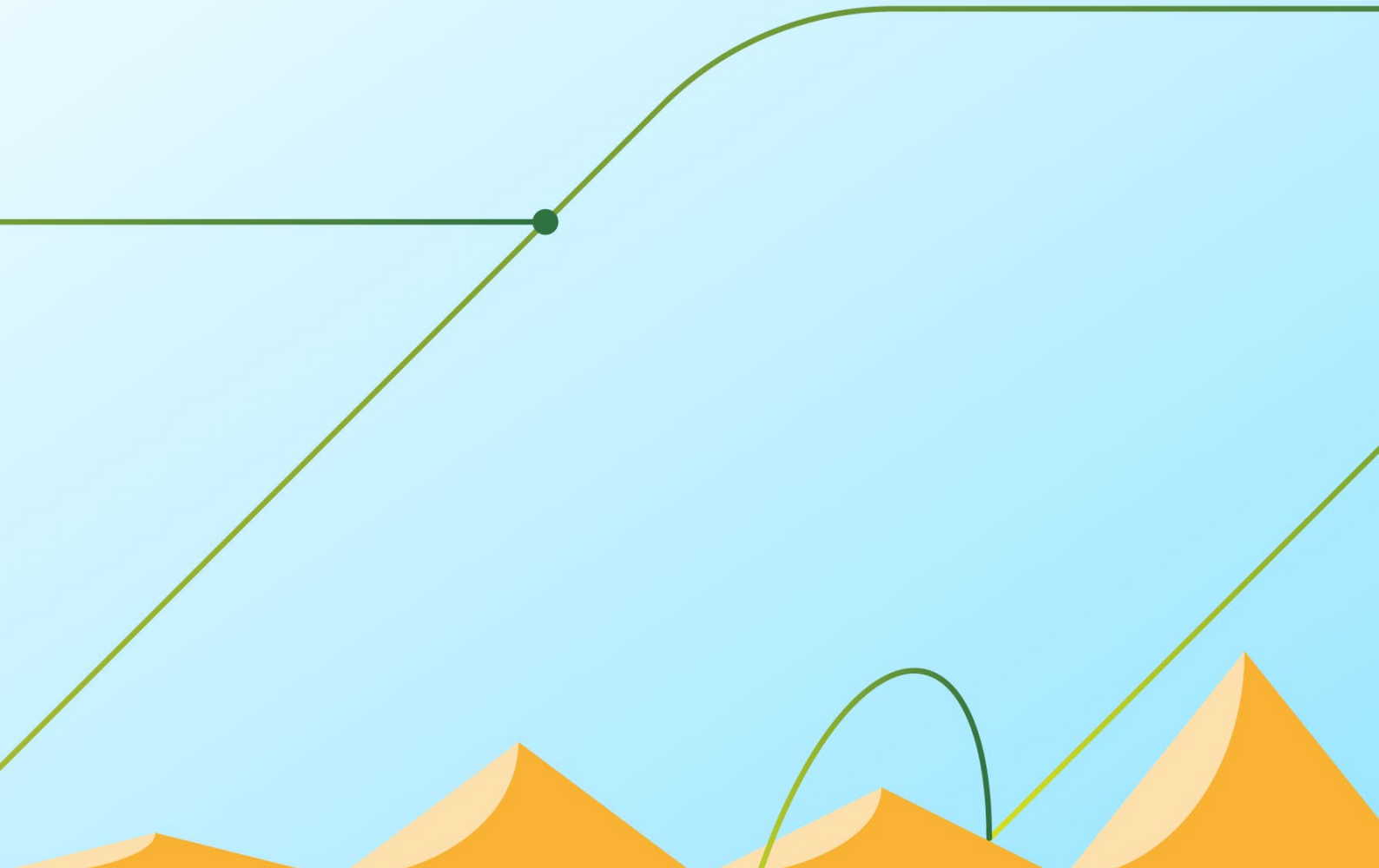
ANPD

sandbox regulatório

4º Relatório Parcial de Monitoramento

Fase de Teste – Ciclo 4



Agência Nacional de Proteção de Dados (ANPD)

Diretor-Presidente
Waldemar Gonçalves Ortunho Junior

Diretores
Iagê Zendron Miola
Lorena Giuberti Coutinho
Miriam Wimmer

Comissão de Sandbox
Diego André Sant'Ana
Fábio Silveira Vidal
Lucas Costa dos Anjos
Maria Carolina Ferreira da Silva
Marcus Vinícius Rossi da Rocha
Jeferson Dias Barbosa

Equipe de Supervisão
Ticianne de Gois Ribeiro Darin
Uberson Rossa

Equipe de Apoio Administrativo
Gabriel Carvalho Moretto
Marcia Valéria Paixão
Raphael Gonçalves Vanderlei

ANPD
SCN, Qd. 6, Conj. A, Ed. Venâncio 3000, Bl. A, 9º andar
Brasília, DF Brasil 70716-900
t. (61) 2025-8101 www.gov.br/anpd

Centro de Inteligência Artificial e Aprendizado de Máquina / Universidade de São Paulo (CIAAM/USP)

Coordenadora:
Cristina Godoy Bernardo de Oliveira

Responsáveis pela Equipe Técnica:
Anarosa Brandão
Juliano Maranhão
Carlos Portugal
João Paulo Cândia Veiga

Pesquisadores da Área Técnica:
Bruno Albertini
Luciano Gallegos

Pesquisadores da Área Regulatória:
Fernanda Valle Versiani
Paloma Saldanha
Rafael Meira Silva
Giovana Carneiro
Thiago Marcílio
Anthony Novaes

Consultores Internacionais
João Pita Costa (IRCAI UNESCO)
Vasilka Sancin (Universidade de Ljubljana e Juíza da Corte Europeia de Direitos Humanos)

Colaboradores HARPIA
Sarajane Peres (Pesquisadora Responsável USP)
Vinícius B. Matos (Pesquisador)
Ricardo Grava (Pesquisador)

Colaboradores Médicos (Testes de LLM na Área de Saúde)
Daniele Thais de Almeida Santos
Paulo Victor Gomes Ferreira
Haroldo Rodrigues
Matheus Ananias
Giuliano Bongiovanni

Sumário

1. Introdução 5

2. Planejamento e estruturação das sessões de orientação para os participantes do Sandbox Regulatório 7

3. Definição dos temas e metodologia para as sessões remotas e presenciais 8

4. Apoio, orientação e supervisão das atividades de testagem 9

4.1 Synapse 9

4.1.1 Documentação principal 10

4.1.2 Fase A 10

4.1.3 Fase B 13

4.1.4 Fase C 17

4.2 Prevvine 19

4.2.1. Documentação Principal 20

4.2.2 Fase A 21

4.2.3 Fase B 25

4.2.4 Fase C 29

4.3 Metatext 32

4.3.1 Documentação principal 32

4.3.2 Fase A 32

4.3.3 Fase B 33

4.3.4 Fase C 36

5. Consulta Pública 38

5.1. Consulta multissetorial 38

5.2. Conclusões gerais da consulta pública 41

6. Conclusões gerais do Ciclo 4 44

1. Introdução

Este relatório apresenta o planejamento, a orientação e a supervisão das atividades do Ciclo 4 do Projeto Piloto de Sandbox Regulatório de Inteligência Artificial e Proteção de Dados, reunindo os resultados técnicos e regulatórios das participantes Synapse[1], com o projeto Trajetto; Prevvine[2], com a plataforma STAiDOC; e Metatext[3], com a solução GuardionAI. Seu objetivo é registrar os avanços, os riscos, as limitações e os aprendizados identificados, fundamentando os encaminhamentos para o Ciclo 5.

O acompanhamento manteve a metodologia[4] estruturada nas Fases A — Pré-teste, B — Teste e C — Pós-teste, com execução adaptada às características e ao estágio de cada solução. As atividades compreenderam orientações, análises documentais, preparação e execução de testes, avaliação de riscos e acompanhamento de condicionantes. As evidências consideradas incluíram as fichas A1, A2, A3 e A4, B1, B2 e C1, pareceres, documentos de arquitetura e governança, avaliações éticas, matrizes de transparência, registros da HarpIA, fichas de êxitos[5] e planos de descontinuidade, conforme disponibilidade para cada participante.

