) Carta CEPEJ

"*4**) Principle of transparency, impartiality and fairness**: make data processing methods accessible and understandable, authorise external audits.*

- A balance must be struck between the intellectual property of certain processing methods and the need for transparency (access to the design process), impartiality (absence of bias), fairness and intellectual integrity (prioritising the interests of justice) when tools are used that may have legal consequences or may significantly affect people's lives. It should be made clear that these measures apply to the whole design and operating chain as the selection process and the quality and organisation of data directly influence the learning phase.
- *The first option is complete technical transparency (for example, open source code and documentation), which is sometimes restricted by the protection of trade secrets. The system could also be explained in clear and familiar language (to describe how results are produced) by communicating, for example, the nature of the services offered, the tools that have been developed, performance and the risks of error. Independent authorities or experts could be tasked with certifying and auditing processing methods or providing advice beforehand. Public authorities could grant certification, to be regularly reviewed."*

3.4) Aplicação no Âmbito do Laboratório

Segundo as Orientações GPAN, o requisito da transparência abrange todos os elementos relevantes para uma solução de IA: os dados, o sistema e os modelos de negócio[50].

O art. 8º da Resolução CNJ parece adotar esse mesmo conceito abrangente ao incluir no requisito da transparência obrigações relacionadas aos dados (inciso I), ao sistema (incisos III a VI) e ao modelo de negócio (inciso II).

Em complementação, a Portaria CNJ determina a adoção de tecnologias, padrões e formatos abertos e livres, bem como estabelece que o SINAPSES será a plataforma oficial de inteligência artificial para fins de assegurar a transparência.

No que se refere aos dados, o termo "divulgação responsável", utilizado no

[50] *Idem*, p. 21, § 75.

inciso I do art. 8º, significa que nenhum acesso a pessoas estranhas à equipe deve ser concedido sem prévia autorização dos órgãos internos competentes. Significa, ainda, que os desenvolvedores devem armazenar e manipular os *datasets* em estrita observância às normas internas de segurança da informação, documentando as ocorrências e reportando eventuais incidentes imediatamente à SETI.

Quanto aos requisitos relacionados ao sistema, as exigências da Resolução CNJ correspondem aos conceitos de rastreabilidade, explicabilidade e auditabilidade formulados nas Orientações GPAN:

a) Rastreabilidade. "Os conjuntos de dados e os processos que produzem a decisão do sistema de IA, incluindo os processos de recolha e etiquetagem dos dados, bem como os algoritmos utilizados, devem ser documentados da melhor forma possível para permitir a rastreabilidade e um aumento da transparência. Isto também se aplica às decisões tomadas pelo sistema de IA. Deste modo, é possível identificar os motivos por que uma decisão de IA foi errada, o que, por sua vez, poderá ajudar a evitar erros futuros. A rastreabilidade facilita, assim, a auditabilidade e a explicabilidade."[51];

b) Explicabilidade. "A explicabilidade diz respeito à capacidade de explicar tanto os processos técnicos de um sistema de IA como as decisões humanas com eles relacionadas (p. ex., os domínios de aplicação de um sistema de IA). A explicabilidade técnica exige que as decisões tomadas por um sistema de IA possam ser compreendidas e rastreadas por seres humanos. Além disso, poderá ser necessário adotar soluções de compromissos entre o reforço da explicabilidade de um sistema (o que poderá reduzir a sua exatidão) ou o aumento da sua exatidão (à custa da sua explicabilidade). Sempre que um sistema de IA tenha um impacto significativo na vida das pessoas, deverá ser possível solicitar uma explicação adequada do respetivo processo de tomada de decisões. Tal explicação deve ser oportuna e adaptada ao nível de especialização da parte interessada em causa (p. ex., leigo, regulador ou investigador). Além disso, devem ser disponibilizadas explicações sobre o grau de influência e de intervenção de um sistema de IA no processo decisório da organização, as opções de conceção do sistema e os fundamentos da sua implantação (assegurando assim a transparência do modelo de negócio)."[52];

c) Auditabilidade. "A auditabilidade implica que seja possibilitada a avaliação de algoritmos, dados e processos de concepção. Tal não implica necessariamente que as informações sobre os modelos de negócios e a propriedade intelectual relacionadas com o sistema de IA tenham de estar sempre publicamente disponíveis. A avaliação por auditores internos e externos e a disponibilidade desses relatórios de avaliação podem

[51] *Idem*, p. 21-22, § 76.
[52] *Idem*, p. 22, § 77.

contribuir para a fiabilidade da tecnologia. Em aplicações que afetem os direitos fundamentais, incluindo aplicações críticas para a segurança, os sistemas de IA devem poder ser objeto de auditorias independentes."[53]

A rastreabilidade e a auditabilidade guardam íntima relação com a documentação do projeto. Significam, em primeiro lugar, que a documentação deve ser completa, incluindo *datasets*, código-fonte, resultados dos testes efetuados, métricas colhidas, requisitos, documentos de aprovação etc. Em segundo lugar, exigem que essa documentação seja acessível para o caso de ser necessário rastrear os processos pelos quais foi produzida uma decisão ou para auditoria por órgão de controle.

Importante notar que os conceitos de rastreabilidade e auditabilidade não se confundem com publicidade. Projetos totalmente rastreáveis e auditáveis podem conter documentação de acesso restrito, no todo ou em parte, seja para proteger dados pessoais, seja para fins de segurança da informação ou segurança cibernética, seja para fins de satisfazer qualquer outra exigência ética ou jurídica.

Nessa linha de raciocínio, em relação ao acesso à documentação do projeto, há duas questões importantes a serem observadas:

1º) Ao assegurar o acesso aos *datasets* para fins de rastreabilidade e auditabilidade, a equipe de projeto deve cuidar para que isso não se dê em contrariedade com as normas de segurança da informação ou com a LGPD. Para tanto, deve procurar orientação da SETI quanto ao modo e local de armazenamento e aos meios de acesso. Se houver dados sigilosos ou dados pessoais sensíveis ou de crianças ou adolescentes, deve buscar também prévia aprovação do CGPDP-3R, conforme mencionado anteriormente no item 1, subitem 1.3, que tratou do respeito aos direitos fundamentais.

2º) O pleno acesso ao código-fonte exigirá da equipe de desenvolvimento atenção especial à escolha das dependências, inclusive quanto à sua licença de uso, que deve ser compatível com a licença do código-fonte e com os requisitos de rastreabilidade e auditabilidade aqui mencionados.

A explicabilidade diz respeito à capacidade de tornar claro e compreensível o funcionamento de uma solução de IA para um determinado público-alvo, mediante o fornecimento, em linguagem adequada para o referido público-alvo, dos detalhes ou razões pelas quais a solução apresentou determinada decisão (um determinado

[53] *Idem*, p. 24, § 88.

output[54]) e não outra[55].

Segundo o Instituto Alan Turing (The Alan Turing Institute), existem seis formas principais de explicar uma decisão de IA[56]:

1) Explicação por justificativa: fornecer as razões que levaram à decisão, em linguagem acessível, não técnica.

2) Explicação por responsabilidade: indicar as pessoas envolvidas no desenvolvimento, gestão e implementação da solução de IA e quem contactar para pedir a revisão da decisão.

3) Explicação pelos dados: indicar quais dados foram usados e como foram usados em uma decisão específica, assim como no treinamento e nos testes da solução de IA.

4) Explicação por equidade (*fairness*): indicar quais as providências adotadas durante o desenvolvimento e a implementação da solução de IA para assegurar que as decisões contempladas são equânimes e não enviesadas e para dizer se um usuário foi ou não tratado com isonomia.

5) Explicação por segurança e performance: indicar quais as providências adotadas durante o desenvolvimento e a implementação da solução de IA para maximizar a acurácia, confiabilidade, segurança e robustez das decisões e

[54] Segundo explicado no guia prático do Instituto Alan Turing, o termo "decisão de IA", quando utilizado numa acepção ampla, envolve *outputs* de qualquer natureza (predições, recomendações ou classificações) utilizados tanto em processos automatizados como em semi-automatizados: "*So, an AI decision can be based on a prediction, a recommendation or a classification. It can also refer to a solely automated process, or one in which a human is involved*" (cf. The Alan Turing Institute, parte 1, p. 6).

[55] Essa definição baseia-se na que foi proposta por BARREDO ARRIETA et al.: "*Given an audience, an explainable Artificial Intelligence is one that produces details or reasons to make its functioning clear or easy to understand*" (p. 85). Apesar dos intensos debates ainda existentes em torno desse conceito, a definição mostra-se, a nosso ver, suficiente para os propósito práticos deste manual.

[56] "*There are different ways of explaining AI decisions. We have identified six main types of explanation:*
• Rationale explanation: the reasons that led to a decision, delivered in an accessible and non-technical way.
• Responsibility explanation: who is involved in the development, management and implementation of an AI system, and who to contact for a human review of a decision.
• Data explanation: what data has been used in a particular decision and how; what data has been used to train and test the AI model and how.
• Fairness explanation: steps taken across the design and implementation of an AI system to ensure that the decisions it supports are generally unbiased and fair, and whether or not an individual has been treated equitably.
• Safety and performance explanation: steps taken across the design and implementation of an AI system to maximise the accuracy, reliability, security and robustness of its decisions and behaviours.
• Impact explanation: the impact that the use of an AI system and its decisions has or may have on an individual, and on wider society." (cf. THE ALAN TURING INSTITUTE, p. 19).

comportamentos.

6) Explicação pelo impacto: indicar o impacto que o uso da solução de IA e suas decisões podem ter sobre um indivíduo e sobre a sociedade em geral.

De modo semelhante, BARREDO ARRIETA et al. sugerem as seguintes orientações metodológicas para assegurar a explicabilidade de soluções e IA[57]:

1) Levar em conta o contexto, impactos potenciais e necessidades específicas do domínio do problema, o que inclui compreender (i) a finalidade do projeto; (ii) a complexidade das explicações exigidas pelo público-alvo; e (iii) a performance e o grau de interpretabilidade da tecnologia, dos modelos e métodos existentes.

[57] BARREDO ARRIETA et al., p. 102-103: "*Given the confluence of multiple criteria and the need for having the human in the loop, some attempts at establishing the procedural guidelines to implement and explain AI systems have been recently contributed. Among them, we pause at the thorough study in [383], which suggests that the incorporation and consideration of explainability in practical AI design and deployment workflows should comprise four major methodological steps:*
1. Contextual factors, potential impacts and domain-specific needs must be taken into account when devising an approach to interpretability: These include a thorough understanding of the purpose for which the AI model is built, the complexity of explanations that are required by the audience, and the performance and interpretability levels of existing technology, models and methods. The latter pose a reference point for the AI system to be deployed in lieu thereof.
2. Interpretable techniques should be preferred when possible: when considering explainability in the development of an AI system, the decision of which XAI approach should be chosen should gauge domain-specific risks and needs, the available data resources and existing domain knowledge, and the suitability of the ML model to meet the requirements of the computational task to be addressed. It is in the confluence of these three design drivers where the guidelines postulated in [383] (and other studies in this same line of thinking [384]) recommend first the consideration of standard interpretable models rather than sophisticated yet opaque modeling methods. In practice, the aforementioned aspects (contextual factors, impacts and domain-specific needs) can make transparent models preferable over complex modeling alternatives whose interpretability require the application of post-hoc XAI techniques. By contrast, black-box models such as those reviewed in this work (namely, support vector machines, ensemble methods and neural networks) should be selected only when their superior modeling capabilities fit best the characteristics of the problem at hand.
3. If a black-box model has been chosen, the third guideline establishes that ethics-, fairness- and safety-related impacts should be weighed. Specifically, responsibility in the design and implementation of the AI system should be ensured by checking whether such identified impacts can be mitigated and counteracted by supplementing the system with XAI tools that provide the level of explainability required by the domain in which it is deployed. To this end, the third guideline suggests 1) a detailed articulation, examination and evaluation of the applicable explanatory strategies, 2) the analysis of whether the coverage and scope of the available explanatory approaches match the requirements of the domain and application context where the model is to be deployed; and 3) the formulation of an interpretability action plan that sets forth the explanation delivery strategy, including a detailed time frame for the execution of the plan, and a clearance of the roles and responsibilities of the team involved in the workflow.
4. Finally, the fourth guideline encourages to rethink interpretability in terms of the cognitive skills, capacities and limitations of the individual human. This is an important question on which studies on measures of explainability are intensively revolving by considering human mental models, the accessibility of the audience to vocabularies of explanatory outcomes, and other means to involve the expertise of the audience into the decision of what explanations should provide."

2) Preferir modelos de IA transparentes[58] sempre que possível, levando em conta os riscos e as necessidades envolvidos, os dados disponíveis, o conhecimento existente e a adequação do modelo de aprendizagem de máquina para a solução do problema computacional a ser resolvido. Modelos opacos, como máquinas de vetores de suporte (*support vector machines)*, métodos de composição (*ensemble)* e redes neurais profundas (*deep learning*) devem ser selecionados somente quando se mostrarem mais adequados à solução do problema.

3) Ao selecionar-se uma solução "caixa preta", deve-se ter atenção redobrada aos potenciais impactos relacionados à ética, equidade e segurança, mediante avaliação cuidadosa das estratégias de explicação, daquilo que deve ser explicado e de como e quando a explicação deve ser comunicada ao público-alvo.

4) As explicações devem ser formuladas levando em conta o indivíduo destinatário, com suas habilidades, capacidades e limitações.

Verifica-se, portanto, que a explicabilidade não envolve apenas o funcionamento da solução de IA em si, mas o contexto mais amplo em que o modelo foi concebido, desenvolvido e implementado e é utilizado[59].

Ademais, apesar dos benefícios inegáveis da explicabilidade tanto para os usuários finais quanto para os próprios desenvolvedores e demais partes interessadas (*stakeholders)*[60], nem sempre é viável atingir esse ideal com plenitude, especialmente

[58] Aqui, utilizamos o conceito de transparência tal como empregado no texto de BARREDO ARRIETA et al.: "*A model is considered to be transparent if by itself it is understandable. Since a model can feature different degrees of understandability, transparent models in Section 3 are divided into three categories: simulatable models, decomposable models and algorithmically transparent models*" (p. 85). Nessa acepção, transparência se opõe a opacidade ou "black-boxness".

[59] "*The Cambridge dictionary defines 'explanation' as: 'The details or reasons that someone gives to make something clear or easy to understand.' While this is a general definition, it remains valid when considering how to explain AI-assisted decisions to the individuals affected (who are often also data subjects). It suggests that you should not always approach explanations in the same way. What people want to understand, and the 'details' or 'reasons' that make it 'clear' or 'easy' for them to do so may differ. Our own research, and that of others, reveals that context is a key aspect of explaining decisions involving AI. Several factors about the decision, the person, the application, the type of data, and the setting, all affect what information an individual expects or finds useful. Therefore, when we talk about explanations in this guidance, we do not refer to just one approach to explaining decisions made with the help of AI, or providing a single type of information to affected individuals. Instead, the context affects which type of explanation you use to make an AI-assisted decision clear or easy for individuals to understand.*" (cf. THE ALAN TURING INSTITUTE, p. 19-20).

[60] Segundo BARREDO ARRIETA et al., a literatura sobre o assunto tem identificado os seguintes benefícios promovidos pela explicabilidade (cf. p. 86-87): (i) aumento da confiabilidade da solução de IA (*trustworthiness*); (ii) capacidade de descobrir relações de causalidade entre as variáveis dos *datasets* (*causality*); (iii) capacidade de transferir conhecimento entre soluções de IA (*transferability*); (iv) capacidade de obter informações sobre a estratégia utilizada pela solução de IA para resolver o problema proposto (*informativeness*); (v) aumento da robustez e da estabilidade da solução de IA (*confidence*); (vi) maior equidade dos resultados obtidos (*fairness*); (vii) maior envolvimento dos usuários finais no aperfeiçoamento da solução de IA (*accessibility*); (viii) maior interação entre os usuários finais e a solução

quando a solução envolve o uso de modelos opacos, como os que são criados com técnicas de aprendizagem profunda. Existem, portanto, diferentes graus de explicabilidade de soluções de IA e o grau de explicabilidade aceitável para cada projeto irá depender essencialmente dos riscos e dos benefícios envolvidos, segundo critérios de razoabilidade[61]. Ademais, existem técnicas de explicação indireta que podem ser utilizadas pelas equipes de projeto para assegurar um nível mínimo de explicabilidade mesmo quando utilizados modelos opacos[62].

---

de IA (*interactivity*); (ix) maior consciência sobre como os dados pessoais dos usuários são processados pela solução de IA, ajudando a evitar potenciais violações de privacidade (*privacy awareness*).

[61] Cf. Orientações GPAN, p. 16: "53) A explicabilidade é crucial para criar e manter a confiança dos utilizadores nos sistemas de IA. Tal significa que os processos têm de ser transparentes, as capacidades e a finalidade dos sistemas de IA abertamente comunicadas e as decisões — tanto quanto possível — explicáveis aos que são por elas afetados de forma direta e indireta. Sem essas informações, não é possível contestar devidamente uma decisão. Nem sempre é possível explicar por que razão um modelo gerou determinado resultado ou decisão (e que combinação de fatores de entrada contribuiu para esse efeito). Estes casos são designados por algoritmos de «caixa negra» e exigem especial atenção. Nessas circunstâncias, podem ser necessárias outras medidas da explicabilidade (p. ex., a rastreabilidade, a auditabilidade e a comunicação transparente sobre as capacidades do sistema), desde que o sistema, no seu conjunto, respeite os direitos fundamentais. O grau de necessidade da explicabilidade depende em grande medida do contexto e da gravidade das consequências de um resultado errado ou inexato."

[62] Essas técnicas, denominadas de explicabilidade "post-hoc" por BARREDO ARRIETA et al., compreendem as seguintes modalidades (cf. p. 88):

"• *Text explanations deal with the problem of bringing explainability for a model by means of learning to generate text explanations that help explaining the results from the model. Text explanations also include every method generating symbols that represent the functioning of the model. These symbols may portrait the rationale of the algorithm by means of a semantic mapping from model to symbols.*

• *Visual explanation techniques for post-hoc explainability aim at visualizing the model's behavior. Many of the visualization methods existing in the literature come along with dimensionality reduction techniques that allow for a human interpretable simple visualization. Visualizations may be coupled with other techniques to improve their understanding, and are considered as the most suitable way to introduce complex interactions within the variables involved in the model to users not acquainted to ML modeling.*

• *Local explanations tackle explainability by segmenting the solution space and giving explanations to less complex solution subspaces that are relevant for the whole model. These explanations can be formed by means of techniques with the differentiating property that these only explain part of the whole system's functioning.*

• *Explanations by example consider the extraction of data examples that relate to the result generated by a certain model, enabling to get a better understanding of the model itself. Similarly to how humans behave when attempting to explain a given process, explanations by example are mainly centered in extracting representative examples that grasp the inner relationships and correlations found by the model being analyzed.*

• *Explanations by simplification collectively denote those techniques in which a whole new system is rebuilt based on the trained model to be explained. This new, simplified model usually attempts at optimizing its resemblance to its antecedent functioning, while reducing its complexity, and keeping a similar performance score. An interesting byproduct of this family of post-hoc techniques is that the simplified model is, in general, easier to be implemented due to its reduced complexity with respect to the model it represents.*

• *Finally, feature relevance explanation methods for post-hoc explain ability clarify the inner functioning of a model by computing a relevance score for its managed variables. These scores quantify the affection (sensitivity) a feature has upon the output of the model. A comparison of the scores among different variables unveils the importance granted by the model to each of such variables when producing its output. Feature relevance methods can be thought to be an indirect method to explain a model.*"

Em termos práticos, para assegurar que as soluções de IA atendam à exigência de explicabilidade em grau adequado, as equipes de projeto devem definir previamente e fazer incluir na documentação:

a) o escopo e a finalidade da solução de IA, descrevendo qual o problema que se pretendeu resolver e por quê;

b) os benefícios esperados que motivaram e justificaram o projeto, se e por que tais benefícios não poderiam ser obtidos por outros meios;

c) o grupo de usuários e o contexto de uso para os quais a solução se destina.

Qualquer alteração nos elementos acima deve ser formalizada na documentação do projeto e submetida a nova aprovação interna.

Além disso, ao longo do desenvolvimento devem ser também documentadas (i) as razões para a adoção das técnicas, ferramentas e *datasets* utilizados, mencionando eventuais alternativas, e porque estão em linha com o escopo e a finalidade do projeto; e (ii) os riscos mapeados para o caso de ocorrer um resultado errado ou inexato, assim como a gravidade das consequências daí decorrentes, levando em conta o uso e o contexto de uso para os quais a solução foi concebida.

Embora não seja responsável pela posterior implementação da solução em ambiente de produção, a equipe de projeto deve manter postura colaborativa, entregando aos responsáveis pela implementação um sumário dos elementos de explicabilidade a serem observados no uso da solução, do qual devem constar:

a) o escopo e a finalidade da solução de IA;

b) o uso e o contexto de uso recomendados, incluindo os tipos de usuários a que a solução se destina;

c) as capacidades e as limitações conhecidas da solução de IA;

d) os riscos mapeados e a gravidade das consequências em caso de resultados errados ou inexatos;

e) as informações a serem dadas aos usuários antes e durante o uso da solução de IA[63].

Sobre o grau de explicabilidade de cada tipo de algoritmo de *machine learning*, ver BARREDO ARRIETA et al., p. 90, tabela 2.

[63] Por exemplo, sempre que houver risco de que o usuário pense interagir com um ser humano quando está na verdade interagindo apenas com o sistema, deve-se dar informações claras a esse respeito ao usuário, oferecendo-lhe, ainda, a opção de comunicar-se com um ser humano se assim desejar. Acerca

4) Governança, Qualidade e Segurança

4.1) Resolução CNJ

> "Art. 9º Qualquer modelo de Inteligência Artificial que venha a ser adotado pelos órgãos do Poder Judiciário deverá observar as regras de governança de dados aplicáveis aos seus próprios sistemas computacionais, as Resoluções e as Recomendações do Conselho Nacional de Justiça, a Lei no 13.709/2018, e o segredo de justiça.
>
> Art. 10. Os órgãos do Poder Judiciário envolvidos em projeto de Inteligência Artificial deverão:
>
> I – informar ao Conselho Nacional de Justiça a pesquisa, o desenvolvimento, a implantação ou o uso da Inteligência Artificial, bem como os respectivos objetivos e os resultados que se pretende alcançar;
>
> II – promover esforços para atuação em modelo comunitário, com vedação a desenvolvimento paralelo quando a iniciativa possuir objetivos e resultados alcançados idênticos a modelo de Inteligência Artificial já existente ou com projeto em andamento;
>
> III – depositar o modelo de Inteligência Artificial no Sinapses.
>
> Art. 11. O Conselho Nacional de Justiça publicará, em área própria de seu sítio na rede mundial de computadores, a relação dos modelos de Inteligência Artificial desenvolvidos ou utilizados pelos órgãos do Poder Judiciário.
>
> Art. 12. Os modelos de Inteligência Artificial desenvolvidos pelos órgãos do Poder Judiciário deverão possuir interface de programação de aplicativos (API) que permitam sua utilização por outros sistemas.
>
> Parágrafo único. O Conselho Nacional de Justiça estabelecerá o padrão de interface de programação de aplicativos (API) mencionado no caput deste artigo.
>
> [...]
>
> Art. 13. Os dados utilizados no processo de treinamento de modelos de Inteligência Artificial deverão ser provenientes de fontes seguras, preferencialmente governamentais.
>
> Art. 14. O sistema deverá impedir que os dados recebidos sejam alterados antes de sua utilização nos treinamentos dos modelos, bem como seja mantida sua cópia (dataset) para cada versão de modelo desenvolvida.
>
> Art. 15. Os dados utilizados no processo devem ser eficazmente protegidos contra os riscos de

desse tema, dizem as Orientações GPAN (p. 22): "78) Os sistemas da IA não se devem apresentar como seres humanos aos utilizadores; os seres humanos têm direito a serem informados de que estão a interagir com um sistema de IA. Tal implica que os sistemas de IA devem ser identificáveis como tal. Além disso, deve ser facultada a opção de decidir contra essa interação a favor da interação humana, sempre que necessário, a fim de garantir que os direitos fundamentais são respeitados. Além disso, as capacidades e limitações do sistema de IA devem ser comunicadas aos profissionais no domínio da IA ou aos utilizadores finais de forma adequada ao caso de utilização em questão. Essa comunicação poderá incluir o nível de exatidão do sistema de IA, bem como as suas limitações."

destruição, modificação, extravio ou acessos e transmissões não autorizados.

Art. 16. O armazenamento e a execução dos modelos de Inteligência Artificial deverão ocorrer em ambientes aderentes a padrões consolidados de segurança da informação."

4.2) Carta CEPEJ

"***3) Principle of quality and security**: with regard to the processing of judicial decisions and data, use certified sources and intangible data with models conceived in a multi-disciplinary manner, in a secure technological environment.*

- *Designers of machine learning models should be able to draw widely on the expertise of the relevant justice system professionals (judges, prosecutors, lawyers, etc.) and researchers/lecturers in the fields of law and social sciences (for example, economists, sociologists and philosophers).*
- *Forming mixed project teams in short design cycles to produce functional models is one of the organisational methods making it possible to capitalise on this multidisciplinary approach.*
- *Existing ethical safeguards should be constantly shared by these project teams and enhanced using feedback.*
- *Data based on judicial decisions that is entered into a software which implements a machine learning algorithm should come from certified sources and should not be modified until they have actually been used by the learning mechanism. The whole process must therefore be traceable to ensure that no modification has occurred to alter the content or meaning of the decision being processed.*
- *The models and algorithms created must also be able to be stored and executed in secure environments, so as to ensure system integrity and intangibility.*"

4.3) Aplicação no Âmbito do Laboratório

Nos projetos do LIAA-3R devem ser observadas as regras de governança de dados e de sistemas computacionais em vigor para a Justiça Federal da 3ª Região.

Além disso, a equipe de projeto deve adotar as seguintes cautelas para preservar a qualidade e segurança dos dados utilizados na criação de modelos de IA, solicitando, sempre que necessário, o auxílio da SETI:

1) utilizar fontes de dados seguras e de qualidade[64], preferencialmente

[64] Segundo o "*Data Management Body of Knowledge*" editado pela DAMA – Data Management Association (DAMA-DMBOK) (apud LIMA, 2019, p. 143-144), os aspectos a serem considerados na aferição da qualidade dos dados são os seguintes: (i) acurácia: quão próximos estão de representar as entidades reais; (ii) completude: se são ou não suficientes para fornecer a informação de que se necessita num dado

governamentais;

2) definir meios adequados para que os dados utilizados sejam protegidos contra os riscos de destruição, alteração, distorção[65], extravio ou acessos e transmissões não autorizados;

3) definir a estrutura, a composição e o modo de armazenamento, preservação da integridade e proteção dos *datasets*[66]; e

4) obter aprovação do CGPDP-3R antes de iniciar o tratamento de dados sigilosos, dados pessoais sensíveis ou de crianças ou adolescentes[67].

Nos casos em que o LIAA-3R participar da implementação da solução de IA para uso em ambientes de homologação ou produção, a equipe de projeto deve também elaborar, em conjunto com a SETI, plano de gestão de dados para todo o ciclo de vida da solução de IA, plano este que deve envolver, além das cautelas acima, os seguintes aspectos adicionais[68]:

---

contexto; (iii) consistência: se apresentam integridade e coerência quando confrontados com outras fontes; (iv) atualidade: se correspondem ao estado atual de coisas; (v) precisão: se refletem ou não o grau de precisão exigido em cada contexto (em casas decimais, por exemplo); (vi) privacidade: se atendem as normas de privacidade e sigilo; (vii) razoabilidade: se são produzidos de modo consistente com as expectativas gerenciais; (viii) integridade referencial: se atendem aos parâmetros de integridade necessários para que sejam considerados confiáveis; (ix) unicidade: se são gerados por uma única "fonte da verdade"; e (x) validade: correspondem ao tipo ou formato adequado.

[65] Exemplificando, os dados recebidos deverão ser protegidos contra alterações mal intencionadas com o objetivo de enviesar o resultado do algoritmo.

[66] Os *datasets* de treinamento, validação e testes integram a documentação do projeto e devem ser preservados para assegurar a rastreabilidade dos resultados obtidos e a auditabilidade da solução como um todo.

[67] No caso de dados pessoais, a autorização pode ser dispensada quando houver anonimização como resultado do processo de tratamento de dados (cf. item IV-3 abaixo).

[68] Cf. Orientações GPAN, p. 21:

"71) Estreitamente ligado ao princípio de prevenção de danos está o direito à privacidade, um direito fundamental que é particularmente afetado pelos sistemas de IA. A prevenção da ameaça à privacidade também exige uma governação adequada dos dados, que assegure a qualidade e a integridade dos dados utilizados, a sua relevância para o domínio em que os sistemas de IA serão implantados, os seus protocolos de acesso e a capacidade de tratar os dados de modo a proteger a privacidade.

72) Privacidade e proteção de dados. Os sistemas de IA devem garantir a privacidade e a proteção de dados ao longo de todo o ciclo de vida de um sistema. Tal inclui as informações inicialmente fornecidas pelo utilizador, bem como as informações produzidas sobre o utilizador ao longo da sua interação com o sistema (p. ex., os resultados gerados pelo sistema de IA para utilizadores específicos ou a forma como os utilizadores responderam a determinadas recomendações). Os registos digitais do comportamento humano podem permitir que os sistemas de IA infiram não só as preferências dos indivíduos, mas também a sua orientação sexual, a sua idade e as suas convicções religiosas ou políticas. Para que as pessoas possam confiar no processo de recolha de dados, deve ser garantido que os dados recolhidos a seu respeito não serão utilizados para as discriminar de forma ilegal ou injusta.

73) Qualidade e integridade dos dados. A qualidade dos conjuntos de dados utilizados é fundamental para o desempenho dos sistemas de IA. Quando são recolhidos, os dados podem conter enviesamentos socialmente construídos, inexatidões, erros e enganos. Esta questão tem de ser resolvida antes de se

1) Ter respeito à privacidade dos usuários, recolhendo e preservando somente os dados pessoais estritamente necessários para o funcionamento da solução desenvolvida.

2) Zelar pela qualidade e integridade dos dados pessoais recolhidos de usuários, mediante o uso de técnicas e ferramentas que ajudem a identificar enviesamentos, inexatidões, erros e enganos, assim como dados maliciosos introduzidos no sistema.

3) Respeitar os protocolos de acesso aos dados definidos pela SETI.