Na Fase A, foi utilizado um *framework* geral baseado em recomendações da UNESCO[6] (doravante tratado como *framework* geral), o qual estrutura a avaliação ética de sistemas de IA em sete eixos ponderados, totalizando 100 pontos:

- Eixo 1 — Centralidade da dignidade humana e dos direitos fundamentais (20 pontos);
- Eixo 2 — Governança e responsabilidade clara (15 pontos);
- Eixo 3 — Avaliação de impactos (15 pontos);
- Eixo 4 — Justiça, não discriminação e vieses algorítmicos (15 pontos);
- Eixo 5 — Transparência, explicabilidade e comunicação (15 pontos);
- Eixo 6 — Segurança técnica e robustez (10 pontos); e
- Eixo 7 — Proporcionalidade e necessidade (10 pontos).

---

[1] A Synapse AI é a desenvolvedora do projeto Trajetto, uma plataforma inteligente de gestão de fluxo de pessoas em transportes ferroviários, baseada na integração de dados em tempo real.

[2] A Prevvine Tecnologia é a desenvolvedora da plataforma STAIDOC, um ecossistema de inteligência artificial médica com "*Large Language Models*" personalizados para o contexto brasileiro.

[3] A Metatext é a desenvolvedora da solução Guardion.AI, uma plataforma de segurança para inteligência artificial, projetada para estabelecer o controle e mitigar riscos em sistemas de agentes autônomos e IA generativa.

[4] A metodologia de testagem desenvolvida para supervisionar e acompanhar os testes do Projeto Piloto de Sandbox Regulatório de Inteligência Artificial e Proteção de Dados está disponível neste link: https://www.gov.br/anpd/pt-br/assuntos/projetos-acoes-iniciativas/sandbox/metodologia-sandbox.pdf/@@display-file/file, acessado em: 21 de outubro de 2026

[5] Para fins do Sandbox Regulatório, o êxito corresponde à produção de evidências e aprendizados sobre os avanços, riscos, limites, condicionantes e requisitos para a evolução segura da solução. Sua classificação restringe-se ao ciclo e ao escopo efetivamente examinado, sem representar aprovação plena da solução, comprovação de conformidade integral ou autorização para uso irrestrito.

[6] UNESCO — *United Nations Educational, Scientific and Cultural Organization* ou Organização das Nações Unidas para a Educação, a Ciência e a Cultura.

O *framework* é um instrumento setorialmente neutro, aplicável a sistemas de naturezas diversas, cabendo ao avaliador ponderar as particularidades do setor, finalidade e criticidade em cada caso concreto.

Na Synapse, o Ciclo 4 aprofundou a análise documental e regulatória, e acrescentou esclarecimentos à matriz de transparência. A pontuação no *framework geral* passou de 68 para 78 pontos, com acréscimos nos Eixos 3, 5 e 6, mantendo-se" a avaliação "suficiente com condicionantes".

Na Prevvine, foram realizadas avaliações clínicas na plataforma HarpIA, acompanhadas de avanços documentais e medidas corretivas. A pontuação decorrente da aplicação do *framework* geral aumentou. O parecer manteve conclusão favorável com condicionantes. Na Metatext, foi atingido o objetivo de iniciar os testes na plataforma HarpIA, com execução do Teste 1, relativo à proteção de dados, e preparação do Teste 2, composto por trinta alarmes destinados à equipe de segurança. Os resultados detalhados do primeiro teste, a execução do segundo e a avaliação conclusiva da Fase C ficaram previstos para o Ciclo 5, considerando a complexidade técnica e regulatória. O parecer recomendou continuidade com condicionantes.

Para o Ciclo 5, os encaminhamentos prioritários foram tratados no item 6 deste relatório, no qual foram detalhadas as medidas sugeridas. Essas medidas abrangem, entre outros temas, transparência, inclusão e acessibilidade, segurança da informação e gestão de terceiros.

A principal contribuição do Ciclo 4 foi ampliar a base de evidências e delimitar as pendências de comprovação. A leitura dos resultados deve considerar o escopo testado e os limites de cada avaliação, distinguindo controles verificados, medidas documentadas ou declaradas e providências ainda sujeitas à validação. A continuidade experimental não equivale à aprovação das soluções. O Ciclo 5 deverá reunir novos resultados, principalmente da Metatext e da Prevvine, permitindo aprofundar a análise técnica e regulatória da efetividade das salvaguardas e dos riscos remanescentes.

O relatório também apresenta a metodologia da Consulta Pública Multissetorial, voltada à identificação de desafios e possíveis aperfeiçoamentos regulatórios. O evento ocorreu via webconferência e teve participação de representantes da academia, setor privado, setor público e sociedade civil. Tal diversidade permitiu a discussão construtiva sobre temas transversais como transparência, governança de dados, impactos desproporcionais e lacunas regulatórias.

Para fins deste relatório, os resultados atribuídos ao Ciclo 4 consideram as atividades e evidências efetivamente produzidas e examinadas nessa etapa. Informações ou documentos recebidos posteriormente são identificados como elementos complementares, cuja avaliação deverá ocorrer nos ciclos subsequentes, preservando a delimitação temporal e a rastreabilidade dos achados.

2. Planejamento e estruturação das sessões de orientação para os participantes do Sandbox Regulatório

Preliminarmente, registra-se que, em razão dos resultados alcançados e da experiência adquirida no Ciclo 3, a estrutura das sessões e das interações foi mantida no Ciclo 4, com ajustes conforme as necessidades de cada participante. O planejamento concentrou-se principalmente na execução dos testes da Prevvine e da Metatext, responsáveis pelas soluções STAiDOC e GuardionAI, respectivamente, além do aprofundamento da análise regulatória mediante solicitações contínuas de documentos e esclarecimentos para avaliação pela equipe do CIAAM/USP.

Nesse sentido, além das reuniões previstas no calendário, o acompanhamento compreendeu encontros individualizados com as participantes, discussões internas entre as equipes técnica e regulatória e intercâmbio de informações pelo Microsoft Teams. As sessões buscaram assegurar diálogo direto e objetivo entre CIAAM/USP, participantes e ANPD, permitindo esclarecer dúvidas, orientar a produção de evidências, ampliar os achados e melhorar os resultados do Sandbox Regulatório de Inteligência Artificial e Proteção de Dados.

As bases estabelecidas nos ciclos anteriores permitiram direcionar as sessões ao acompanhamento da testagem e à análise das evidências. Na Prevvine, o foco esteve nas avaliações realizadas na HarpIA e na verificação das condicionantes técnicas e regulatórias. Na Metatext, as atividades abrangeram a preparação e a execução do Teste 1, relativo à proteção de dados, e a preparação do Teste 2, destinado à equipe de segurança, cuja execução ocorrerá no Ciclo 5. Para a Synapse, manteve-se o aprofundamento documental e regulatório, especialmente quanto às medidas ainda dependentes de comprovação.

Assim, o planejamento e a estruturação das sessões no Ciclo 4 articularam orientação, execução de testes, solicitação de documentos e análise técnica e regulatória. As entregas compreenderam o calendário de acompanhamento, os instrumentos de testagem, os pareceres e os registros consolidados neste relatório. Esse processo permitiu delimitar as pendências e organizar as próximas atividades, especialmente a análise dos resultados detalhados do Teste 1 da Metatext, a execução do Teste 2 e o aprofundamento dos testes técnicos e regulatórios da Prevvine no Ciclo 5.

3. Definição dos temas e metodologia para as sessões remotas e presenciais

No âmbito do quarto ciclo da fase de testes, foi mantida a base metodológica de supervisão e acompanhamento adotada nos produtos anteriores, estruturada nas Fases A — Pré-teste, B — Teste e C — Pós-teste. Nesse contexto, a metodologia foi aprimorada mediante interação constante entre as participantes, os pesquisadores das equipes técnica e regulatória do CIAAM/USP e a Comissão de Sandbox da ANPD, considerando as dificuldades, os resultados e as necessidades identificadas durante a experimentação.