Uma vez aprovado o projeto internamente, a equipe deve providenciar sua inscrição no CNJ e, ao final dos trabalhos, depositar os modelos de IA no SINAPSES. Tanto quanto possível, deve também atender aos padrões e boas práticas definidos pelo CNJ para a criação, armazenamento e proteção dos *datasets*, codificação, versionamento e entrega dos algoritmos, construção das APIs etc.

5) Controle do Usuário

5.1) Resolução CNJ

> "Art. 17. O sistema inteligente deverá assegurar a autonomia dos usuários internos, com uso de modelos que:
>
> I – proporcione incremento, e não restrição;
>
> II – possibilite a revisão da proposta de decisão e dos dados utilizados para sua elaboração, sem que haja qualquer espécie de vinculação à solução apresentada pela Inteligência Artificial.
>
> Art. 18. Os usuários externos devem ser informados, em linguagem clara e precisa, quanto à utilização de sistema inteligente nos serviços que lhes forem prestados.
>
> Parágrafo único. A informação prevista no caput deve destacar o caráter não vinculante da proposta de solução apresentada pela Inteligência Artificial, a qual sempre é submetida à análise da autoridade competente.

---

treinar o sistema com um determinado conjunto de dados. Além disso, há que assegurar a integridade dos dados. A introdução de dados maliciosos num sistema de IA pode alterar o seu comportamento, em especial no caso dos sistemas com autoaprendizagem. Os processos e conjuntos de dados utilizados devem ser testados e documentados em cada uma das etapas, nomeadamente de planeamento, treino, ensaio e implantação. O mesmo se aplica aos sistemas de IA que não foram desenvolvidos a nível interno, mas sim adquiridos externamente.

74) Acesso aos dados. Em qualquer organização que trate dados pessoais (independentemente de pertencerem a um utilizador ou a um não utilizador do sistema), devem ser adotados protocolos de governação do acesso aos dados. Estes protocolos devem indicar quem pode aceder aos dados e em que circunstâncias o pode fazer. O acesso a dados pessoais só deverá ser permitido a pessoal devidamente qualificado, que tenha competência e necessidade de aceder aos mesmos."

> Art. 19. Os sistemas computacionais que utilizem modelos de Inteligência Artificial como ferramenta auxiliar para a elaboração de decisão judicial observarão, como critério preponderante para definir a técnica utilizada, a explicação dos passos que conduziram ao resultado.
>
> Parágrafo único. Os sistemas computacionais com atuação indicada no caput deste artigo deverão permitir a supervisão do magistrado competente."

5.2) Carta CEPEJ

> "***5) Principle 'under user control'**: preclude a prescriptive approach and ensure that users are informed actors and in control of their choices.*
>
> - *User autonomy must be increased and not restricted through the use of artificial intelligence tools and services.*
> - *Professionals in the justice system should, at any moment, be able to review judicial decisions and the data used to produce a result and continue not to be necessarily bound by it in the light of the specific features of that particular case.*
> - *The user must be informed in clear and understandable language whether or not the solutions offered by the artificial intelligence tools are binding, of the different options available, and that s/he has the right to legal advice and the right to access a court. S/he must also be clearly informed of any prior processing of a case by artificial intelligence before or during a judicial process and have the right to object, so that his/her case can be heard directly by a court within the meaning of Article 6 of the ECHR.*
> - *Generally speaking, when any artificial intelligence-based information system is implemented there should be computer literacy programmes for users and debates involving professionals from the justice system.*"

5.3) Aplicação no Âmbito do Laboratório

Os arts. 17 a 19 da Resolução CNJ e o 5º princípio da Carta CEPEJ visam à proteção da autonomia individual.

Segundo explicam as Orientações GPAN, "os sistemas de IA devem ajudar os indivíduos a fazerem escolhas mais corretas e fundamentadas em conformidade com os seus objetivos", o que se opõe a qualquer tentativa de "moldar e influenciar o comportamento humano", mediante "manipulação, engano, arregimentação e condicionamento". Por conseguinte, "o princípio geral da autonomia do utilizador deve estar no centro da funcionalidade do sistema"[69].

A proteção da autonomia individual se dá por meio da supervisão humana, a

[69] Orientações GPAN, pág. 19, § 64.

qual pode ocorrer segundo três modelos de intervenção sobre o sistema de IA[70]:

a) *human-in-the-loop — HITL*, no qual o ser humano intervém em todos os ciclos de decisão do sistema;

b) *human-on-the-loop — HOTL*, no qual o ser humano intervém apenas nos ciclos de concepção e de acompanhamento do funcionamento do sistema; e

c) *human-in-command — HIC*, no qual o ser humano intervém em toda a atividade do sistema de IA, podendo "decidir quando e como utilizar o sistema em qualquer situação específica" e até mesmo optar por "não utilizar um sistema de IA numa determinada situação, de estabelecer níveis de apreciação humana durante a utilização do sistema, ou de assegurar a capacidade de anular uma decisão tomada por um sistema".

A Resolução CNJ parece ter inspiração no terceiro modelo ou alguma variante dele. É possível resumir os seus preceitos relacionados a esse tema em quatro regras de validação ético-jurídica:

1ª) A solução de IA nunca deve restringir a autonomia decisória do ser humano.

2ª) A solução de IA deve sempre permitir que o ser humano rejeite por completo a proposta de decisão por ela apresentada, sem qualquer espécie de vinculação.

3ª) A solução de IA deve conferir a mesma autonomia aos usuários externos, informando-lhes sobre a natureza inteligente do sistema e sobre o caráter não vinculativo da proposta de decisão apresentada, submetendo a proposta também à análise da autoridade competente.

4ª) As soluções de IA destinadas a auxiliar na elaboração de decisão judicial devem ser explicáveis, preferencialmente pela enumeração dos passos que conduziram ao resultado, e estar submetidas à supervisão do magistrado competente.

Essa última regra está ligada também à explicabilidade, de que falamos no item 3 ("Publicidade e Transparência").

6) Pesquisa, Desenvolvimento e Implantação de Serviços de IA

[70] Cf. Orientações GPAN, p. 19, § 65.

6.1) Resolução CNJ

"Art. 20. A composição de equipes para pesquisa, desenvolvimento e implantação das soluções computacionais que se utilizem de Inteligência Artificial será orientada pela busca da diversidade em seu mais amplo espectro, incluindo gênero, raça, etnia, cor, orientação sexual, pessoas com deficiência, geração e demais características individuais.

§ 1º A participação representativa deverá existir em todas as etapas do processo, tais como planejamento, coleta e processamento de dados, construção, verificação, validação e implementação dos modelos, tanto nas áreas técnicas como negociais.

§ 2º A diversidade na participação prevista no caput deste artigo apenas será dispensada mediante decisão fundamentada, dentre outros motivos, pela ausência de profissionais no quadro de pessoal dos tribunais.

§ 3º As vagas destinadas à capacitação na área de Inteligência Artificial serão, sempre que possível, distribuídas com observância à diversidade.

§ 4º A formação das equipes mencionadas no caput deverá considerar seu caráter interdisciplinar, incluindo profissionais de Tecnologia da Informação e de outras áreas cujo conhecimento científico possa contribuir para pesquisa, desenvolvimento ou implantação do sistema inteligente.

Art. 21. A realização de estudos, pesquisas, ensino e treinamentos de Inteligência Artificial deve ser livre de preconceitos, sendo vedado:

I – desrespeitar a dignidade e a liberdade de pessoas ou grupos envolvidos em seus trabalhos;

II – promover atividades que envolvam qualquer espécie de risco ou prejuízo aos seres humanos e à equidade das decisões;

III – subordinar investigações a sectarismo capaz de direcionar o curso da pesquisa ou seus resultados.

Art. 22. Iniciada pesquisa, desenvolvimento ou implantação de modelos de Inteligência Artificial, os tribunais deverão comunicar imediatamente ao Conselho Nacional de Justiça e velar por sua continuidade.

§ 1º As atividades descritas no caput deste artigo serão encerradas quando, mediante manifestação fundamentada, for reconhecida sua desconformidade com os preceitos éticos estabelecidos nesta Resolução ou em outros atos normativos aplicáveis ao Poder Judiciário e for inviável sua readequação.

§ 2º Não se enquadram no caput deste artigo a utilização de modelos de Inteligência Artificial que utilizem técnicas de reconhecimento facial, os quais exigirão prévia autorização do Conselho Nacional de Justiça para implementação.

Art. 23. A utilização de modelos de Inteligência Artificial em matéria penal não deve ser estimulada, sobretudo com relação à sugestão de modelos de decisões preditivas.

§ 1º Não se aplica o disposto no caput quando se tratar de utilização de soluções computacionais destinadas à automação e ao oferecimento de subsídios destinados ao cálculo de penas, prescrição, verificação de reincidência, mapeamentos, classificações e triagem dos autos para fins de gerenciamento de acervo.

§ 2º Os modelos de Inteligência Artificial destinados à verificação de reincidência penal não devem indicar conclusão mais prejudicial ao réu do que aquela a que o magistrado chegaria sem sua utilização.

Art. 24. Os modelos de Inteligência Artificial utilizarão preferencialmente software de código aberto que:

I – facilite sua integração ou interoperabilidade entre os sistemas utilizados pelos órgãos do Poder Judiciário;

II – possibilite um ambiente de desenvolvimento colaborativo;

III – permita maior transparência;

IV – proporcione cooperação entre outros segmentos e áreas do setor público e a sociedade civil."

6.2) Portaria CNJ

"Art. 10. O desenvolvimento de modelos de inteligência artificial no âmbito do Poder Judiciário deverá ser feito pela plataforma oficial de disponibilização de modelos de inteligência artificial.

§ 1º O Sinapses é a plataforma oficial de disponibilização de modelos de inteligência artificial e está disponível no endereço <https://sinapses.ia.pje.jus.br/>.

§ 2º O desenvolvimento de modelos de inteligência artificial no âmbito do Poder Judiciário deverá respeitar as diretrizes previstas na Resolução CNJ nº 332/2020 e o disposto nesta normatização, sendo obrigatória a comunicação ao Conselho Nacional de Justiça.

Art. 11. O desenvolvimento e registro de modelos na plataforma deve ser precedido da instalação do módulo extrator para assegurar que os dados que lhe servem de base constem do repositório central, englobando a capa do processo judicial (metadados), suas movimentações processuais e os documentos devidamente convertidos em formato de texto simples.

§ 1º Os dados utilizados para treinamento no modelo devem estar disponibilizados junto aos recursos do modelo.

§ 2º É responsabilidade do órgão criador e/ou mantenedor de cada modelo de inteligência artificial a adoção de medidas, durante o processo de disponibilização de dados, que assegurem a preservação do sigilo e do segredo de justiça, adotando-se quanto aos dados sensíveis, medidas de ocultação ou anonimização."

6.3) Aplicação no Âmbito do Laboratório

Os preceitos contidos nos arts. 20 a 24 não têm similar na Carta CEPEJ. Tratam de requisitos específicos a serem atendidos para a aprovação dos projetos.

Esses requisitos podem ser divididos em quatro grupos:

a) requisitos de formação das equipes de pesquisa, desenvolvimento e implantação das soluções computacionais que se utilizem de Inteligência Artificial (art. 20);

b) requisitos deontológicos relacionados à vedação ao preconceito (art. 21);

c) requisitos de governança (arts. 22 e 23); e

d) requisitos de qualidade técnica (art. 24).

No que se refere à formação das equipes, a resolução preceitua que ela deve ser orientada pela busca da mais ampla diversidade em todas as etapas do processo e pela multidisciplinaridade. A exigência de diversidade pode, todavia, ser dispensada por "decisão fundamentada, dentre outros motivos, pela ausência de profissionais no quadro de pessoal dos tribunais".

As equipes do laboratório deverão zelar pelo cumprimento desse requisito, cabendo-lhes pleitear a dispensa, quando necessário, no momento de apresentação do projeto para aprovação interna, nos termos descritos mais adiante no Capítulo IV, item 1.

Quanto aos requisitos deontológicos, caberá a cada membro da equipe zelar pela lisura dos trabalhos realizados, comunicando qualquer incidente à coordenação do laboratório.

No que tange aos requisitos de governança, deve a equipe de projeto:

1º) Obter prévia e expressa autorização do CNJ antes de iniciar o desenvolvimento de modelos de IA que envolvam reconhecimento facial.

2º) Evitar projetos em matéria penal, sobretudo com relação à sugestão de modelos de decisões preditivas, ressalvados os projetos que visem à automação e ao oferecimento de subsídios destinados ao cálculo de penas, prescrição, verificação de reincidência, mapeamentos, classificações e triagem dos autos para fins de gerenciamento de acervo.

Por fim, a equipe de projeto deverá zelar pelo atendimento de todos os requisitos de qualidade técnica, mediante a utilização preferencial de software de código aberto que permita (i) a adoção dos padrões de interoperabilidade definidos pelo CNJ para o SINAPSES[71]; (ii) a utilização de ambiente de desenvolvimento colaborativo definido pelo LIAA-3R; (iii) a observância das exigências de rastreabilidade, auditabilidade e explicabilidade decorrentes do dever de transparência; e (iv) a

[71] A Portaria CNJ determina que o desenvolvimento de modelos de IA no âmbito do Poder Judiciário deverá ser realizado na plataforma SINAPSES, de modo a assegurar o compartilhamento dos modelos e o trabalho colaborativo. Tal determinação, porém, não impede as equipes de projeto de utilizarem editores, IDEs e outras ferramentas, técnicas ou tecnologias consideradas padrão de indústria, inclusive para treinamento dos modelos, desde que respeitadas as normas de segurança da informação aplicáveis à Justiça Federal da 3ª Região.

cooperação entre outros segmentos e áreas do setor público e a sociedade civil.

A equipe de projeto deverá, ainda, documentar todas as licenças das dependências utilizadas no projeto e definir, juntamente com a SETI, a licença a ser aplicada à solução de IA desenvolvida.

7) Prestação de Contas e Responsabilização

7.1) Resolução CNJ

> "Art. 25. Qualquer solução computacional do Poder Judiciário que utilizar modelos de Inteligência Artificial deverá assegurar total transparência na prestação de contas, com o fim de garantir o impacto positivo para os usuários finais e para a sociedade.
>
> Parágrafo único. A prestação de contas compreenderá:
>
> I – os nomes dos responsáveis pela execução das ações e pela prestação de contas;
>
> II – os custos envolvidos na pesquisa, desenvolvimento, implantação, comunicação e treinamento;
>
> III – a existência de ações de colaboração e cooperação entre os agentes do setor público ou desses com a iniciativa privada ou a sociedade civil;
>
> IV – os resultados pretendidos e os que foram efetivamente alcançados;
>
> V – a demonstração de efetiva publicidade quanto à natureza do serviço oferecido, técnicas utilizadas, desempenho do sistema e riscos de erros.
>
> Art. 26. O desenvolvimento ou a utilização de sistema inteligente em desconformidade aos princípios e regras estabelecidos nesta Resolução será objeto de apuração e, sendo o caso, punição dos responsáveis.
>
> Art. 27. Os órgãos do Poder Judiciário informarão ao Conselho Nacional de Justiça todos os registros de eventos adversos no uso da Inteligência Artificial."

7.2) Aplicação no Âmbito do Laboratório

A obrigação de prestar contas passa a existir a partir do início do projeto.

Para tanto, é necessário, inicialmente, uma definição clara de papéis e das atribuições dos membros da equipe participante dos projetos, considerando as competências técnicas exigidas para o sucesso da iniciativa.

Nesse sentido, a Portaria SINAPSES, no item 5 de seu Anexo, havia definido os atores e respectivos perfis desejados para compor as equipes de projeto: coordenador,

gestor técnico, cientista de dados, cientista de inteligência artificial, engenheiro de inteligência artificial, analista desenvolvedor *full-stack* e curadoria. Apesar de revogada a portaria, essa relação pode ainda servir como elemento de orientação para a formação das equipes.

No âmbito do LIAA-3R, nem sempre as equipes poderão contar com uma pessoa específica para desempenhar cada um dos perfis acima delineados. Uma mesma pessoa poderá exercer um ou mais desses perfis, respeitada a segregação de funções conflitantes, como será abordado adiante no Capítulo IV, item 4, e adotando-se o cuidado necessário para que as responsabilidades sejam atribuídas conforme a capacitação e o perfil técnico de cada membro das equipes de projetos do laboratório.

É desejável, ainda, que as equipes participem de treinamentos internos e externos. Os membros do laboratório poderão realizar cursos para os demais participantes, buscando sempre disseminar o conhecimento técnico relativo aos modelos de IA.

Outro aspecto relativo à prestação de contas e à responsabilização diz respeito ao gerenciamento de riscos do projeto. É recomendável que as equipes documentem os riscos identificados com base nas diretrizes deste documento e, sempre que possível, sugiram às equipes técnicas, responsáveis pela implantação, os meios e ferramentas adequados para monitorá-los e mitigá-los e corrigir eventuais falhas ou resultados indesejados.

Em suma, toda a documentação do projeto deve estar sempre em ordem e atualizada para permitir a prestação de contas a qualquer tempo, assim como a gestão dos riscos envolvidos, nos termos acima.

IV - DIRETRIZES ESPECÍFICAS DE CONFORMIDADE

1) Aprovação e Registro

Nos termos do art. 4º, caput, da Portaria Instituidora, três regras devem ser observadas quando da propositura de um novo projeto:

1ª) Abrir expediente eletrônico específico no SEI, com observância dos procedimentos e boas práticas em vigor (preenchimento e assinatura de FIP, documentação completa do projeto etc).

2ª) Obter aprovação da coordenação do LIAA-3R, que dará ciência a outros órgãos internos, especialmente ADEG, AGES e SETI, assim como ao CNJ, e providenciará, se necessário, a autorização da Presidência, da Comissão de Informática, da CGPDP-3R ou de quaisquer outros órgãos internos ou externos de controle. Durante o processo de aprovação, a equipe deve atentar para o disposto nos §§ 2º, 3º e 4º do art. 3º da Portaria Instituidora[72], verificando se existem outros projetos similares em andamento e se há conflito entre eles e o projeto que se pretende desenvolver. Deve também cuidar da observância da LGPD e das normas internas de segurança da informação. Havendo

[72] "Art. 3º [...]
[...]
§ 2º As atividades do LIAA-3R deverão ser desempenhadas de modo a não interferir com outras iniciativas das áreas técnicas do Tribunal Regional Federal da 3.ª Região.
§ 3º A criação de mais de um modelo de inteligência artificial, por equipes diferentes, para a solução de um mesmo tipo de problema, não significará, por si, a existência de conflito, interferência ou retrabalho.
§ 4º Na hipótese do parágrafo anterior, o coordenador do LIAA-3R e os coordenadores de projeto, em interlocução com os órgãos técnicos do Tribunal Regional Federal da 3ª Região, deverão zelar para que a criação dos modelos em paralelo se dê de modo harmônico, a fim de que facilite e acelere a identificação dos pontos fortes e fracos de cada uma das abordagens utilizadas e enriqueça, desse modo, o repertório das equipes envolvidas, mediante a troca experiências e o mútuo aprendizado."

dúvida sobre a compatibilidade do projeto com esses preceitos normativos, é recomendável que solicite parecer prévio dos órgãos internos competentes.

3º) No curso do desenvolvimento do projeto, surgindo a necessidade ou dúvida sobre a necessidade de dar ciência do projeto a outros órgãos ou de obter novas aprovações, a equipe deverá encaminhar a solicitação à coordenação do LIAA-3R, que adotará as providências que entender pertinentes.

4ª) Providenciar registro no PGP3R.

O processo acima e as providências adicionais a serem tomadas pela equipe após a aprovação e registro do projeto estão visualmente descritos no fluxograma do **Anexo IX**.

2) Documentação

As equipes de projeto deverão manter os códigos-fonte e os datasets nos repositórios designados pela coordenação do LIAA-3R, ouvida a SETI. Deverão, ainda, incluir no expediente SEI do projeto todos os demais artefatos de documentação, assim como registrar no referido expediente: (i) lista completa de dependências, com suas respectivas licenças de uso; (ii) os testes realizados e seus respectivos resultados; (iii) os meios de comunicação utilizados para troca de informações pela equipe e com atores externos; e (iv) eventual participação de atores externos, com menção ao papel que tiveram no projeto e eventual acesso a dados pessoais ou sigilosos.

Ao término de cada um dos ciclos de anotação de datasets, a equipe de anotadores deverá elaborar relatório sucinto, seguindo o modelo descrito no Anexo IV, sem prejuízo da inclusão de quaisquer informações e documentos adicionais que entender necessários para a documentação da atividade.

De modo similar, a equipe de desenvolvedores elaborará relatórios sucintos sobre cada ciclo de desenvolvimento, seguindo o modelo do Anexo V. Reconhecemos que a definição do que seja um “ciclo de desenvolvimento” é problemática, o que deixa à própria equipe de desenvolvimento a tarefa de definir os marcos (checkpoints) adequados para a elaboração dos relatórios.

Além desses documentos, as equipes de anotadores e desenvolvedores devem atentar, ainda, para os documentos exigidos em caso de tratamento de dados pessoais, mencionados no Capítulo V (“Diretrizes Referentes à LGPD”).

Ao final do projeto, as equipes de anotadores e desenvolvedores elaborarão,

em conjunto, um relatório sobre a formação dos datasets, segundo o modelo descrito no Anexo VI, e a equipe de desenvolvedores responderá ao questionário do Anexo VII, fazendo remissão ao código-fonte, com todas as suas dependências, e aos datasets utilizados na criação dos modelos de IA.

3) Segurança da Informação

Como condição para participarem em projetos de IA conduzidos no âmbito do LIAA-3R, cada um dos integrantes das equipes de projeto, inclusive os anotadores e atores externos, deverá firmar termo de ciência e confidencialidade e conflito de interesses, a ser juntado ao expediente do projeto, conforme o modelo do **Anexo I**.

Eventuais incidentes envolvendo segurança da informação devem ser prontamente comunicados à Comissão Local de Resposta a Incidentes e à Comissão Local de Segurança da Informação.

4) Conflito de Interesses

Os membros das equipes de projeto devem zelar para que não ocorram conflitos de interesse e declarar expressamente os potenciais conflitos que identificarem, submetendo a informação à coordenação do LIAA-3R.

Para tanto, cada um dos integrantes das equipes de projeto, inclusive os anotadores e atores externos, deverão firmar o termo de ciência e confidencialidade e conflito de interesses, conforme o modelo do Anexo II, de modo a possibilitar análise pela coordenação do LIAA-3R e pelos órgãos de controle, se necessário.

No que diz respeito à validação ética e jurídica dos modelos de IA, embora eventual participação de membros do GVEJ nas atividades de desenvolvimento ou anotação não caracterize, necessariamente, conflito de interesses, recomenda-se que os participantes das atividades de desenvolvimento e anotação não participem do processo de validação e vice-versa, a fim de evitar o enviesamento e a parcialidade do processo de validação. De qualquer modo, tal não impede que membros da equipe de desenvolvedores sejam convidados pelo GVEJ para prestar esclarecimentos, participar das reuniões e colaborar na produção de documentos, nem que membros do GVEJ auxiliem as equipes de anotadores e desenvolvedores sobre assuntos relacionados à documentação, à conformidade e à auditabilidade de suas atividades.

V - DIRETRIZES REFERENTES À LGPD

1) Definições

"Art. 5º Para os fins desta Lei, considera-se:

I - dado pessoal: informação relacionada a pessoa natural identificada ou identificável;

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

III - dado anonimizado: dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento;

IV - banco de dados: conjunto estruturado de dados pessoais, estabelecido em um ou em vários locais, em suporte eletrônico ou físico;

V - titular: pessoa natural a quem se referem os dados pessoais que são objeto de tratamento;

VI - controlador: pessoa natural ou jurídica, de direito público ou privado, a quem competem as decisões referentes ao tratamento de dados pessoais;

VII - operador: pessoa natural ou jurídica, de direito público ou privado, que realiza o tratamento de dados pessoais em nome do controlador;

VIII - encarregado: pessoa indicada pelo controlador e operador para atuar como canal de comunicação entre o controlador, os titulares dos dados e a Autoridade Nacional de Proteção de Dados (ANPD); (Redação dada pela Lei nº 13.853, de 2019)

IX - agentes de tratamento: o controlador e o operador;

X - tratamento: toda operação realizada com dados pessoais, como as que se referem a coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração;

XI - anonimização: utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo;

> XII - consentimento: manifestação livre, informada e inequívoca pela qual o titular concorda com o tratamento de seus dados pessoais para uma finalidade determinada;
>
> XIII - bloqueio: suspensão temporária de qualquer operação de tratamento, mediante guarda do dado pessoal ou do banco de dados;
>
> XIV - eliminação: exclusão de dado ou de conjunto de dados armazenados em banco de dados, independentemente do procedimento empregado;
>
> XV - transferência internacional de dados: transferência de dados pessoais para país estrangeiro ou organismo internacional do qual o país seja membro;
>
> XVI - uso compartilhado de dados: comunicação, difusão, transferência internacional, interconexão de dados pessoais ou tratamento compartilhado de bancos de dados pessoais por órgãos e entidades públicos no cumprimento de suas competências legais, ou entre esses e entes privados, reciprocamente, com autorização específica, para uma ou mais modalidades de tratamento permitidas por esses entes públicos, ou entre entes privados;
>
> XVII - relatório de impacto à proteção de dados pessoais: documentação do controlador que contém a descrição dos processos de tratamento de dados pessoais que podem gerar riscos às liberdades civis e aos direitos fundamentais, bem como medidas, salvaguardas e mecanismos de mitigação de risco;
>
> XVIII - órgão de pesquisa: órgão ou entidade da administração pública direta ou indireta ou pessoa jurídica de direito privado sem fins lucrativos legalmente constituída sob as leis brasileiras, com sede e foro no País, que inclua em sua missão institucional ou em seu objetivo social ou estatutário a pesquisa básica ou aplicada de caráter histórico, científico, tecnológico ou estatístico; e (Redação dada pela Lei nº 13.853, de 2019)
>
> XIX - autoridade nacional: órgão da administração pública responsável por zelar, implementar e fiscalizar o cumprimento desta Lei em todo o território nacional. (Redação dada pela Lei nº 13.853, de 2019)"

Comentários

Uma vez que as equipes do LIAA-3R atuarão principalmente na fase de desenvolvimento dos modelos de IA, raramente haverá coleta de novos dados pessoais. Na maior parte do tempo, serão utilizados dados já armazenados nos bancos de dados institucionais.

De qualquer forma, uma vez que a definição de "tratamento" é bastante ampla, o uso de dados pessoais já coletados e armazenados no âmbito da Justiça Federal da 3ª Região, estruturados ou não, também está sujeita às normas da LGPD.

Por conseguinte, sempre que tiver acesso e manipular dados pessoais, a equipe de projeto será considerada "operador" e deverá observar as obrigações atribuídas pela LGPD a essa modalidade de agente de tratamento, assim como as determinações do CGPDP-3R, que é o órgão controlador na Justiça Federal da 3ª Região[73].

[73] "Art. 39. O operador deverá realizar o tratamento segundo as instruções fornecidas pelo controlador, que verificará a observância das próprias instruções e das normas sobre a matéria."

2) Princípios

"Art. 2º A disciplina da proteção de dados pessoais tem como fundamentos:

I - o respeito à privacidade;

II - a autodeterminação informativa;

III - a liberdade de expressão, de informação, de comunicação e de opinião;

IV - a inviolabilidade da intimidade, da honra e da imagem;

V - o desenvolvimento econômico e tecnológico e a inovação;

VI - a livre iniciativa, a livre concorrência e a defesa do consumidor; e

VII - os direitos humanos, o livre desenvolvimento da personalidade, a dignidade e o exercício da cidadania pelas pessoas naturais.

[...]

Art. 6º As atividades de tratamento de dados pessoais deverão observar a boa-fé e os seguintes princípios:

I - finalidade: realização do tratamento para propósitos legítimos, específicos, explícitos e informados ao titular, sem possibilidade de tratamento posterior de forma incompatível com essas finalidades;

II - adequação: compatibilidade do tratamento com as finalidades informadas ao titular, de acordo com o contexto do tratamento;

III - necessidade: limitação do tratamento ao mínimo necessário para a realização de suas finalidades, com abrangência dos dados pertinentes, proporcionais e não excessivos em relação às finalidades do tratamento de dados;

IV - livre acesso: garantia, aos titulares, de consulta facilitada e gratuita sobre a forma e a duração do tratamento, bem como sobre a integralidade de seus dados pessoais;

V - qualidade dos dados: garantia, aos titulares, de exatidão, clareza, relevância e atualização dos dados, de acordo com a necessidade e para o cumprimento da finalidade de seu tratamento;

VI - transparência: garantia, aos titulares, de informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de tratamento, observados os segredos comercial e industrial;

VII - segurança: utilização de medidas técnicas e administrativas aptas a proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou difusão;

VIII - prevenção: adoção de medidas para prevenir a ocorrência de danos em virtude do tratamento de dados pessoais;

IX - não discriminação: impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos;

X - responsabilização e prestação de contas: demonstração, pelo agente, da adoção de medidas eficazes e capazes de comprovar a observância e o cumprimento das normas de proteção de dados pessoais e, inclusive, da eficácia dessas medidas."

Comentários

Os princípios da LGPD alinham-se em grande medida aos princípios da Resolução CNJ, da Carta CEPEJ e das Orientações GPAN.

Os princípios da finalidade, adequação e necessidade requerem que as equipes de desenvolvimento utilizem dados pessoais de forma bastante parcimoniosa, tão somente na medida e pelo tempo necessários para o desenvolvimento dos modelos de IA.

Todavia, essa exigência deve ser compreendida segundo a razoabilidade, levando em consideração a natureza experimental das atividades desempenhadas no laboratório de inovação. Assim, é admissível que os dados pessoais inicialmente considerados necessários para a criação de modelos de IA sejam considerados depois desnecessários numa fase posterior do projeto e vice-versa.

Em qualquer caso, é fundamental justificar por escrito o tratamento ou a cessação do tratamento de dados pessoais. A equipe deve sempre elaborar fundamentação por escrito, de preferência antes de iniciar as atividades de tratamento, juntando-a ao expediente do projeto, conforme o modelo do **Anexo II**. Deve, ainda, registrar no expediente o encerramento das operações de tratamento e justificar a conservação dos *datasets*, conforme o modelo do **Anexo III**. Veja também os itens 4, 5 e 6 abaixo.

A prestação de informações claras e atualizadas pelas equipes de projeto quanto ao tratamento de dados pessoais contribui, também, para o atendimento do princípio do livre acesso.