A definição dos temas combinou eixos comuns, especialmente, proteção de dados, segurança da informação, governança, transparência, explicabilidade, supervisão humana e direitos dos titulares, com questões específicas de cada solução. Para a Synapse, destacaram-se geolocalização, acessibilidade, integrações com terceiros e reidentificação indireta. Para a Prevvine, foram priorizadas a qualidade das respostas clínicas e bibliográficas, a supressão de identificadores e as condicionantes de segurança e continuidade. Para a Metatext, o foco compreendeu a proteção de dados nas mensagens, as informações destinadas à equipe de segurança e o esclarecimento dos fluxos e das responsabilidades. Esses temas orientaram as solicitações documentais, os instrumentos de testagem, as matrizes de transparência e os pareceres apresentados no relatório.

Assim, as sessões continuaram a combinar orientação, análise documental, discussão dos testes e registro de encaminhamentos, prioritariamente por reuniões remotas individualizadas e comunicação pelo Microsoft Teams. A construção metodológica congregou objetivos complementares: apoiar as participantes na avaliação de suas soluções inovadoras; permitir aos pesquisadores examinar conjuntamente os aspectos técnicos e regulatórios; e produzir evidências e aprendizados úteis à atuação da ANPD. O diálogo direto entre esses grupos possibilitou ajustar perguntas, solicitar documentos adicionais, esclarecer divergências e aperfeiçoar os procedimentos conforme os achados.

Ademais, a organização do calendário permaneceu flexível para acomodar o desenvolvimento dos testes e as demandas de avaliação, preservando a relação entre os Planos de Sandbox, as atividades realizadas e as evidências disponíveis. A Consulta Multissetorial descrita no item 5 reuniu perspectivas da academia, do setor privado, do setor público e da sociedade civil sobre transparência e supervisão, anonimização e governança de dados, discriminação e impactos desproporcionais e lacunas regulatórias. Sua metodologia complementou o acompanhamento individualizado ao ampliar o diálogo sobre os desafios identificados nas soluções.

Finalmente, no Ciclo 4, a metodologia articulou a definição dos temas às entregas efetivamente realizadas: análises documentais e regulatórias, avaliações clínicas da Prevvine na plataforma HarpIA, execução do Teste 1 da Metatext, preparação do Teste 2, atualização das matrizes e consolidação dos pareceres e encaminhamentos. Essa construção contínua permitiu distinguir atividades concluídas de avaliações ainda pendentes e organizar o aprofundamento previsto para o Ciclo 5, especialmente quanto aos resultados dos testes técnicos e regulatórios da Prevvine e da Metatext.

4. Apoio, orientação e supervisão das atividades de testagem

Neste item do relatório são apresentadas as atividades de apoio, orientação e supervisão desenvolvidas pela equipe do CIAAM/USP e pela ANPD durante o Ciclo 4, em continuidade ao acompanhamento dos ciclos anteriores. Essas atividades compreenderam a preparação e a execução dos testes, a análise documental e o acompanhamento das condicionantes técnicas e regulatórias, conforme o estágio de cada participante.

O objeto e o escopo dos testes foram definidos considerando as características da inovação, o nível de maturidade tecnológica e os riscos identificados em cada solução. Assim, os resultados alcançados no presente Ciclo 4 devem ser interpretados à luz das condições de testagem e das evidências disponíveis, distinguindo as atividades executadas das avaliações ainda pendentes de consolidação no Ciclo 5.

4.1 Synapse

O relatório do CIAAM/USP referente ao Ciclo 4 apresenta os principais dados, informações e aprendizados relativos à solução Trajetto, considerando os seguintes pilares da metodologia de avaliação: dimensão técnico-funcional: análise das fichas A1 e A2, referentes às características da solução, à arquitetura e aos fluxos de dados;

- dados, privacidade e segurança: análise das informações e evidências registradas nas fichas das subfases A3, A4, B1 e B2, incluindo as salvaguardas declaradas e sua comprovação;
- maturidade tecnológica (TRL[7] e ETC[8]): consideração da autoavaliação apresentada pela participante nas fichas A1 e A2 e da análise realizada pela equipe do CIAAM/USP nas fichas B1 e B2, conforme as evidências disponíveis;
- realismo experimental e benefícios esperados: análise dos testes e das evidências da Fase B, preservando as conclusões e os limites dos ensaios técnicos encerrados no Ciclo 3, sem atribuir ao Ciclo 4 novos resultados desses testes;
- riscos, mitigação e impacto residual: acompanhamento das avaliações de riscos da subfase B2, das medidas de mitigação e das pendências de comprovação de sua efetividade;
- governança, transparência e aprendizagem regulatória: análise da documentação, da estrutura organizacional, dos papéis e responsabilidades, da explicabilidade e da matriz de transparência, especialmente no âmbito da subfase B1.

[7] TRL — *Technology Readiness Level* ou Nível de Maturidade Tecnológica.
[8] ETC — Elementos Tecnológicos Críticos.

Considerando o acompanhamento da Synapse nas Fases A, B e C do Ciclo 4, apresentam-se, a seguir, os principais aspectos identificados em cada etapa, distinguindo os avanços registrados das medidas ainda dependentes de comprovação.

4.1.1 Documentação principal

Preliminarmente, observa-se que os principais documentos referidos neste item do relatório encontram-se organizados conforme a descrição a seguir, por fases, para atendimento dos requisitos pertinentes ao Ciclo 4 deste Sandbox Regulatório:

- Fase A:
  - Fichas A1 e A2 — Aplicação e Matriz TRL;
  - Fichas A3 e A4 — Tabela resumo da evolução dos ciclos; e
  - Parecer regulatório dos documentos enviados pela Synapse.
- Fase B:
  - Ficha B1 — *Framework* geral;
  - Ficha B2 — Classificação de riscos;
  - Parecer técnico consolidado; e
  - Matriz de transparência.
- Fase C:
  - Ficha C1 — Êxitos para a ANPD e para a empresa participante; e
  - Plano de descontinuidade.

4.1.2 Fase A

Inicia-se a análise das fichas técnicas de aplicação (A1 e A2), sob responsabilidade da empresa. Quanto à arquitetura do modelo, a Synapse informou que, entre os Ciclos 3 e 4, passou a trabalhar na experiência do usuário e na solução para criação de mapas geolocalizados, esclarecendo que essa funcionalidade ainda não havia sido aplicada. Em relação ao fluxo de dados e às API, a participante declarou não ter havido mudanças entre os dois ciclos. A ficha também reúne respostas a questionamentos sobre o processamento de rotas na ausência de dados, o risco de identificação indireta e o tratamento das imagens provenientes de câmeras. Essas respostas esclarecem aspectos da solução, mas ainda demandam verificação quanto à implementação das medidas descritas, especialmente no que se refere aos mapas geolocalizados, à variável relativa à condição de pessoa com deficiência e ao fluxo de dados das API e de terceiros.

Da mesma maneira, a empresa declarou que os protocolos de segurança, art. 46 da Lei Geral de Proteção de Dados Pessoais (LGPD), não sofreram alterações entre os Ciclos 3 e 4. A autenticação com fator adicional ainda foi descrita como funcionalidade em desenvolvimento; permanecem pendentes o detalhamento das cópias de segurança e a definição do prazo de guarda das rotas salvas. Em relação à Ficha A2, referente à avaliação da maturidade tecnológica e ao mapeamento dos ETC, o quadro preenchido pela participante indica que a validação técnica dos algoritmos centrais permaneceu em TRL 4 no Ciclo 4. Quanto à pseudonimização e à segurança dos bancos de dados apartados, foi mantido o TRL 6 registrado para o Ciclo 3. Esses níveis correspondem à avaliação apresentada pela empresa na ficha e devem ser considerados em conjunto com as evidências produzidas nos testes.

No que tange às fichas A3 e A4, a análise regulatória e documental do Ciclo 4 não identificou alterações estruturais nos instrumentos de governança apresentados pela Synapse no ciclo anterior. Permanecem, portanto, as conclusões e recomendações decorrentes da análise desses documentos. A indicação de futuras funcionalidades de mapas geolocalizados e de autenticação de fator adicional suscitou novos questionamentos, mas sua implementação ainda depende de verificação. Na avaliação pelo *framework* geral registrada nas fichas, a pontuação passou de 68, no Ciclo 3, para 78 pontos, no Ciclo 4, sem que esse resultado, isoladamente, comprove a efetividade das salvaguardas declaradas. Destacam-se os seguintes aspectos:

A análise regulatória da equipe do CIAAM/USP continuou a examinar os instrumentos de governança, proteção de dados, direitos humanos, segurança da informação e inteligência artificial (IA) adotados no projeto, sem identificar, no Ciclo 4, mudança estrutural que afastasse as recomendações anteriores.