O princípio da qualidade dos dados envolve principalmente o processo de tratamento que se dá fora do âmbito do laboratório, pelos órgãos judiciais e administrativos da Justiça Federal. Todavia, é necessário que as equipes de desenvolvimento atentem também a esse princípio, de modo a conservar e até mesmo melhorar, quando possível, a qualidade dos dados por elas utilizados. Sobre esse tópico, veja o item IV-4.

Os princípios da transparência, da segurança e da prevenção são abordados nos itens 7 e 8 abaixo.

Quanto à não discriminação, à responsabilização e à prestação de contas, aplica-se o que já foi dito nos itens IV-2 e IV-7.

3) Abrangência

"Art. 3º Esta Lei aplica-se a qualquer operação de tratamento realizada por pessoa natural ou por pessoa jurídica de direito público ou privado, independentemente do meio, do país de sua sede ou do país onde estejam localizados os dados, desde que:

I - a operação de tratamento seja realizada no território nacional;

II - a atividade de tratamento tenha por objetivo a oferta ou o fornecimento de bens ou serviços ou o tratamento de dados de indivíduos localizados no território nacional; ou (Redação dada pela Lei nº 13.853, de 2019)

III - os dados pessoais objeto do tratamento tenham sido coletados no território nacional.

§ 1º Consideram-se coletados no território nacional os dados pessoais cujo titular nele se encontre no momento da coleta.

§ 2º Excetua-se do disposto no inciso I deste artigo o tratamento de dados previsto no inciso IV do caput do art. 4º desta Lei.

Art. 4º Esta Lei não se aplica ao tratamento de dados pessoais:

I - realizado por pessoa natural para fins exclusivamente particulares e não econômicos;

II - realizado para fins exclusivamente:

a) jornalístico e artísticos; ou

b) acadêmicos, aplicando-se a esta hipótese os arts. 7º e 11 desta Lei;

III - realizado para fins exclusivos de:

a) segurança pública;

b) defesa nacional;

c) segurança do Estado; ou

d) atividades de investigação e repressão de infrações penais; ou

IV - provenientes de fora do território nacional e que não sejam objeto de comunicação, uso compartilhado de dados com agentes de tratamento brasileiros ou objeto de transferência internacional de dados com outro país que não o de proveniência, desde que o país de proveniência proporcione grau de proteção de dados pessoais adequado ao previsto nesta Lei.

§ 1º O tratamento de dados pessoais previsto no inciso III será regido por legislação específica, que deverá prever medidas proporcionais e estritamente necessárias ao atendimento do interesse público, observados o devido processo legal, os princípios gerais de proteção e os direitos do titular previstos nesta Lei.

[...]

Art. 12. Os dados anonimizados não serão considerados dados pessoais para os fins desta Lei, salvo quando o processo de anonimização ao qual foram submetidos for revertido, utilizando exclusivamente meios próprios, ou quando, com esforços razoáveis, puder ser revertido.

§ 1º A determinação do que seja razoável deve levar em consideração fatores objetivos, tais como custo e tempo necessários para reverter o processo de anonimização, de acordo com as

tecnologias disponíveis, e a utilização exclusiva de meios próprios.

§ 2º Poderão ser igualmente considerados como dados pessoais, para os fins desta Lei, aqueles utilizados para formação do perfil comportamental de determinada pessoa natural, se identificada.

§ 3º A autoridade nacional poderá dispor sobre padrões e técnicas utilizados em processos de anonimização e realizar verificações acerca de sua segurança, ouvido o Conselho Nacional de Proteção de Dados Pessoais.

[...]

Art. 23. O tratamento de dados pessoais pelas pessoas jurídicas de direito público referidas no parágrafo único do art. 1º da Lei nº 12.527, de 18 de novembro de 2011 (Lei de Acesso à Informação), deverá ser realizado para o atendimento de sua finalidade pública, na persecução do interesse público, com o objetivo de executar as competências legais ou cumprir as atribuições legais do serviço público, desde que:

I - sejam informadas as hipóteses em que, no exercício de suas competências, realizam o tratamento de dados pessoais, fornecendo informações claras e atualizadas sobre a previsão legal, a finalidade, os procedimentos e as práticas utilizadas para a execução dessas atividades, em veículos de fácil acesso, preferencialmente em seus sítios eletrônicos;

[...]"

Comentários

Considerando a amplitude do conceito de "tratamento" dado pelo art. 5º da LGPD, qualquer operação realizada com dados pessoais, sensíveis ou não, salvo quando integralmente anonimizados (art. 12), submete as equipes de desenvolvimento às obrigações da lei na condição de "operadoras".

A fim de viabilizar e facilitar o cumprimento do disposto no inciso I do art. 23 pelos órgãos de administração da Justiça Federal da 3ª Região, as equipes de desenvolvimento devem providenciar a juntada no expediente do projeto de termo de justificativa de uso de dados pessoais, conforme o modelo do **Anexo VI**, mantendo sempre atualizadas as informações ali prestadas, mediante juntada de novos termos de justificativa sempre que necessário.

4) Tratamento de Dados Pessoais

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

I - mediante o fornecimento de consentimento pelo titular;

II - para o cumprimento de obrigação legal ou regulatória pelo controlador;

III - pela administração pública, para o tratamento e uso compartilhado de dados necessários à

execução de políticas públicas previstas em leis e regulamentos ou respaldadas em contratos, convênios ou instrumentos congêneres, observadas as disposições do Capítulo IV desta Lei;

IV - para a realização de estudos por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais;

V - quando necessário para a execução de contrato ou de procedimentos preliminares relacionados a contrato do qual seja parte o titular, a pedido do titular dos dados;

VI - para o exercício regular de direitos em processo judicial, administrativo ou arbitral, esse último nos termos da Lei nº 9.307, de 23 de setembro de 1996 (Lei de Arbitragem) ;

VII - para a proteção da vida ou da incolumidade física do titular ou de terceiro;

VIII - para a tutela da saúde, exclusivamente, em procedimento realizado por profissionais de saúde, serviços de saúde ou autoridade sanitária; (Redação dada pela Lei nº 13.853, de 2019)

IX - quando necessário para atender aos interesses legítimos do controlador ou de terceiro, exceto no caso de prevalecerem direitos e liberdades fundamentais do titular que exijam a proteção dos dados pessoais; ou

X - para a proteção do crédito, inclusive quanto ao disposto na legislação pertinente.

§ 1º (Revogado)

§ 2º (Revogado).

§ 3º O tratamento de dados pessoais cujo acesso é público deve considerar a finalidade, a boa-fé e o interesse público que justificaram sua disponibilização.

§ 4º É dispensada a exigência do consentimento previsto no caput deste artigo para os dados tornados manifestamente públicos pelo titular, resguardados os direitos do titular e os princípios previstos nesta Lei.

§ 5º O controlador que obteve o consentimento referido no inciso I do caput deste artigo que necessitar comunicar ou compartilhar dados pessoais com outros controladores deverá obter consentimento específico do titular para esse fim, ressalvadas as hipóteses de dispensa do consentimento previstas nesta Lei.

§ 6º A eventual dispensa da exigência do consentimento não desobriga os agentes de tratamento das demais obrigações previstas nesta Lei, especialmente da observância dos princípios gerais e da garantia dos direitos do titular.

§ 7º O tratamento posterior dos dados pessoais a que se referem os §§ 3º e 4º deste artigo poderá ser realizado para novas finalidades, desde que observados os propósitos legítimos e específicos para o novo tratamento e a preservação dos direitos do titular, assim como os fundamentos e os princípios previstos nesta Lei. (Incluído pela Lei nº 13.853, de 2019)

[...]

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

I - quando o titular ou seu responsável legal consentir, de forma específica e destacada, para finalidades específicas;

II - sem fornecimento de consentimento do titular, nas hipóteses em que for indispensável para:

a) cumprimento de obrigação legal ou regulatória pelo controlador;

b) tratamento compartilhado de dados necessários à execução, pela administração pública, de políticas públicas previstas em leis ou regulamentos;

c) realização de estudos por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais sensíveis;

d) exercício regular de direitos, inclusive em contrato e em processo judicial, administrativo e arbitral, este último nos termos da Lei nº 9.307, de 23 de setembro de 1996 (Lei de Arbitragem);

e) proteção da vida ou da incolumidade física do titular ou de terceiro;

f) tutela da saúde, exclusivamente, em procedimento realizado por profissionais de saúde, serviços de saúde ou autoridade sanitária; ou (Redação dada pela Lei nº 13.853, de 2019)

g) garantia da prevenção à fraude e à segurança do titular, nos processos de identificação e autenticação de cadastro em sistemas eletrônicos, resguardados os direitos mencionados no art. 9º desta Lei e exceto no caso de prevalecerem direitos e liberdades fundamentais do titular que exijam a proteção dos dados pessoais.

§ 1º Aplica-se o disposto neste artigo a qualquer tratamento de dados pessoais que revele dados pessoais sensíveis e que possa causar dano ao titular, ressalvado o disposto em legislação específica.

§ 2º Nos casos de aplicação do disposto nas alíneas “a” e “b” do inciso II do caput deste artigo pelos órgãos e pelas entidades públicas, será dada publicidade à referida dispensa de consentimento, nos termos do inciso I do caput do art. 23 desta Lei.

[...]

Art. 14. O tratamento de dados pessoais de crianças e de adolescentes deverá ser realizado em seu melhor interesse, nos termos deste artigo e da legislação pertinente.

§ 1º O tratamento de dados pessoais de crianças deverá ser realizado com o consentimento específico e em destaque dado por pelo menos um dos pais ou pelo responsável legal.

§ 2º No tratamento de dados de que trata o § 1º deste artigo, os controladores deverão manter pública a informação sobre os tipos de dados coletados, a forma de sua utilização e os procedimentos para o exercício dos direitos a que se refere o art. 18 desta Lei.

§ 3º Poderão ser coletados dados pessoais de crianças sem o consentimento a que se refere o § 1º deste artigo quando a coleta for necessária para contatar os pais ou o responsável legal, utilizados uma única vez e sem armazenamento, ou para sua proteção, e em nenhum caso poderão ser repassados a terceiro sem o consentimento de que trata o § 1º deste artigo.

§ 4º Os controladores não deverão condicionar a participação dos titulares de que trata o § 1º deste artigo em jogos, aplicações de internet ou outras atividades ao fornecimento de informações pessoais além das estritamente necessárias à atividade.

§ 5º O controlador deve realizar todos os esforços razoáveis para verificar que o consentimento a que se refere o § 1º deste artigo foi dado pelo responsável pela criança, consideradas as tecnologias disponíveis.

§ 6º As informações sobre o tratamento de dados referidas neste artigo deverão ser fornecidas de maneira simples, clara e acessível, consideradas as características físico-motoras, perceptivas, sensoriais, intelectuais e mentais do usuário, com uso de recursos audiovisuais quando adequado, de forma a proporcionar a informação necessária aos pais ou ao responsável legal e adequada ao entendimento da criança.”

Comentários

Antes de iniciarem operações de tratamento de dados pessoais, as equipes de projeto devem certificar-se de que os requisitos dos arts. 7º, 11 e 14 da LGPD estão devidamente atendidos.

Em geral, uma vez que os modelos de IA desenvolvidos no âmbito do LIAA-3R destinam-se à melhoria dos serviços judiciais ou da administração judiciária, o tratamento de dados pessoais, inclusive os sensíveis, justifica-se nos termos dos arts. 7º, incisos II e/ou III, e 11, inciso II, alíneas "a", "b" ou "g". Necessário, contudo, que as equipes de desenvolvimento indiquem com clareza, por escrito, na documentação do projeto, o preceito legal que as autoriza a realizar as operações de tratamento de dados pessoais pretendidas (cf. item 3 acima), bem como declarar que o projeto não implica outra restrição regulada de tratamento de dados. Devem informar, em especial, de modo fundamentado, eventual dispensa de consentimento, a fim de subsidiar a prestação de informações pelos órgãos administrativos da Justiça Federal da 3ª Região, nos termos do § 2º do art. 11, combinado com o art. 32, inciso I, da LGPD.

As operações de tratamento de dados pessoais sensíveis e de dados pessoais de crianças e adolescentes somente devem ser realizadas após obtenção de autorização específica do CGPDP-3R, conforme já mencionado no Capítulo V, itens 1 e 2.

5) Transferência Internacional de Dados Pessoais

> "Art. 33. A transferência internacional de dados pessoais somente é permitida nos seguintes casos:
>
> I - para países ou organismos internacionais que proporcionem grau de proteção de dados pessoais adequado ao previsto nesta Lei;
>
> II - quando o controlador oferecer e comprovar garantias de cumprimento dos princípios, dos direitos do titular e do regime de proteção de dados previstos nesta Lei, na forma de:
>
> a) cláusulas contratuais específicas para determinada transferência;
>
> b) cláusulas-padrão contratuais;
>
> c) normas corporativas globais;
>
> d) selos, certificados e códigos de conduta regularmente emitidos;
>
> III - quando a transferência for necessária para a cooperação jurídica internacional entre órgãos públicos de inteligência, de investigação e de persecução, de acordo com os instrumentos de direito internacional;
>
> IV - quando a transferência for necessária para a proteção da vida ou da incolumidade física do

titular ou de terceiro;

V - quando a autoridade nacional autorizar a transferência;

VI - quando a transferência resultar em compromisso assumido em acordo de cooperação internacional;

VII - quando a transferência for necessária para a execução de política pública ou atribuição legal do serviço público, sendo dada publicidade nos termos do inciso I do caput do art. 23 desta Lei;

VIII - quando o titular tiver fornecido o seu consentimento específico e em destaque para a transferência, com informação prévia sobre o caráter internacional da operação, distinguindo claramente esta de outras finalidades; ou

IX - quando necessário para atender às hipóteses previstas nos incisos II, V e VI do art. 7º desta Lei.

Parágrafo único. Para os fins do inciso I deste artigo, as pessoas jurídicas de direito público referidas no parágrafo único do art. 1º da Lei nº 12.527, de 18 de novembro de 2011 (Lei de Acesso à Informação), no âmbito de suas competências legais, e responsáveis, no âmbito de suas atividades, poderão requerer à autoridade nacional a avaliação do nível de proteção a dados pessoais conferido por país ou organismo internacional."

Comentários

As equipes de projeto devem manter os dados e *datasets* armazenados nos meios que lhes forem disponibilizados pela SETI, abstendo-se de transferir os dados e *datasets* para qualquer outro meio físico ou virtual, incluindo repositórios privados ou dispositivos móveis, próprios ou institucionais, sem prévia autorização por escrito da SETI ou do CGPDP-3R.

6) Término do Tratamento de Dados

"Art. 15. O término do tratamento de dados pessoais ocorrerá nas seguintes hipóteses:

I - verificação de que a finalidade foi alcançada ou de que os dados deixaram de ser necessários ou pertinentes ao alcance da finalidade específica almejada;

II - fim do período de tratamento;

III - comunicação do titular, inclusive no exercício de seu direito de revogação do consentimento conforme disposto no § 5º do art. 8º desta Lei, resguardado o interesse público; ou

IV - determinação da autoridade nacional, quando houver violação ao disposto nesta Lei.

Art. 16. Os dados pessoais serão eliminados após o término de seu tratamento, no âmbito e nos limites técnicos das atividades, autorizada a conservação para as seguintes finalidades:

I - cumprimento de obrigação legal ou regulatória pelo controlador;

> II - estudo por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais;
>
> III - transferência a terceiro, desde que respeitados os requisitos de tratamento de dados dispostos nesta Lei; ou
>
> IV - uso exclusivo do controlador, vedado seu acesso por terceiro, e desde que anonimizados os dados."

Comentários

Seguindo os princípios da finalidade, adequação e necessidade (cf. item 2), as equipes de projeto devem limitar o tratamento de dados pessoais ao necessário para o desenvolvimento dos modelos de IA, cessando o tratamento assim que esgotada a sua finalidade. Todavia, os *datasets* efetivamente utilizados para treinamento, validação e testes dos modelos finais deverão ser integralmente conservados em repositório previamente apontado pela Coordenação do LIAA-3R, de modo a manter a auditabilidade, a rastreabilidade e a explicabilidade da solução de IA desenvolvida.

As equipes de projeto indicarão na documentação o local de armazenamento dos *datasets*, com descrição de suas características e conteúdo, assim como a justificativa legal para a sua conservação, conforme modelo do **Anexo III**. Como forma de garantir a integridade dos datasets e a segurança do projeto, as equipes poderão utilizar técnicas de assinatura digital, criptografia ou geração de *hash*.

7) Transparência

> "Art. 20. O titular dos dados tem direito a solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, incluídas as decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade. (Redação dada pela Lei nº 13.853, de 2019)
>
> § 1º O controlador deverá fornecer, sempre que solicitadas, informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada, observados os segredos comercial e industrial.
>
> § 2º Em caso de não oferecimento de informações de que trata o § 1º deste artigo baseado na observância de segredo comercial e industrial, a autoridade nacional poderá realizar auditoria para verificação de aspectos discriminatórios em tratamento automatizado de dados pessoais.
>
> § 3º (VETADO). (Incluído pela Lei nº 13.853, de 2019)
>
> [...]
>
> Art. 37. O controlador e o operador devem manter registro das operações de tratamento de dados pessoais que realizarem, especialmente quando baseado no legítimo interesse."

Comentários

O comando do art. 20 destina-se precipuamente ao controlador. Todavia, o laboratório tem papel importante no cumprimento dessa obrigação legal ao assegurar a transparência dos modelos de IA ali desenvolvidos. Por conseguinte, o disposto no art. 20 é motivo adicional para que as equipes de projeto zelem pela auditabilidade, pela rastreabilidade e pela explicabilidade dos modelos de IA.

Nos termos do art. 37, as equipes de projeto, na condição de operadores, devem manter, na documentação, registro de todas as operações de tratamento de dados pessoais que realizarem.

8) Segurança e Prevenção

> "Art. 46. Os agentes de tratamento devem adotar medidas de segurança, técnicas e administrativas aptas a proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou qualquer forma de tratamento inadequado ou ilícito.
>
> § 1º A autoridade nacional poderá dispor sobre padrões técnicos mínimos para tornar aplicável o disposto no caput deste artigo, considerados a natureza das informações tratadas, as características específicas do tratamento e o estado atual da tecnologia, especialmente no caso de dados pessoais sensíveis, assim como os princípios previstos no caput do art. 6º desta Lei.
>
> § 2º As medidas de que trata o caput deste artigo deverão ser observadas desde a fase de concepção do produto ou do serviço até a sua execução.
>
> Art. 47. Os agentes de tratamento ou qualquer outra pessoa que intervenha em uma das fases do tratamento obriga-se a garantir a segurança da informação prevista nesta Lei em relação aos dados pessoais, mesmo após o seu término.
>
> [...]
>
> Art. 49. Os sistemas utilizados para o tratamento de dados pessoais devem ser estruturados de forma a atender aos requisitos de segurança, aos padrões de boas práticas e de governança e aos princípios gerais previstos nesta Lei e às demais normas regulamentares.
>
> Art. 50. Os controladores e operadores, no âmbito de suas competências, pelo tratamento de dados pessoais, individualmente ou por meio de associações, poderão formular regras de boas práticas e de governança que estabeleçam as condições de organização, o regime de funcionamento, os procedimentos, incluindo reclamações e petições de titulares, as normas de segurança, os padrões técnicos, as obrigações específicas para os diversos envolvidos no tratamento, as ações educativas, os mecanismos internos de supervisão e de mitigação de riscos e outros aspectos relacionados ao tratamento de dados pessoais.
>
> § 1º Ao estabelecer regras de boas práticas, o controlador e o operador levarão em consideração, em relação ao tratamento e aos dados, a natureza, o escopo, a finalidade e a

> probabilidade e a gravidade dos riscos e dos benefícios decorrentes de tratamento de dados do titular.
>
> [...]"

Comentários

Cada um dos membros das equipes de projeto deve procurar conhecer, por si, as regras de tratamento de dados e os padrões de boas práticas e governança editados pelo LIAA-3R, assim como pelos órgãos administrativos da Justiça Federal da 3ª Região e pelos órgãos de controle internos e externos, e observá-los fielmente. Não devem, portanto, limitar-se ao que está escrito no presente documento.

Devem também seguir as orientações que lhes forem dadas pelo CGPDP-3R, considerado "controlador" para os fins da LGPD, conforme deixa claro o art. 39 da lei:

> Art. 39. O operador deverá realizar o tratamento segundo as instruções fornecidas pelo controlador, que verificará a observância das próprias instruções e das normas sobre a matéria.

Por fim, cabe também aos membros da equipe de projeto comunicar imediatamente aos órgãos internos competentes quaisquer "situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou qualquer forma de tratamento inadequado ou ilícito" de dados pessoais.

VI - REFERÊNCIAS

ALPAYDIN, Ethem. ***Machine Learning***. Cambridge, MA: MIT Press, 2016.

BARREDO ARRIETA, Alejandro; DÍAZ-RODRÍGUEZ, Natalia; DEL SER, Javier; et al. ***Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI***. Information Fusion, v. 58, p. 82–115, 2020. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1566253519308103>. Acesso em 19 out. 2020.

COELHO, Alexandre Zavaglia. Tecnologia e Design da Justiça Brasileira: o pioneirismo do iJuspLab. In: Inovação no Judiciário: Conceito, Criação e Práticas do Primeiro Laboratório de Inovação do Poder Judiciário. São Paulo: Blucher, 2019, p. 211-222.

CONSELHO EUROPEU. Comissão Europeia para a Eficiência da Justiça (European Commission for the Efficiency of Justice - CEPEJ). ***European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment***. Estrasburgo, 3-4 dez. 2019. Disponível em: <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>. Acesso em: 13 set. 2020.

DOURADO, Gabriela. Design Thinking: Por que utilizar? In: Inovação no Judiciário: Conceito, Criação e Práticas do Primeiro Laboratório de Inovação do Poder Judiciário. São Paulo: Blucher, 2019, p. 79-94.

ESTADOS UNIDOS DA AMÉRICA. U.S. General Services Administration. ***Usability.gov. Improving the User Experience***. Site do governo norte-americano que reúne informações sobre metodologias, diretrizes e modelos para melhorar a experiência do usuário. Disponível em <https://www.usability.gov/>. Acesso em: 13 set. 2020.

FACELI, Katti; LORENA, Ana Carolina; GAMA, João; ALMEIDA, Tiago Agostinho de; CARVALHO, André C. P. L. F. de. Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina. Rio de Janeiro: LTC, 2021, 2ª edição.

GREGÓRIO, Álvaro. Um Laboratório de Inovação no Judiciário: Por quê e o quê. In: Inovação no Judiciário: Conceito, Criação e Práticas do Primeiro Laboratório de Inovação do Poder Judiciário. São Paulo: Blucher, 2019, p. 59-78.

LIMA, Caio Moysés de. Introduzindo a Cultura de Inovação Tecnológica no Poder Judiciário: A Experiência do iJuspLab. In: Inovação no Judiciário: Conceito, Criação e Práticas do Primeiro Laboratório de Inovação do Poder Judiciário. São Paulo: Blucher, 2019, p. 127-158.

PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. **Inteligência Artificial e Direito. Volume 1**. Curitiba: Alteridade, 2019.

RIES, Eric. ***What is the minimum viable product?*** (entrevista para Venture Hacks). 23 mar. 2009. Disponível em: <https://venturehacks.com/minimum-viable-product>. Acesso em: 13 set. 2020.

_____. **A Startup Enxuta: Como os Empreendedores Atuais Utilizam a Inovação Contínua para Criar Empresas Extremamente bem-sucedidas**. São Paulo: Lua de Papel, 2012, e-book Kindle.

RUSSEL, Stuart; NORVIG, Peter. Inteligência Artificial. Rio de Janeiro: Elsevier, 2013, 3ª edição.

THE ALAN TURING INSTITUTE. ***Explaining decisions made with AI. Draft guidance for consultation. Part 1. The basics of explaining AI.*** Versão 1.0. Disponível em: <https://ico.org.uk/media/2616434/explaining-ai-decisions-part-1.pdf>. Acesso em: 14 nov. 2020.

TRIBUNAL REGIONAL FEDERAL DA 3ª REGIÃO. Projeto Sigma, do TRF3, Ganha Prêmio Innovare 2021. Disponível em: <http://web.trf3.jus.br/noticias-intranet/Noticiar/ExibirNoticia/412508-projeto-sigma-do-trf3-ganha-premio-innovare-2021>. Acesso em: 6 jan. 2022.

UNIÃO EUROPEIA. Comissão Europeia. Grupo de Peritos de Alto Nível sobre a Inteligência Artificial - GPAN IA. **Orientações Éticas para uma IA de Confiança**. Publicado em 8 abr. 2019. Versão em língua portuguesa. Disponível em: <https://op.europa.eu/s/oizr>. Acesso em: 12 set. 2020.

ZANONI, Luciana Ortiz Tavares Costa. A Mudança Cultural da Gestão Judicial: Inovação como Base da Busca da Excelência do Serviço Público. In: Inovação no Judiciário: Conceito, Criação e Práticas do Primeiro Laboratório de Inovação do Poder Judiciário. São Paulo: Blucher, 2019, p. 41-58.

ANEXO I - TERMO DE CIÊNCIA E CONFIDENCIALIDADE

TERMO DE CIÊNCIA E CONFIDENCIALIDADE E DE CONFLITO DE INTERESSES

Projeto	<nome do projeto>
Expediente SEI	<número do expediente>
Declarante	<nome, RF, cargo, órgão e função na equipe>
Data de Ingresso	<data de ingresso na equipe>

1) Ciência

Declaro ter lido e compreendido o teor do documento intitulado "Diretrizes de Auditabilidade e Conformidade no Desenvolvimento e Testes de Modelos de IA no Âmbito do LIAA-3R", o qual me comprometo a seguir.

Declaro ter plena ciência do teor das normas de tratamento de dados, de desenvolvimento de modelos de inteligência artificial e de segurança da informação a seguir relacionadas, as quais prometo cumprir fielmente, solicitando auxílio à coordenação do LIAA-3R em caso de dúvida, antes de praticar qualquer ato que possa resultar em violação das referidas normas:

- Lei Geral de Proteção de Dados (Lei nº 13.709/2018)
- Resolução CNJ nº 332, de 21 de agosto de 2020
- Portaria CNJ nº 271, de 4 de dezembro de 2020
- Resolução CJF nº 6, de 7 de abril de 2008
- Resolução CNJ nº 360, de 17 de dezembro de 2020
- Resolução CNJ nº 361, de 17 de dezembro de 2020
- Resolução CNJ nº 362, de 17 de dezembro de 2020
- Portaria CNJ nº 292, de 17 de dezembro de 2020
- Resolução CNJ nº 363, de 12 de janeiro de 2021

2) Confidencialidade

Prometo manter sigilo quanto ao teor dos dados a que tiver acesso no desenvolvimento de modelos de inteligência artificial no âmbito do LIAA-3R, não podendo realizar qualquer operação de tratamento de dados pessoais, incluindo a coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração, que não esteja no escopo do projeto a que se refere o presente expediente.

3) Conflito de Interesses

[Declaro que não exerço atividade, nem ocupo cargo ou função em órgão ou entidade pública ou privada, que resulte em conflito de interesses com a realização do presente projeto.]

ou

[Declarar projetos, incentivos, trabalhos, atividades ou quaisquer outros eventos, ações ou situações que possam ensejar conflito de interesses com o projeto.]

<ASSINATURA ELETRÔNICA>

ANEXO II - TERMO DE JUSTIFICATIVA DE USO DE DADOS PESSOAIS

TERMO DE JUSTIFICATIVA DE USO DE DADOS PESSOAIS

Projeto	<nome do projeto>
Expediente SEI	<número do expediente>
Declarante	<nome, RF, cargo, órgão>, Coordenador(a) da Equipe de Desenvolvimento do Projeto

Na condição de coordenador(a) do projeto a que se refere o presente expediente, declaro que serão realizadas operações de **<especificar operações de tratamento>** dos dados armazenados nas colunas **<especificar>** das tabelas **<especificar>** do banco de dados **<especificar>**.

Referidas operações se destinam a **<explicar>** e serão feitas com base nos arts. **<especificar fundamentos normativos, incluindo, no caso de dados sigilosos ou de dados pessoais sensíveis ou de crianças e adolescentes, número do documento SEI em que foi aprovado o tratamento>**.

Comprometo-me a registrar as operações de tratamento no presente expediente, a informar aqui quaisquer alterações no plano de tratamento de dados, a obter as aprovações prévias eventualmente necessárias e a firmar termo de encerramento das operações de tratamento de dados tão logo estejam concluídos os trabalhos.

Tendo em vista os fundamentos legais mencionados anteriormente, declaro que as operações de tratamento de dados pessoais a serem realizadas dispensam o consentimento dos respectivos titulares.

<ASSINATURA ELETRÔNICA>

ANEXO III - TERMO DE ENCERRAMENTO DO TRATAMENTO E DE JUSTIFICATIVA DA CONSERVAÇÃO DE DADOS PESSOAIS

TERMO DE ENCERRAMENTO DO TRATAMENTO E DE JUSTIFICATIVA DA CONSERVAÇÃO DE DADOS PESSOAIS

Projeto	<nome do projeto>
Expediente SEI	<número do expediente>
Declarante	<nome, RF, cargo, órgão>, Coordenador(a) da Equipe de Desenvolvimento do Projeto

Na condição de coordenador(a) do projeto a que se refere o presente expediente, declaro que foram concluídas em **<data>** as operações de tratamento de dados pessoais a que se refere o documento **<número do documento SEI do termos de justificativa>**.

Tendo em vista o disposto no art. 16, inciso I, da Lei nº 13.709, de 14 de agosto de 2018 (LGPD), combinado com a parte final do art. 14 da Resolução CNJ nº 332, de 21 de agosto de 2020, os *datasets* resultantes das operações de tratamento de dados, compostos por **<descrever os datasets, com seu formato, teor e tamanho>** serão conservados em **<especificar local>**, conforme instruções contidas no documento **<número do documento SEI onde se encontram as orientações dada pelos órgão competente>**.

<ASSINATURA ELETRÔNICA>

ANEXO IV - MODELO DE RELATÓRIO PARCIAL DAS ATIVIDADES DE ANOTAÇÃO

(Equipe de Anotadores)

PROJETO XXX

RELATÓRIO DO XXX CICLO/ETAPA DE ANOTAÇÃO DE DADOS

1) Objetivo

2) Descrição da fonte de dados

[Descrever sucintamente origem, formato, conteúdo, quantidade, método de extração e local de armazenamento.]