Permanecem como documentos centrais da avaliação o Relatório de Impacto à Proteção de Dados Pessoais (RIPD), a Avaliação de Impacto em Direitos Humanos (AIDH), o documento relativo à atuação do encarregado pelo tratamento de dados pessoais, as políticas de governança de IA e de dados, o plano de resposta a incidentes e os registros das operações de tratamento apresentados à ANPD. Sua entrega supriu parte das lacunas documentais identificadas nos ciclos anteriores, mas ainda é necessária a comprovação material das medidas neles descritas.

Os documentos relativos à arquitetura de privacidade e criptografia, ao inventário de API e terceiros, ao fluxo de dados ponta a ponta, à explicabilidade em camadas e aos testes de robustez e estresse integram a avaliação técnica. Seu estágio de comprovação não é uniforme: o fluxo de dados e o protocolo de explicabilidade foram descritos apenas parcialmente; o inventário de API e terceiros e o relatório de robustez e estresse permanecem insuficientemente comprovados; e segue pendente o teste contra falsificação de informações.

Nos Acordos de Cooperação Técnica (ACT) com entidades privadas, permanece a recomendação de delimitar com maior precisão o ambiente de operação, as entregas, os usos proibidos e a finalidade técnica, científica e regulatória da experimentação.

Os ACT também devem especificar os dados tratados, as respectivas finalidades e bases legais, os papéis dos agentes de tratamento, as responsabilidades das partes e as medidas técnicas e organizacionais de segurança.

Permanece a recomendação de incluir nos ACT regras verificáveis de governança de IA, supervisão humana, avaliação de impacto, revisão periódica de riscos, versionamento do modelo, rastreabilidade e validação técnica.

Nos ACT com entidades públicas, devem continuar expressos a finalidade pública da cooperação, sua relação com as competências do órgão, os deveres relativos à LGPD, a eventual incidência da Lei de Acesso à Informação (LAI) e o regime de recursos e custos.

A experiência do usuário continua a exigir aperfeiçoamentos de acessibilidade e transparência. As respostas apresentadas no Ciclo 4 à matriz de transparência esclarecem alguns aspectos do uso da solução, mas ainda não demonstram, de modo completo, quais informações são efetivamente exibidas aos diferentes grupos de usuários.

Permanece a necessidade de verificar alternativas aos elementos predominantemente visuais do aplicativo, incluindo descrições textuais, orientação por áudio, contraste adequado e compatibilidade com tecnologias assistivas. Na matriz de transparência, a empresa informou que usuários com deficiência visual não integram,

neste momento, o público atendido pela versão em desenvolvimento. Essa limitação de escopo deve ser comunicada com clareza e acompanhada de planejamento para sua revisão.

Recomenda-se manter a disponibilização de rotas acessíveis como opção para todos os usuários, com filtros funcionais relativos a escadas, elevadores, obstáculos e condições do trajeto, sem exigir autodeclaração de deficiência quando ela não for necessária à prestação da funcionalidade.

O tratamento de dados de localização e deslocamento continua a apresentar riscos de identificação indireta, exposição de rotinas e inferência de hábitos. A possível ampliação do uso de mapas geolocalizados requer esclarecimento sobre dados coletados, compartilhamento com terceiros, período de conservação e riscos adicionais para os titulares.

A interface deve distinguir o aceite dos termos e condições de uso, a autorização para uso da localização e o recebimento de comunicações. Permanecem relevantes mecanismos que permitam ao usuário controlar o compartilhamento de localização, encerrá-lo imediatamente e identificar quem tem acesso a essa informação. Devem ser apresentados os textos e as telas efetivamente utilizados.

A AIDH mantém a identificação de riscos relacionados à opacidade algorítmica e aos impactos sobre grupos vulneráveis. A elevação da pontuação no *framework* geral não dispensa evidências de transparência acessível, indicação de obstáculos nos trajetos e monitoramento contínuo da precisão das recomendações.

O RIPD permanece, em linhas gerais, compatível com a estrutura avaliada no ciclo anterior, mas deve incorporar expressamente os riscos de recomendações inadequadas de acessibilidade e esclarecer a hipótese legal aplicável ao eventual tratamento de informações que revelem deficiência de pessoa identificada ou identificável. Nessa hipótese, trata-se de dado pessoal sensível, sujeito às exigências do art. 11 da LGPD, não podendo seu tratamento ser fundamentado no legítimo interesse. A oferta de rotas acessíveis a todos os usuários deve ser considerada como medida de minimização da coleta.

Permanece a recomendação de assegurar ao encarregado pelo tratamento de dados pessoais autonomia técnica e condições para atuar sem conflito de interesses, conforme a Resolução CD/ANPD nº 18/2024. A adequação da estrutura institucional dessa função deverá ser verificada para além do ambiente experimental.

A política de governança de IA ainda precisa demonstrar como são executados, na prática, os procedimentos de comunicação com usuários, supervisão humana, intervenção, monitoramento e responsabilização. As declarações de governança devem ser acompanhadas de responsáveis, registros e evidências de sua aplicação.

Os *model cards* e os mecanismos de explicação devem informar uso pretendido, critérios principais de recomendação, métricas de desempenho, limitações e cenários de falha. Na matriz de transparência do Ciclo 4, a empresa especificou que a explicação da rota será apresentada no aplicativo a cada escolha; contudo, ainda faltam o conteúdo exato das mensagens, a definição completa dos momentos de exibição dos avisos e as evidências da interface implementada. Também deve ser esclarecida a informação prestada ao usuário diante de atualizações relevantes nas API e no tratamento de imagens.

A política de governança de dados deve demonstrar a implementação das salvaguardas declaradas, especialmente quanto à minimização, ao tratamento de informações relativas à deficiência, à transparência e à mitigação da identificação indireta

decorrente da combinação de localização, de destino e de atributos de acessibilidade. A afirmação de que a referida identificação não seria possível requer comprovação técnica.

O plano de resposta a incidentes permanece considerado suficiente em sua estrutura, mas ainda deve definir com precisão o conceito de "rota crítica" e os critérios empregados para sua identificação e priorização. A avaliação deve verificar, igualmente, a execução dos procedimentos previstos, e não apenas a sua descrição documental.

O relatório de viés e *fairness* continua a representar avanço no exame da distribuição espacial das recomendações, do monitoramento de *drift* e da supervisão humana. Esse documento, contudo, não substitui a comprovação de acessibilidade, transparência na interface, bases legais aplicáveis, controle dos dados compartilhados com terceiros e implementação das medidas de governança. As respostas acrescentadas à matriz no Ciclo 4 deverão ser confrontadas com a versão da solução submetida a testes.

Os documentos apresentados pela Synapse foram analisados nos aspectos técnicos e regulatórios pelas equipes específicas, sendo apontadas as seguintes evidências como pendentes:

- arquitetura de privacidade e criptografia;
- inventário de API e terceiros;
- mapa de fluxo de dados ponta a ponta;
- protocolo de explicabilidade em camadas;
- comprovação material do relatório e testes de robustez e estresse.

Quanto aos resultados da equipe técnica sintetizados nas fichas A3 e A4, o Ciclo 4 mantém o diagnóstico de avanço na arquitetura de proteção da solução Trajetto e de necessidade de comprovação de sua implementação prática, uma vez que ainda não há evidências que confirmem as declarações da Synapse. A síntese técnica menciona processamento de visão computacional na borda[9], transmissão de coordenadas e uso de protocolo de segurança na camada de transporte, mas ainda requer evidências sobre a não retenção de imagens, o inventário e os contratos relativos às API, bem como os avisos e explicações efetivamente apresentados no aplicativo. Também identifica risco de reidentificação pela combinação de geolocalização com atributos de acessibilidade. Há, ainda, uma divergência documental a esclarecer: a síntese técnica menciona autenticação com dois fatores, enquanto a Ficha A1 a descreve como funcionalidade em desenvolvimento. Assim, sua presença na versão testada deve ser confirmada antes de ser considerada controle implementado.

4.1.3 Fase B

A) *Framework* geral baseado em recomendações da UNESCO

No que tange à aplicação do *framework* geral, manteve-se no Ciclo 4 a metodologia empregada nos ciclos anteriores. Assim, a avaliação é processual e iterativa. No Ciclo 1, a própria organização realiza autoavaliação com base em suas evidências; nos ciclos seguintes, o especialista designado reexamina criticamente as respostas, confrontando-as com a documentação, as entrevistas e o plano de sandbox apresentado. O escore

[9] Processamento de visão computacional na borda — extração de coordenadas a partir do processamento de imagens e vídeos de câmeras feito na própria infraestrutura dos parceiros, sem trânsito de imagens para a infraestrutura do Trajetto.

agregado, limitado a 100 pontos, é convertido em quatro faixas de maturidade que orientam os encaminhamentos desde a continuidade plena no piloto (≥ 85) até a inaptidão para prosseguir no formato atual (< 50), passando por estágios intermediários sujeitos a planos de ação ou de mitigação. O resultado é um diagnóstico replicável e auditável, capaz de equilibrar o estímulo à inovação responsável e a proteção efetiva dos titulares de dados.

No Ciclo 4, a pontuação da Synapse passou de 68 para 78 pontos. O acréscimo de dez pontos decorreu dos seguintes elementos: identificação de grupos vulneráveis afetados, somando 4 pontos no Eixo 3 (Avaliação de impactos); indicação de documentação acessível, somando 4 pontos no Eixo 5 (Transparência, explicabilidade e comunicação); e avaliação de que o modelo não gera resultados perigosos fora de seu escopo, somando 2 pontos no Eixo 6 (Segurança técnica e robustez). O Eixo 4 (Justiça, não discriminação e vieses algorítmicos) permaneceu sem pontuação, uma vez que a documentação necessária e os ajustes na aplicação relativos à acessibilidade das rotas ainda não foram entregues. A identificação dos grupos vulneráveis ainda é preliminar, e a documentação apresentada requer ajustes. A elevação da pontuação, portanto, registra a evolução da solução Trajetto nesses critérios, mas não comprova, por si só, a efetividade das salvaguardas, a mitigação do risco de reidentificação indireta ou a acessibilidade da versão submetida a testes.

B) Testes da equipe técnica do CIAAM/USP

Os quatro testes da equipe técnica foram encerrados no Ciclo 3, mantendo-se as conclusões alcançadas nos limites e nas condições então documentados. Desse modo, não se atribuem ao Ciclo 4 novas conclusões técnicas decorrentes desses ensaios. No Ciclo 4, registrou-se apenas a verificação do protocolo de segurança na camada de transporte descrita no item D, feita no parecer técnico

C) Matriz de transparência

A matriz de transparência apresentada para o Ciclo 4 acrescentou respostas da participante para diferentes grupos de interlocutores. A Synapse informou que a explicação da rota aparecerá no aplicativo a cada escolha, esclareceu que o usuário geral visualiza a rota e a razão da recomendação, sem comparação de fluxos, e descreveu informações agregadas destinadas ao operador. Também detalhou, para a equipe técnica, perfis de acesso, registros e periodicidade de revisão. Essas respostas especificam parte dos meios de comunicação e dos controles internos, mas ainda não apresentam as telas e os textos exatos que serão exibidos aos usuários, nem definem completamente quando serão prestadas as informações sobre o uso de IA, os dados tratados e os critérios da recomendação. A participante declarou, ainda, que usuários com deficiência visual não integram o escopo da versão em desenvolvimento; essa limitação precisa constar de modo claro da descrição da solução e ser formalizada no acompanhamento do Sandbox.

D) Governança, segurança da informação e privacidade em IA

Em relação à governança, à segurança da informação e à privacidade, o parecer técnico consolidado do Ciclo 4 manteve a classificação “suficiente com condicionantes” e recomendou o prosseguimento dos testes acompanhado de solicitação de complementações. A avaliação, utilizando a metodologia descrita na ISO 27001, registra risco inerente agregado alto e risco residual agregado médio, considerando as medidas criptográficas examinadas e as pendências de comprovação dos demais controles. O

parecer reconhece a segregação lógica das bases, o emprego de criptografia autenticada com gerenciamento de chaves, além da verificação de uso de protocolo de segurança na camada de transporte no percurso auditado no Ciclo 4. Esses elementos não afastam a necessidade de avaliar o risco de identificação indireta e os fluxos que ficaram fora daquele teste.

Com o foco de sistematizar os indicadores de avaliação, a seguir apresenta-se o quadro atualizado a partir do parecer técnico consolidado do Ciclo 4:

Quadro 1: Classificação de Indicadores do sistema Trajetto

Indicador de Avaliação	***Classificação***
Conclusão Preliminar:	Suficiente com condicionantes.
Recomendação Principal:	Prosseguir com testes e solicitar complementação.
Nível Agregado de Risco Inerente:	Alto (devido ao tratamento contínuo de geolocalização e padrões de mobilidade).
Nível Agregado de Risco Residual:	Médio (devido aos controles criptográficos fortes empregados, mas impactado pela ausência de planos testados).
Mitigações Documentadas:	• Visão Computacional na borda: processamento de imagens e vídeos na infraestrutura dos parceiros, com recebimento exclusivo de coordenadas pelo Trajetto (não retenção de imagens pendente de comprovação).
Mitigações Planejadas Críticas:	Pendentes: • Mitigação do risco de reidentificação por atributos de acessibilidade: mecanismos para evitar que a combinação de coordenadas com a indicação de rota acessível permita a reidentificação do titular.

- Memorial descritivo de visão na borda: comprovação técnica de não retenção de imagens em *cache* local ou memória nas API de câmeras parceiras.

- Transparência material e mensagens no aplicativo (Art. 9º e 20 da LGPD): apresentação das telas, avisos e textos exatos a serem lidos pelo usuário no aplicativo.

- Formalização de *roadmap* para deficiência visual: registro formal no Sandbox Regulatório do cronograma de inclusão de usuários com deficiência visual, justificando sua exclusão temporária da versão em desenvolvimento.

Fonte: ANPD

No plano documental, o projeto Trajetto apresenta arquitetura orientada ao princípio de *privacy by design*, fundamentado na segregação de ambientes (cofres de identidade e telemetria), pseudonimização baseada em *hash*, com gerenciamento de chaves em provedor de serviços em nuvem, e criptografia em repouso. Na camada de transporte, a implementação de protocolo de segurança e a validação de resistência contra interceptação em testes de tunelamento reforçaram a proteção do canal de rede.

Contudo, a avaliação consolidada até o Ciclo 4 atesta que o projeto transitou de uma lacuna documental para uma lacuna probatória de implementação prática, restando atender aos seguintes elementos:

- Transparência material ao titular (Art. 9º e 20 da LGPD): a Matriz de Transparência atesta que a empresa entregou diretrizes conceituais, mas ainda não apresentou os protótipos de interface, os textos exatos das mensagens, os avisos (*banners*) e as explicações contextuais de IA a serem lidos pelo usuário final e grupos vulneráveis no aplicativo.
- Risco de reidentificação em dados de acessibilidade: identificou-se risco elevado no cruzamento de geolocalização dinâmica contínua com a indicação de rota acessível direcionada a locais específicos ou restritos com listas públicas de membros (exemplo: clubes ou academias), exigindo salvaguardas de agregação para evitar a reidentificação do titular.
- Memorial técnico e terceiros: persiste a necessidade de entrega de memorial descritivo atestando a não retenção de imagens em memória/*cache* nas API das câmeras parceiras, além do detalhamento

contratual auditado do inventário de dados compartilhados com o Google Maps e provedores de autenticação.

- Enquadramento normativo de alto risco: o tratamento de geolocalização contínua combinado com atributos de vulnerabilidade enquadra a solução como de alto risco, afastando a aplicação das flexibilizações para agentes de pequeno porte (Art. 4º da Resolução ANPD nº 02/2022) e sujeitando o projeto aos requisitos da Resolução CD/ANPD nº 15/2024.