3) Participantes e seus papéis

4) Atividades realizadas

[Mencionar datas, participantes, breve descrição.]

5) Ferramentas e procedimentos de anotação

[Descrever sucintamente as ferramentas utilizadas e os procedimentos adotados. Pode ser feita remissão a outros documentos.]

6) Problemas, incidentes e providências

[Informar datas, eventuais prejuízos, pessoas envolvidas e quaisquer outros dados relevantes. Pode ser feita remissão a outros documentos.]

7) Datasets resultantes

[Descrever sucintamente formato, conteúdo, quantidade, local de armazenamento e cautelas para preservação da integridade e controle de acesso.]

[data e assinatura pelos coordenadores da equipe de anotadores]

ANEXO V - MODELO DE RELATÓRIO PARCIAL DAS ATIVIDADES DE DESENVOLVIMENTO

(Equipe de Desenvolvedores)

PROJETO XXX

RELATÓRIO PARCIAL DAS ATIVIDADES DE DESENVOLVIMENTO Nº XXX

1) Descrição do(s) problema(s)

2) Datasets utilizados

[Mencionar origem, local de armazenamento, conteúdo, formato, finalidade do uso]

3) Abordagem(ns) utilizadas

[Mencionar algoritmos, pipelines, dependências, etc. Incluir remissão ao checkpoint correspondente no repositório do código-fonte. Ilustrar com fluxogramas, gráficos, slides ou o que mais se mostrar útil para a compreensão.]

4) Resultados e testes

[Descrever os testes adotados, se for o caso, mencionando as razões da escolha e os resultados obtidos. Incluir métricas de desempenho e fazer uma avaliação sucinta dos resultados obtidos em termos de evolução, expectativas e prognósticos.]

[data e assinatura pelos coordenadores da equipe de desenvolvedores]

ANEXO VI - MODELO DE RELATÓRIO FINAL SOBRE A FORMAÇÃO DOS DATASETS

(Equipes de Anotadores e Desenvolvedores)

PROJETO XXX

RELATÓRIO SOBRE A FORMAÇÃO DOS DATASETS

1) Introdução

[Explicar o objetivo do projeto como um todo, como a atividade de anotação se insere no contexto geral. Local adequado também para delimitação do escopo, alertas e ressalvas.]

2) Descrição dos datasets produzidos

[Mencionar formato, quantidade, local de armazenamento, meios de assegurar integridade e controle de acesso e quaisquer outras informações relevantes.]

3) Descrição das atividades, papéis e critérios de anotação

[Incluir prints de tela, guidelines, remissão a documentos e quaisquer outros recursos que facilitem a compreensão das atividades realizadas.]

3.1) Etapa de extração dos dados

3.2) Etapa de pré-processamento

3.3) Etapa de anotação

3.4) Etapa de curadoria

4) Ciclos de anotação e curadoria realizados

[Mencionar datas/períodos, participantes, modo de inclusão na equipe.]

5) Aspectos de integridade e segurança

[Mencionar os meios adotados para assegurar integridade e controle de acesso.]

[data e assinatura pelos coordenadores das equipes de desenvolvedores e anotadores]

ANEXO VII - LISTA DE QUESTÕES À EQUIPE DE DESENVOLVEDORES

1. Questões relacionadas ao escopo e à finalidade do projeto:

1.1. Qual problema pretendeu-se resolver com a solução desenvolvida?

1.2. Qual o comportamento esperado da solução desenvolvida?

1.3. Quais comportamentos não se esperam da solução desenvolvida? Quais as limitações existentes?

1.4. Quais cautelas devem ser adotadas pelos usuários dos modelos de IA para que a sua finalidade não seja desvirtuada?

2. Questões relacionadas aos usuários e ao contexto de uso:

2.1. A qual grupo ou a quais grupos de usuários a solução se destina?

2.2. Qual o contexto de uso para o qual a solução foi concebida?

3. Questões relacionadas aos *datasets* utilizados:

3.1. Quais as fontes dos *datasets* de treinamento, de validação e de testes? Quais os critérios utilizados para formação dos *datasets*?

3.2. Qual a estrutura dos *datasets* utilizados?

3.3. Onde os *datasets* estão armazenados? Quais medidas foram tomadas para assegurar a sua proteção e preservação de sua integridade para eventual auditoria pelos órgãos de controle?

3.4. Os *datasets* contêm dados sigilosos e/ou pessoais? Em caso afirmativo, foram obtidas autorizações ou pareceres para respaldar o seu uso no projeto?

3.5. Houve compartilhamento de dados com agentes externos à Justiça Federal da 3ª Região? Em caso afirmativo, quais as medidas tomadas para assegurar que os dados não fossem utilizados para finalidades estranhas ao escopo do projeto?

3.6. O uso da solução exigirá a coleta de dados de usuários? Que tipo de dados? Para qual finalidade?

4. Questões relacionadas à arquitetura e tecnologias adotadas:

4.1. Quantos modelos de IA foram desenvolvidos e qual a funcionalidade oferecida por cada um deles?

4.2. Quais espécies de algoritmos foram empregados em cada um dos modelos?

4.3. Onde está armazenado o código-fonte? Como se pretende manter e preservar a integridade do código-fonte para eventual auditoria pelos órgãos de controle?

4.4. Quais as dependências da solução? Descreva cada uma delas, com as respectivas versões, url dos repositórios utilizados e licenças de uso.

4.5. A reprodutibilidade do resultado (*output*) da solução depende de algum modo de contextos ou condições específicas ou depende exclusivamente dos dados (*input*) fornecidos?

5. Questões relacionadas aos testes da solução:

5.1. A solução desenvolvida foi testada no laboratório? Em caso afirmativo, quais foram as métricas utilizadas e os resultados obtidos?

5.2. Foram empregados testes automatizados no projeto? Em caso afirmativo, de que tipo? Onde estão armazenados os códigos-fonte utilizados para os testes?

5.3. Os resultados dos testes estão documentados? De que forma?

5.4. Quem participou dos testes e qual a função desempenhada?

ANEXO VIII - LISTA DE VERIFICAÇÃO PARA O GVEJ[74]

1. Respeito aos Direitos Fundamentais

1.1. Liberdades e direitos individuais

1.1.1. A solução de IA impede ou dificulta o acesso do jurisdicionado e demais atores do sistema judiciário ao juiz da causa?

1.1.2. A solução de IA impede ou dificulta que o juiz da causa decida com independência?

1.1.3. A solução de IA desrespeita o cidadão, tratando como "coisa" a ser examinada, triada, classificada, arregimentada, condicionada ou manipulada?

1.1.4. A solução de IA despreza ou ameaça a integridade física ou mental dos seres humanos, o seu sentido de identidade pessoal e cultural e a satisfação das suas necessidades essenciais?

1.1.5. A solução de IA despreza ou ameaça a autonomia individual (o direito de cada indivíduo de controlar a própria vida e decidir por si mesmo, sem coerção indevida ou manipulação)?

1.1.6. A solução de IA discrimina injustamente certos indivíduos ou grupos?

1.1.7. A solução de IA impede ou ameaça as liberdades de expressão, de crença, de empresa, de reunião ou de associação?

1.1.8. A solução de IA é contraditória com os valores do Estado Democrático de Direito?

1.1.9. A solução de IA tem algum outro tipo de impacto nos direitos fundamentais? Quais os potenciais conflitos constatados entre os diferentes princípios e direitos? Esses potenciais conflitos foram documentados?

1.2. Direitos coletivos e sociais

1.2.1. Foram criados mecanismos para medir o impacto ambiental do desenvolvimento, da implantação e da utilização da solução de IA (p. ex.,

[74] Esta lista de verificação baseia-se em grande parte na lista proposta pelo GPAN, mas foi bastante modificada. Segue a estrutura dos capítulos deste manual.

quantidade e tipo de energia utilizada pelo data center)?

1.2.2. Quais foram as medidas adotadas para reduzir o impacto ambiental do ciclo de vida da solução de IA?

1.2.3. Caso a solução de IA interaja diretamente com seres humanos:

a) Os seres humanos são estimulados a desenvolver laços e empatia com a solução?

b) A solução de IA demonstra de forma clara que a sua interação social é simulada e que não tem qualquer capacidade para "compreender" ou "sentir"?

1.2.4. Os impactos sociais da solução de IA são bem compreendidos? Há, por exemplo, risco de perda de postos de trabalho ou de perda de competências da mão de obra? Que medidas foram adotadas para combater tais riscos?

1.2.5. Foi avaliado o impacto social mais geral da utilização da solução de IA, para além do usuário final individual, como, por exemplo, sobre as partes interessadas que poderão ser indiretamente afetadas?

2. Não Discriminação

2.1. Qual estratégia e quais procedimentos foram adotados para evitar criar ou reforçar enviesamentos injustos na solução de IA, tanto no que respeita à utilização de dados de entrada como à concepção do algoritmo?

a) Foram avaliadas e reconhecidas as eventuais limitações decorrentes da composição dos datasets utilizados?

b) Foi considerada a diversidade e a representatividade dos usuários nos dados?

c) Foram realizados testes em relação a populações específicas ou a casos de uso problemáticos?

d) Quais ferramentas foram utilizadas para melhorar a compreensão dos dados, do modelo e do desempenho?

e) Quais processos foram criados para testar e controlar potenciais enviesamentos durante as fases de desenvolvimento, implantação e utilização da solução?

f) Se a solução se destina a apoiar decisões judiciais, quem avaliou ou avaliará a existência de enviesamentos? Foi tomada alguma medida para neutralizá-los ou

eliminá-los? Quem tomou ou determinou essas medidas estava legitimado para fazê-lo?

2.2. Existe algum mecanismo que permita que outras pessoas apontem eventual enviesamento, discriminação ou mau desempenho da solução de IA?

a) Qual o modo de suscitar essas questões? A quem podem ser apresentadas? O meio de fazê-lo é comunicado de forma clara?

b) Teve-se em conta, além dos usuários finais, outras pessoas que possam ser indiretamente afetadas pela solução de IA? Quais e em que casos de uso?

2.3. Verificou-se se pode ocorrer variabilidade das decisões em condições idênticas? Em caso afirmativo, quais as possíveis causas? Qual é o mecanismo de medição ou avaliação de potencial impacto dessa variabilidade nos direitos fundamentais?

2.4. Existe uma definição operacional adequada de "equidade" (fairness) que se aplique à concepção da solução de IA?

a) A definição é comumente utilizada? Foram avaliadas outras definições antes de se escolher a que está sendo utilizada?

b) Como é medida e testada a aplicação da definição de equidade escolhida? Quais os parâmetros utilizados para tanto? Existe análise quantitativa?

c) Quais mecanismos foram estabelecidos para garantir a equidade da solução de IA? Quais outros mecanismos foram considerados?

2.5. A solução de IA abrange uma gama adequada de preferências e capacidades individuais?

a) A solução de IA pode ser utilizada por pessoas com necessidades especiais ou deficiência, ou pessoas em risco de exclusão? De que forma foi essa possibilidade incorporada na concepção da solução e como é verificada?

b) As informações sobre a solução de IA também estão acessíveis a usuários com necessidades especiais?

c) Essa comunidade esteve envolvida ou foi consultada durante a fase de desenvolvimento da solução?

2.6. Foi considerado o impacto da solução de IA no grupo potencial de usuários?

a) A equipe de projeto é representativa do público-alvo de usuários? É representativa da população em geral, considerando também outros grupos que possam ser

indiretamente afetados?

b) Avaliou-se se poderão existir pessoas ou grupos desproporcionalmente afetados pelas implicações negativas?

c) Foram colhidas observações de outras equipes ou outros grupos que representam diferentes contextos e experiências?

2.7. Foi empregado algum mecanismo para incluir a participação das diferentes partes interessadas no desenvolvimento e na utilização da solução de IA?

2.8. Os servidores e magistrados afetados foram previamente informados e envolvidos no processo de criação e implantação da solução de IA?

3. Publicidade e Transparência

3.1. Rastreabilidade

3.1. Quais medidas foram adotadas para garantir a rastreabilidade? Verificar se foram documentados:

a) no caso de soluções de IA baseadas em regras: (i) o método de programação ou a forma como o modelo foi construído; e (ii) os cenários ou casos utilizados para testes e validação.

b) no caso de soluções de IA baseadas na aprendizagem de máquina: (i) o método de treinamento do algoritmo, incluindo os dados de entrada que foram extraídos e selecionados, e a forma como isso foi feito; e (ii) as informações sobre os dados utilizados para testes e validação.

c) em qualquer dos casos acima, os resultados ou as decisões tomadas pelo algoritmo, bem como outras decisões potenciais que resultariam de casos diferentes (p. ex., para outros subgrupos de usuários).

3.2 Auditabilidade:

3.2.1 Quais mecanismos foram criados para facilitar a auditabilidade da solução por auditores internos e/ou independentes, especificamente quanto à rastreabilidade e ao registo dos processos e resultados da solução de IA?

3.2.2. Houve a abertura de um expediente para documentação do projeto? O expediente é público ou sigiloso?

3.2.3. A documentação recomendada no Manual GVEJ foi produzida pela equipe

do projeto? Onde está armazenada? Existem mecanismos para evitar a sua perda ou adulteração intencional ou acidental?

4.3. Explicabilidade:

4.3.1. Em que medida as decisões e os resultados produzidos pela solução de IA podem ser compreendidos?

4.3.2. Em que medida a decisão da solução de IA influencia os processos de tomada de decisões nas unidades judiciárias ou administrativas?

4.3.3. Por que razão é necessário utilizar a solução de IA no domínio específico de negócios para a qual foi desenvolvida?

4.3.4. Qual o modelo de negócios da solução (p. ex., como é que ele cria valor para a Justiça Federal, para o jurisdicionado e para outros usuários externos)?

4.3.5. É possível tornar compreensíveis a todos os usuários os motivos por que a solução fez determinada escolha que levou a um determinado resultado?

4.3.6. A interpretabilidade da solução de IA foi levada em consideração desde o início do projeto?

a) Houve preocupação de investigar e utilizar o modelo mais simples e fácil de interpretar para o uso pretendido?

b) Os dados utilizados durante o treinamento e os testes são passíveis de análise? É possível alterá-los e atualizá-los ao longo do tempo?

c) Quais as formas de interpretabilidade consideradas para a solução desenvolvida e quais os critérios de escolha adotados?

4. Governança, Qualidade e Segurança

4.1. Resiliência perante ataques e segurança

4.1.1. Foram avaliadas as potenciais formas de ataque a que a solução de IA seria vulnerável? Em particular, foram considerados diferentes tipos de vulnerabilidades, como as relacionadas à poluição de dados, à infraestrutura física ou a ataques cibernéticos?

4.1.2. Quais medidas ou sistemas foram adotados para garantir a integridade e a resiliência da solução de IA contra potenciais ataques?

4.1.3. Avaliou-se como se comporta a solução em situações e ambientes inesperados?

4.1.4. Ponderou-se se a solução podia ou não, e até que ponto, ser utilizada de forma maliciosa, para fins não desejados ou não contemplados no projeto?

4.2. Plano de contingência e segurança geral

4.2.1. Existe um plano de contingência adequado para o caso de ataques maliciosos ou outras situações inesperadas (p. ex., solicitação da intervenção de um operador humano antes de prosseguir)?