Assim, a principal pendência da Fase B no Ciclo 4 é a comprovação material dos controles descritos. Em especial, deverão ser confrontados com a versão testada os avisos e explicações apresentados no aplicativo, a arquitetura de processamento das imagens, o compartilhamento de dados com terceiros e as medidas destinadas a evitar a reidentificação pela combinação de localização e atributos de acessibilidade. A recomendação de continuidade dos testes constante do parecer técnico está sujeita ao acompanhamento das condicionantes.

4.1.4 Fase C

A elaboração do relatório de êxitos integra a Fase C da metodologia desenvolvida pelo CIAAM/USP e validada pela Comissão de Sandbox da ANPD. Essa etapa possui função complementar aos pareceres técnicos: sistematiza os aprendizados produzidos durante a testagem e examina os resultados obtidos para a autoridade reguladora e para a empresa participante.

Para orientar essa avaliação no Ciclo 4, o relatório manteve os campos destinados à ANPD e à empresa. Em cada dimensão, o resultado observado deve ser relacionado à evidência correspondente, distinguindo-se as medidas implementadas, as declarações da participante e as providências ainda pendentes de comprovação.

A) Resultados do Ciclo 4 sob a perspectiva da ANPD

O acompanhamento permitiu à Agência aprofundar entendimentos aplicáveis a soluções de IA de mobilidade baseadas em geolocalização, tais como: i) qualificação, no estágio atual de desenvolvimento, da solução Trajetto como sistema de recomendação, e não como sistema de decisão exclusivamente automatizada com efeitos significativos, o que deslocaria, potencialmente, a incidência do art. 20 da LGPD para modelo de explicabilidade proporcional ao risco, estruturado em duas camadas (explicação acessível ao titular e explicação técnico-regulatória, sob sigilo, à Agência); ii) a pseudonimização baseada em hash, com gerenciamento de chaves em provedor de serviços em nuvem, e criptografia em repouso não equivale à anonimização e coordenadas associadas a uma sessão de uso permanecem dados pessoais pseudonimizados; iii) a combinação de geolocalização contínua com atributos de acessibilidade constitui vetor de reidentificação indireta e caracterizaria, potencialmente, o tratamento como de alto risco, afastando as flexibilizações da Resolução CD/ANPD nº 2/2022, nos termos de seu art. 4º, além de sujeitar o projeto ao cumprimento integral dos requisitos de segurança da informação da Resolução CD/ANPD nº 15/2024; e iv) em agentes inovadores de pequeno porte, a lacuna

predominante tende a migrar de documental para probatória, isto é, da ausência de instrumentos de governança para a ausência de evidência material de sua implementação.

O ambiente experimental induziu a adoção ou a formalização de salvaguardas em benefício dos titulares, como: i) transferência do processamento de imagens de câmeras para a infraestrutura dos parceiros, com recebimento exclusivo de coordenadas pelo Trajetto, o que mitiga o risco de tratamento de dados biométricos; ii) entrega do RIPD, da AIDH, do plano de resposta a incidentes, do relatório de viés e fairness e do relatório de testes de robustez e estresse, inexistentes no início da testagem; iii) implementação de protocolo de segurança na camada de transporte; iv) esclarecimento de que a condição de acessibilidade é tratada como atributo da rota (rota acessível), dissociado do cadastro e vinculado a sessão renovada a cada acesso; e v) recomendações de interface encaminhadas à empresa: separação de consentimentos por finalidade, painel de privacidade e oferta de rotas acessíveis a todos os usuários, sem exigência de autodeclaração de deficiência.

Ressalva-se, contudo, que parte dessas medidas ainda depende de comprovação material e validação prática, especialmente quanto aos testes contra falsificação de informações, ao inventário de API e terceiros, à transparência material no aplicativo, ao memorial técnico de visão computacional na borda e à mitigação do risco de reidentificação por atributos de acessibilidade.

O caso fornece subsídios preliminares para futura atuação normativa, orientativa e supervisória da Agência, em especial: i) regulamentação do art. 20 da LGPD e orientações sobre transparência algorítmica: a Matriz de Transparência por interlocutor (por quê, o quê, quando e como) mostrou-se instrumento apto a distinguir transparência declarada de transparência efetiva; ii) orientações sobre anonimização e pseudonimização de dados de geolocalização e de mobilidade; iii) tratamento de dados relativos à deficiência em serviços de acessibilidade, com primazia de soluções por padrão (rotas acessíveis ofertadas a todos) sobre a coleta de dado pessoal sensível mediante consentimento, à luz dos arts. 6º, inciso III, e 11 da LGPD; e iv) critérios de caracterização de tratamento de alto risco em soluções que combinem geolocalização contínua, padrões de mobilidade e atributos de vulnerabilidade.

Os efeitos no mercado ainda são prospectivos, uma vez que a solução permanece em estágio de desenvolvimento e a funcionalidade de mapas geolocalizados ainda não foi implementada no Ciclo 4. Identificam-se, contudo, efeitos indiretos relevantes: i) constituição de referencial de documentação de governança — RIPD, AIDH, políticas de IA e de dados, plano de incidentes e relatório de fairness — replicável por outros agentes do segmento de mobilidade; ii) demonstração de que agente inovador de pequeno porte, vinculado a instituição de ensino, pode adotar arquitetura robusta de privacidade desde a concepção, desde que acompanhada de evidências materiais suficientes; iii) redução de incerteza jurídica para parcerias público-privadas no setor, por meio do mapeamento normativo e de recomendações de cláusulas para ACT com entes públicos e privados; e iv) preservação da isonomia concorrencial, na medida em que o Sandbox não deve operar como mecanismo de flexibilização de exigências.

B) Resultados do Ciclo 4 sob a perspectiva da Synapse

Sob a perspectiva da empresa Synapse, o Ciclo 4 resultou em melhorias na usabilidade e na experiência do usuário da solução Trajetto. Nesse sentido, a participante informou que trabalhou no aprimoramento da interface e que continuou o

aperfeiçoamento da comunicação no site, no âmbito das medidas de transparência e explicabilidade.

Já no campo regulatório, a empresa informou que os aprendizados permaneceram conforme a atualização anterior. Da mesma forma, não foram relatadas mudanças em relação aos direitos dos titulares no decorrer do Ciclo 4.

A partir das informações coletadas, verifica-se que o Ciclo 4 apresentou avanços na maturidade regulatória e técnica do Trajetto, especialmente na consolidação da arquitetura de privacy by design, na verificação dos controles de segurança, na evolução da pontuação no *framework* geral e na identificação dos riscos associados à geolocalização, à acessibilidade e à reidentificação indireta. Por outro lado, cumpre ressaltar que evidências documentais e registros das alterações declaradas não foram entregues ou apresentados na Ficha C1 pela empresa. O êxito do ciclo, contudo, é parcial, pois persistem lacunas probatórias quanto à transparência material ao titular, ao inventário de API e terceiros, aos testes contra falsificação de informações, ao memorial técnico de visão computacional, à mitigação do risco de reidentificação por atributos de acessibilidade e à comprovação operacional das medidas de descontinuidade.

Nesse passo, constata-se que a evolução do projeto deverá ser orientada pela apresentação e validação progressiva dessas evidências. Devem ser priorizados a comprovação visual e funcional das mensagens de transparência no aplicativo; o detalhamento contratual e técnico das integrações com Google Maps, provedores de autenticação e demais API de terceiros; os testes contra falsificação de informações, de robustez e de estresse; o memorial descritivo do processamento de imagens na borda e de sua não retenção em cache ou memória; o protocolo de mitigação do risco de reidentificação por atributos de acessibilidade; a atualização do RIPD para contemplar os riscos de acessibilidade; e a atualização do Plano de Descontinuidade, com testes de acionamento, procedimentos de comunicação com provedores, matriz de riscos residuais e evidências por ciclo.

Finalmente, o Plano de Descontinuidade registra, para o Ciclo 3, a segregação das informações no banco de dados, o mapeamento das API de terceiros e a dificuldade de formalizar cláusulas contratuais com alguns fornecedores. Para o Ciclo 4, acrescenta a declaração de que foi realizado um teste técnico controlado de suspensão, bloqueio ou eliminação em ambiente delimitado. No entanto, os testes técnicos declarados pela empresa permanecem pendentes quanto à comprovação e entrega dos respectivos documentos que evidenciam as medidas citadas. Assim, o documento registra a realização declarada do teste, mas não permite confirmar sua conclusão nem seus resultados, permanecendo necessária a apresentação dos registros de execução e a atualização dos quadros de acompanhamento do plano.

4.2 Prevvine

O relatório do CIAAM/USP referente ao Ciclo 4 de testagem da solução STAiDOC, apresentada pela Prevvine, consolida as informações, os registros e as evidências das Fases A e B, bem como a avaliação da ANPD disponível para a Fase C. A análise abrange a caracterização técnico-funcional, a maturidade tecnológica, os testes realizados, os riscos, as medidas de mitigação e a governança. A seguir, apresentam-se os principais dados, informações e aprendizados do Ciclo 4, considerando os seis pilares da metodologia de avaliação:

- Dimensão técnico-funcional: análise das fichas A1 e A2, que descrevem a solução, sua arquitetura, os componentes tecnológicos, os fluxos operacionais e as características do modelo de linguagem utilizado;
- Dados, privacidade e segurança: análise das informações e evidências registradas nas subfases A3, A4, B1 e B2, incluindo o tratamento de dados pessoais, as medidas de segurança, os controles de acesso, a transparência e a efetividade das salvaguardas declaradas;
- Maturidade tecnológica (TRL e ETC): consideração da autoavaliação apresentada pela Prevvine nas fichas A1 e A2 e da análise da equipe do CIAAM/USP nas fichas B1 e B2, à luz das evidências e dos resultados da testagem. A manutenção dos registros anteriores não permite presumir evolução dos níveis de maturidade no Ciclo 4;
- Realismo experimental e benefícios esperados: avaliação dos testes e das evidências da Fase B, especialmente das interações clínicas examinadas na plataforma HarpIA, considerando a qualidade das respostas, a fundamentação bibliográfica, as limitações observadas e a correspondência com os usos e benefícios esperados no contexto da saúde;
- Riscos, mitigação e impacto residual: análise das avaliações da subfase B2, contemplando os riscos identificados, sua probabilidade e impacto, as medidas de mitigação, as condicionantes e as pendências de comprovação. A classificação agregada do risco residual deve ser considerada em conjunto com os riscos individuais e as evidências disponíveis;
- Governança, transparência e aprendizagem regulatória: análise da documentação institucional, dos papéis e responsabilidades, da supervisão humana, da explicabilidade e da matriz de transparência, especialmente na subfase B1. Esses elementos permitem identificar avanços e lacunas, distinguindo os resultados clínicos da comprovação dos controles técnicos e regulatórios.

Considerando o acompanhamento do Ciclo 4 em suas três fases: (i) Fase A (pré-teste); (ii) Fase B (teste); e (iii) Fase C (pós-teste), são apresentados a seguir os principais aspectos identificados em cada uma dessas etapas. Contudo, na Fase C, a consolidação considera a ficha preenchida pela ANPD, permanecendo pendentes a manifestação da Prevvine e, em consequência, a conclusão conjunta.

4.2.1. Documentação Principal

Preliminarmente, observa-se que os principais documentos referidos neste item do relatório encontram-se organizados conforme a descrição a seguir, por fases, para atendimento dos requisitos pertinentes ao Ciclo 4 deste Sandbox Regulatório:

- Fase A:
    - Fichas A1 e A2; e

    - Fichas A3 e A4.
- Fase B
    - Ficha B1 — Framework geral;
    - Ficha B2 — Classificação de riscos;
    - Planilha de Testes com 30 Vinhetas;
    - Matriz de Transparência; e
    - Parecer Técnico Consolidado.
- Fase C
    - Ficha C1 — Êxitos para a ANPD; e
    - Plano de descontinuidade.

4.2.2 Fase A

Inicialmente, no que se refere à Ficha A1, ao se analisarem os itens relativos à arquitetura do modelo, ao fluxo de dados, às API, à inteligência artificial e aos protocolos de segurança, a participante Prevvine informou que não houve alterações entre os Ciclos 3 e 4 de testagem. Essa declaração registra a continuidade das características descritas pela empresa, sem representar, isoladamente, comprovação da implementação ou da efetividade dos controles apresentados.

No que tange à Ficha A2, permaneceram identificados quatro ETC:

- motor de IA;
- camada de supressão de identificadores por expressões regulares (REGEX), denominada pela participante “camada de anonimização”;
- sistema NTAV, responsável pela verificação bibliográfica no PubMed; e
- infraestrutura de conformidade LGPD.

Quanto à classificação de maturidade tecnológica, a matriz da Ficha A2 não apresenta preenchimento na coluna correspondente ao Ciclo 4. Assim, os últimos níveis registrados são os do Ciclo 3: TRL 7 para o motor de IA, a camada de supressão de identificadores e a infraestrutura de conformidade, e TRL 6 para o sistema NTAV, cuja classificação foi reduzida em razão dos resultados dos testes realizados na plataforma HarpIA. Não há, portanto, registro suficiente nessa matriz para afirmar nova classificação ou recuperação da maturidade do NTAV no Ciclo 4. Ademais, o detalhamento individual desse componente ainda menciona TRL 7, divergência que deverá ser harmonizada com o quadro consolidado e com as observações da própria ficha.

Segundo o registro da Ficha A2, a redução da classificação do NTAV ao final do Ciclo 3 decorreu da identificação de alucinações, mistura de dados reais com informações inventadas, *links* quebrados e pesquisas realizadas pelo modelo sem resultados úteis. Também foi apontada a ausência de informação explícita ao usuário sobre possíveis falhas, incluindo a geração de referências inexistentes. Esses achados permanecem como antecedentes da avaliação do Ciclo 4, sem que a declaração de ausência de alterações permita concluir que as limitações foram superadas.

Em relação à camada de supressão de identificadores (denominada pela participante “camada de anonimização”), a ficha descreve a supressão de identificadores por REGEX, o descarte do texto original e a manutenção de *hashes*. Entretanto, o descarte do original não comprova, por si só, a impossibilidade de identificação indireta a partir do conteúdo clínico remanescente. Dessa forma, a técnica deve ser descrita como supressão

de identificadores, permanecendo pendente a comprovação de sua eficácia para caracterizar anonimização nos termos dos arts. 5º e 12 da LGPD. Tampouco se deve concluir automaticamente pela pseudonimização sem verificar a existência de informação adicional que permita restabelecer a associação ao titular.

Diante do exposto, embora a participante não tenha declarado mudanças nos componentes descritos na Ficha A1 entre os Ciclos 3 e 4, permanecem necessárias a atualização e a harmonização da matriz de maturidade tecnológica, bem como a comprovação da efetividade das medidas de desidentificação. A avaliação do Ciclo 4 deve, portanto, confrontar as declarações da participante com as evidências da testagem, distinguindo os níveis anteriormente registrados, as medidas descritas e os resultados efetivamente verificados.

Já em relação às Fichas A3 e A4 da Prevvine, procedeu-se à comparação das informações apuradas no Ciclo 3 com os registros apresentados para o Ciclo 4. A análise abrangeu o marco regulatório aplicável, os controles internos, os processos operacionais e a arquitetura tecnológica da solução de inteligência artificial em saúde, distinguindo os achados registrados, as atividades previstas e as pendências de comprovação.

Para avaliar a incidência normativa, foram mantidos três aspectos:

- a relação da norma com o setor de saúde, com a ferramenta (um *software* conversacional que trata dados e produz orientações) e com o contexto clínico de uso, especialmente quando suas recomendações podem influenciar decisões assistenciais;
- as normas transversais de proteção de dados, segurança da informação e defesa do consumidor; e
- as regras aplicáveis a atividades específicas da área da saúde.

No âmbito da proteção de dados, a análise manteve como referências a LGPD e a Resolução CD/ANPD nº 4/2023, sobre dosimetria das sanções, e aprofundou a relação entre as falhas identificadas e o Regulamento de Comunicação de Incidente de Segurança, aprovado pela Resolução CD/ANPD nº 15/2024. A avaliação concentrou-se na proteção dos dados de saúde, na qualidade das informações produzidas e na comprovação das salvaguardas. Como os ensaios descritos utilizaram dados sintéticos, seus resultados constituem evidências de vulnerabilidades e riscos, sem demonstrar, por si só, incidente envolvendo dados pessoais reais. Eventual dever de comunicação em ambiente de produção depende da verificação dos requisitos legais e regulamentares.