4.2.2. Qual o nível de risco para o caso do item anterior?

a) Há algum processo para medir e avaliar os riscos e a segurança?

b) Que informações são dadas aos usuários quanto aos riscos para a integridade física dos seres humanos?

c) Houve aquisição de apólice de seguro para cobrir eventuais danos causados pela solução de IA?

d) Foram identificados os potenciais riscos de segurança de outros usos previsíveis da tecnologia, incluindo o mau uso acidental ou doloso? Existe algum plano para atenuar ou gerir esses riscos?

4.2.3. Existe alguma probabilidade de que o sistema de IA cause danos ou prejuízos aos usuários ou a terceiros? Em caso afirmativo, qual a probabilidade, os potenciais danos, o público afetado e a gravidade?

a) Caso existam riscos de danos, analisou-se a regulamentação em matéria de responsabilidade e de defesa do usuário da solução? De que modo se teve em conta essa regulamentação? Houve análise da cadeia de imputabilidade?

b) A análise de risco abrangeu condições de segurança de informação e de infraestrutura (p. ex., potenciais riscos para a segurança cibernética) que, associados a um comportamento não intencional da solução de IA, poderiam pôr em risco a segurança dos usuários ou lhes causar danos?

4.2.4. Houve estimativa do impacto provável de uma falha da solução de IA que a leve a fornecer resultados inválidos, que a torne indisponível ou que a faça fornecer resultados inaceitáveis do ponto de vista social (p. ex., práticas discriminatórias)?

a) Foram definidos protocolos para acionar planos alternativos ou de contingência caso se verifiquem os cenários acima?

b) Houve definição, implementação e testes de planos de contingência?

4.3. Desempenho

4.3.1. Qual o nível de desempenho necessário no contexto em que a solução de IA será utilizada?

a) Como será medido e assegurado esse nível de desempenho?

b) Quais medidas adotadas para assegurar que os dados utilizados são suficientemente abrangentes e atualizados?

c) Quais medidas adotadas para avaliar se são necessários dados adicionais, por exemplo para melhorar a acurácia ou eliminar os enviesamentos?

4.3.2. Como são avaliados os danos que podem ser causados se a solução de IA fizer previsões inválidas?

4.3.3. Quais são as formas de medir se a solução está produzindo um número inaceitável de previsões inválidas?

4.3.4. Quais medidas foram adotadas para melhorar a acurácia da solução?

4.3.5. As medidas adotadas são validadas por pessoa legitimada a decidir a questão?

4.4. Confiabilidade e reprodutibilidade:

4.4.1. Que estratégia foi adotada para controlar e testar se a solução de IA cumpre os seus objetivos, finalidades e usos previstos?

a) É necessário ter em conta contextos ou condições específicas de uso para garantir a reprodutibilidade dos resultados obtidos (output)?

b) Quais processos ou métodos de verificação foram adotados para medir e assegurar os diferentes aspectos da confiabilidade e da reprodutibilidade?

c) Quais processos foram adotados para definir se a solução de IA falha em certos contextos de uso?

d) Os processos adotados para testar e verificar a confiabilidade da solução de IA foram documentados e implementados?

e) Quais mecanismos ou formas de comunicação foram adotados para garantir aos usuários finais a confiabilidade da solução de IA?

4.5. Qualidade e integridade dos dados

4.5.1. A solução está em harmonia com as normas de segurança e gestão de dados em vigor na Justiça Federal da 3ª Região?

4.5.2. Foram criados mecanismos de supervisão para a extração, a conservação, o tratamento e a utilização de dados? Quais?

4.5.3. Existe controle da qualidade dos dados obtidos de fontes externas?

4.5.4. Quais processos foram adotados para garantir a qualidade e a integridade dos dados? Foram avaliados outros processos? Qual solução foi adotada para verificar se os datasets não foram comprometidos ou objeto de pirataria informática?

4.6. Acesso aos dados:

4.6.1. Que protocolos, processos e procedimentos foram seguidos para assegurar a gestão adequada dos dados?

a) Avaliou-se quem pode acessar os dados dos usuários e em que circunstâncias?

b) Verificou-se se essas pessoas são qualificadas, se realmente necessitam de acesso aos dados, se têm as competências necessárias para compreender a política de proteção de dados em detalhes?

c) Existe algum mecanismo de supervisão para registrar quando, onde, como, por quem e para que fim os dados foram acessados?

5. Controle do Usuário

5.1. Ação humana:

5.1.1. A solução de IA interage com a tomada de decisões por usuários finais humanos (p. ex., recomendação de ações ou decisões a tomar, apresentação de opções)? Nesses casos, existe algum risco de que a solução de IA afete a autonomia humana, interferindo com o processo decisório do utilizador final de forma não intencional?

5.2.1. Em se tratando de solução de IA introduzida num processo de trabalho,

ponderou-se a distribuição de tarefas entre a solução de IA e os trabalhadores humanos no que diz respeito a interações significativas e a uma supervisão e um controle adequados por seres humanos?

a) A solução de IA melhora ou aumenta as capacidades humanas?

b) Quais as salvaguardas adotadas para evitar o excesso de confiança ou o excesso de dependência face à solução de IA nos processos de trabalho?

5.2. Supervisão humana:

5.2.1. Qual seria o nível adequado de controle humano para a solução de IA?

a) Qual o nível de controle ou envolvimento humano, se aplicável? Quem é o "ser humano no controle" e quais são os momentos ou os mecanismos para a intervenção humana?

b) Quais mecanismos e medidas foram criados para assegurar esse potencial controle ou supervisão por seres humanos, ou para garantir que as decisões sejam tomadas sob a responsabilidade global de seres humanos?

c) Quais medidas foram adotadas para permitir a auditoria e corrigir questões relacionadas à governança da autonomia da IA?

5.2.2. No caso de utilização de técnicas de aprendizagem de máquina, quais mecanismos específicos de controle e de supervisão foram adotados?

a) Quais mecanismos de detecção e de resposta foram estabelecidos para avaliar se algo poderia falhar?

b) Foi previsto um "botão de parada" ou procedimento para abortar a operação de forma segura, se necessário? Esse procedimento aborta o processo por completo, parcialmente ou delega o controle a um ser humano?

c) Foram previstos mecanismos para que o ser humano assuma o controle da atividade se necessário?

5.3. Informação:

5.3.1. Os usuários são informados de que estão interagindo com uma solução de IA e não com outro ser humano?

a) Ponderou-se se a solução de IA deveria comunicar aos usuários que uma decisão, um conteúdo, um conselho ou um resultado provém de uma decisão

algorítmica?

b) Caso a solução de IA inclua um sistema de conversação automática (chatbot), os usuários finais humanos foram informados do fato de estarem interagindo com um agente não humano?

5.3.2. Quais mecanismos foram criados para informar os usuários acerca das razões e dos critérios subjacentes aos resultados da solução de IA?

a) Essa informação é transmitida de forma clara e inteligível aos usuários?

b) Existem meios para receber sugestões e comentários de usuários e para aperfeiçoar a solução a partir dessas sugestões e comentários?

c) Os usuários são informados dos riscos potenciais ou percebidos, tais como os riscos de enviesamento?

d) Foi considerada a necessidade de prestar informações e oferecer transparência a terceiros e ao público em geral?

5.3.3. Há explicação clara sobre a finalidade da solução de IA e sobre quais são os seus eventuais beneficiários?

a) Foram claramente especificados os casos de uso aos quais a solução se destina?

b) A explicação está formulada de forma clara e de fácil assimilação e compreensão, levando em conta os seus destinatários e a necessidade de eliminar enviesamentos cognitivos?

5.3.4. As características, as limitações e as potenciais insuficiências da solução de IA são informadas de modo claro seja a quem irá implantar a solução seja ao usuário final?

6. Pesquisa, Desenvolvimento e Implantação de Serviços de IA

6.1. Equipe do projeto

6.1.1 Como foi selecionada a equipe do projeto? Quais os critérios adotados?

6.1.2. Houve divisão de atribuições? Quais subgrupos foram criados e qual a função de cada um? Qual o critério adotado para atribuição das tarefas?

6.1.3. Houve preocupação em orientar a formação da equipe pela busca da mais

ampla diversidade em todas as etapas do processo e pela multidisciplinaridade? Em caso negativo, essa posição está respaldada em decisão fundamentada (p.ex, pela ausência de profissionais no quadro de pessoal)?

6.2. Governança

6.2.1 Se a solução contemplar o reconhecimento facial, houve obtenção de prévia autorização do CNJ?

6.2.2. Se o projeto envolver matéria penal, especialmente sugestão de modelos de decisões preditivas (ressalvados os projetos que visem à automação e ao oferecimento de subsídios destinados ao cálculo de penas, prescrição, verificação de reincidência, mapeamentos, classificações e triagem dos autos para fins de gerenciamento de acervo), ponderou-se acerca da recomendação do CNJ de que projetos dessa espécie fossem evitados? Quais as razões que motivaram a continuação do projeto?

6.3. Qualidade Técnica

6.3.1. Houve utilização preferencial de software de código aberto com características que permitam:

a) A adoção dos padrões de interoperabilidade definidos pelo CNJ para o SINAPSES?

b) A utilização de ambiente de desenvolvimento colaborativo definido pelo LIAA-3R?

c) Observar as exigências de rastreabilidade, auditabilidade e explicabilidade decorrentes do dever de transparência?

d) A cooperação entre outros segmentos e áreas do setor público e a sociedade civil?

6.3.2. Foi definida alguma licença a ser aplicada à solução de IA desenvolvida?

6.3.3. Foram documentadas todas as licenças das dependências utilizadas no projeto? Elas são compatíveis com a licença aplicada à solução de IA desenvolvida?

7. Prestação de Contas e Responsabilização

7.1 Minimização e comunicação dos impactos negativos

7.1.1 Foi realizada avaliação de riscos ou de impacto da solução de IA que leve em conta as diferentes partes direta ou indiretamente afetadas?

7.1.2 Houve ações de capacitação relacionadas a práticas de prestação de contas? Quais servidores, magistrados e terceiros, membros ou não da equipe, estão envolvidos? Essas ações vão além da fase de desenvolvimento? Essas ações de capacitação abrangem aspectos jurídicos do projeto?

7.1.3 Complementarmente às iniciativas ou aos órgãos de supervisão internos, há algum tipo de orientação externa? Foram também criados processos de auditoria?

7.1.4 Existem meios para que servidores, magistrados e quaisquer terceiros comuniquem eventuais vulnerabilidades, riscos ou enviesamentos na solução de IA?

7.2. Documentação de soluções de compromisso (tradeoffs):

7.2.1 Houve necessidade de sacrificar interesses e valores para promover outros considerados mais importantes?

7.2.2 O sopesamento dos interesses e valores em jogo foi documentado? Como foram feitas as escolhas? Por quem?

8. Diretrizes Específicas de Conformidade

8.1. Aprovações

8.1.1. O projeto recebeu autorização e/ou parecer favorável de quais órgãos internos e externos? As cópias das autorizações e dos pareceres integram a documentação do projeto?

8.1.2. Houve comunicação ao CNJ?

8.1.3. Houve registro no PDP3R?

8.1.4. Haverá necessidade de novas autorizações? Em caso afirmativo, quais e por quê? Como e em que prazo serão obtidas?

8.1.5. A solução encontra-se em produção? Em caso afirmativo, qual foi o órgão responsável pela implantação? Foram obtidas as aprovações necessárias para tanto? Quais e onde estão documentadas? Quais as recomendações comunicadas a esse órgão?

8.2. Documentação

8.2.1. O código-fonte e os datasets são mantidos em repositórios indicados pela coordenação do LIAA-3R?

8.2.2. Foram incluídos no expediente SEI do projeto todos os artefatos de documentação previstos no Manual GVEJ, assim como registrados (i) a lista completa de dependências, com suas respectivas licenças de uso; (ii) os testes realizados e seus respectivos resultados; (iii) os meios de comunicação utilizados para troca de informações pela equipe e com atores externos; e (iv) eventual participação de atores externos, com menção ao papel que tiveram no projeto e eventual acesso desses atores a dados pessoais ou sigilosos?

8.3. Conflito de interesses

8.3.1. A solução desenvolvida pode de algum modo beneficiar pessoalmente algum dos membros da equipe de projeto, do LIAA-3R ou da Justiça Federal da 3ª Região ou algum de seus familiares, cônjuges, companheiros ou amigos íntimos?

8.3.2. Houve atuação de integrantes do GVEJ em alguma das outras equipes do projeto? Em caso afirmativo, foram identificados potenciais conflitos de interesse quanto à validação ético-jurídica dos modelos de IA?

9. Diretrizes Referentes à LGPD

9.1. Qual o tipo e o escopo dos dados que compõem os datasets (p. ex., se contêm dados pessoais)?

9.2. Foram avaliadas formas de desenvolver a solução de IA sem a utilização ou com uma utilização mínima de dados potencialmente sensíveis ou pessoais?

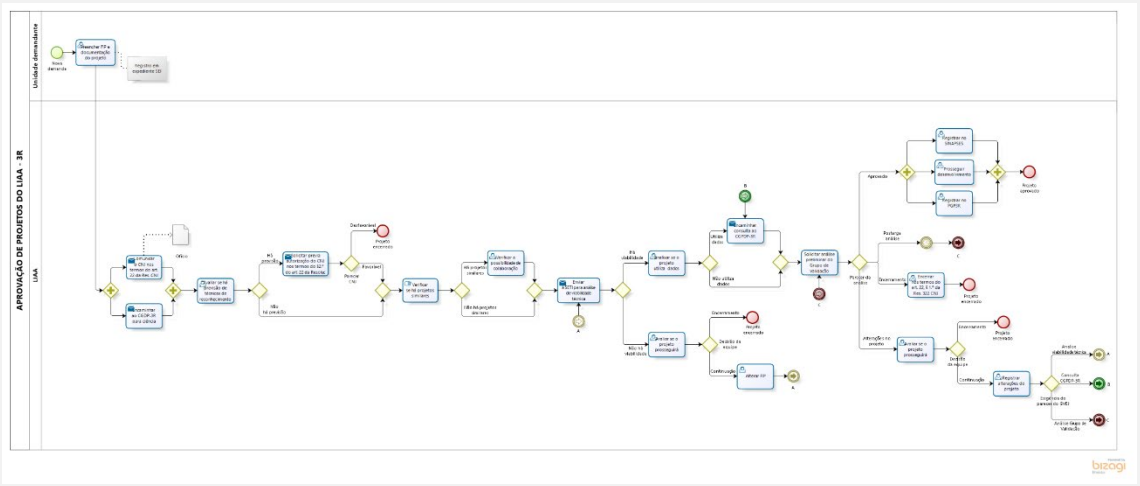
9.3. Quais os mecanismos criados para identificar e controlar dados pessoais em cada caso de uso (tais como o consentimento válido e a possibilidade de revogação, quando aplicável)?

9.4. Foram adotadas medidas para aumentar a privacidade, tais como a encriptação, a anonimização e a agregação?

9.5. Foram obtidas as devidas autorizações para tratamento de dados pessoais ou sigilosos?

9.6. Quais mecanismos foram criados para que outras pessoas possam informar problemas de privacidade ou proteção de dados relacionados com os processos de extração (para treinamento e funcionamento) e de tratamento de dados?

ANEXO IX - FLUXOGRAMA DE APROVAÇÃO DE PROJETOS



Clique na imagem para acessar o fluxograma na internet
ou use o QRCode para acessá-lo pelo celular.