Quanto à saúde digital, foram mantidas as seguintes referências:

- Lei nº 14.510/2022, que disciplina a telessaúde;
- Lei nº 12.842/2013, relativa ao exercício da medicina; e
- normas profissionais aplicáveis aos médicos e, quando pertinente, à enfermagem.

Nesse sentido, a análise reafirmou que a solução deve permanecer delimitada como ferramenta de apoio, sem substituir o julgamento clínico do profissional. As normas relativas à telenfermagem e ao exercício da enfermagem continuam com incidência condicionada à utilização nesse contexto. O Ciclo 4 também reforçou a necessidade de verificar a não discriminação por meio de testes, examinando se mudanças no perfil do paciente alteram recomendações sem justificativa clínica.

Também foi mantida a análise do possível enquadramento da solução como *Software as a Medical Device* (SaMD). A RDC ANVISA[10] nº 657/2022 deve ser considerada em conjunto com a RDC nº 751/2022, relativa à classificação de risco e aos regimes de regularização, e com a RDC nº 848/2024, sobre requisitos essenciais de segurança e desempenho. A necessidade de regularização e a classe de risco dependem da finalidade pretendida e das funcionalidades efetivamente oferecidas. A classificação sanitária pode subsidiar a interlocução entre autoridades, preservadas as diferenças entre os riscos à saúde e os riscos aos direitos dos titulares.

No exame dos documentos e das políticas internas, o Ciclo 4 registrou a previsão de análise comparativa das versões das políticas de privacidade de consumidor e do *System Card* do *Large Language Model* (LLM) utilizado. Essas atividades buscam examinar alterações nas condições de tratamento e atualizar os instrumentos de teste para o modelo efetivamente utilizado. Sua aplicação à STAiDOC exige distinguir as condições dos serviços de consumo daquelas pertinentes ao uso via API. Também foram solicitados o mapeamento de dados, o RIPD, a avaliação de impacto em direitos humanos, as políticas de privacidade e de retenção e descarte, o registro de atividades de tratamento, o inventário de sistemas e ativos de IA e o plano de resposta a incidentes. A ficha não registra a conclusão dessas análises nem permite considerar encerradas as pendências de implementação, inclusive as relativas aos avisos sobre limites de uso e situações de emergência, tendo sido entregues no curso do ciclo o RIPD, a avaliação de impacto em direitos humanos e o plano de resposta a incidentes

Além disso, a análise dos dados e dos fluxos de tratamento registrou a manutenção do fluxo funcional e do escopo exclusivamente textual da testagem. Foi descrito um conjunto de sessenta vinhetas sintéticas para avaliar o filtro de privacidade, com identificadores diretos, informações sensíveis, atributos e vínculos assistenciais, incluindo erros de grafia e valores implausíveis. Na análise de dez conversas, foram registradas resistência às solicitações diretas de extração e exposição de nomes fictícios de filiação materna em tentativas de indução à correção. Esses resultados limitam a afirmação de que todos os tratamentos são anonimizados. As respostas do modelo sobre acesso ao histórico, retenção e descarte apresentaram inconsistências e precisam ser confrontadas com evidências da arquitetura, pois sua autodescrição não comprova o funcionamento desses controles.

Paralelamente, a definição das bases legais e dos papéis dos agentes de tratamento permaneceu dependente da consolidação do mapeamento. O hospital pode atuar como controlador quando determina as finalidades e os elementos essenciais do tratamento, enquanto a fornecedora pode atuar como operadora quando segue suas instruções. A eventual controladoria conjunta exige verificar as decisões efetivamente compartilhadas, tema indicado para reunião específica. Quanto aos dados pessoais sensíveis de saúde, devem ser examinadas as hipóteses do art. 11 da LGPD, sem estender a eles eventual legítimo interesse considerado para metadados bibliográficos.

Ademais, a análise aprofundou a distinção entre a transferência de dados do Brasil ao fornecedor estrangeiro e a eventual incidência de legislação estrangeira sobre dados de autores citados. A nacionalidade do autor não determina, isoladamente, essa incidência: devem ser considerados os elementos de conexão pertinentes, como estabelecimento dos agentes, localização dos titulares e mercados atendidos. Quanto aos fluxos internacionais da solução, permanecem necessárias a identificação dos dados efetivamente enviados, a avaliação da possibilidade de identificação e a demonstração do mecanismo aplicável nos termos do art. 33 da LGPD e da Resolução CD/ANPD nº 19/2024.

[10] ANVISA — Agência Nacional de Vigilância Sanitária

Dados de terceiros identificáveis presentes em artigos, especialmente informações de saúde, demandam análise própria.

Em seguida, no campo da transparência e da conformidade, a avaliação aprofundou as falhas de referências bibliográficas, incluindo artigos inexistentes, atribuições incorretas, nomes truncados e filiações deslocadas. Foram examinadas hipóteses de falha de exatidão e de integridade, cuja distinção depende de esclarecer se a referência foi gerada pelo modelo ou recuperada e alterada durante o processamento. Permanecem relevantes a explicação dos critérios de seleção das fontes, a identificação das referências efetivamente consultadas e a informação ao usuário sobre limitações. A divergência entre controles declarados e comportamento observado reforça a necessidade de comprovação técnica das medidas de transparência e proteção de dados.

Quanto à coordenação regulatória, a ficha não registra atualização substancial da lista de normas setoriais, mas apresenta avanços na formulação de instrumentos de cooperação. Foram elaboradas propostas de cláusulas para eventual acordo entre a ANPD e a ANVISA, tomando como referência o acordo com a SENACON. As propostas abrangem intercâmbio de informações, indicadores conjuntos, coordenação no tratamento de incidentes e reclamações, capacitação e fiscalização, preservadas as competências de cada autoridade. Também foram sugeridas aproximações entre os registros de tecnovigilância e as comunicações de incidentes, com informações específicas sobre o sistema de IA e a reprodutibilidade das falhas. Essas propostas não equivalem à celebração do acordo ou à realização das reuniões previstas.

Em suma, as Fichas A3 e A4 do Ciclo 4 mostram que a avaliação regulatória da solução STAiDOC avançou no detalhamento dos seguintes pontos:

- Riscos de extração, uso e identificação de informações no contexto conversacional;
- Qualidade das referências bibliográficas e limites das salvaguardas declaradas; e
- Instrumentos possíveis de coordenação entre a proteção de dados e a regulação sanitária.

Contudo, permaneceram pendentes:

- A consolidação do mapeamento de dados e a confirmação técnica dos fluxos, da retenção e do descarte;
- A definição fundamentada das bases legais e dos papéis dos agentes de tratamento;
- A conclusão das análises documentais e a comprovação da implementação das políticas e avaliações de impacto;
- A demonstração dos limites de uso e da efetividade dos controles de privacidade, segurança e não discriminação; e
- A comprovação das medidas de transparência, governança e regularidade dos fluxos internacionais.

Diante da análise, recomendou-se à ANPD:

- Promover interlocução com a ANVISA e, conforme o contexto de utilização, com a ANS[11] e os conselhos profissionais, considerando as propostas de cooperação apresentadas.
- Manter um mapeamento integrado das normas aplicáveis, distinguindo regras setoriais, transversais e condicionadas ao uso da solução.
- Solicitar a complementação do mapeamento de dados e a comprovação técnica dos mecanismos de supressão, retenção e descarte, com identificação dos agentes e discussão da eventual controladoria conjunta.
- Verificar, à luz das funcionalidades e da finalidade pretendida, o eventual enquadramento como SaMD e a classe de risco correspondente.
- Verificar a implementação de avisos claros sobre emergências, limitações e necessidade de validação profissional das recomendações e referências.
- Solicitar evidências das políticas, avaliações de impacto, mitigações e mecanismos de transparência, distinguindo os resultados clínicos do cumprimento das condicionantes.
- Avaliar os mecanismos aplicáveis às transferências internacionais e esclarecer a origem e o processamento das referências bibliográficas, para fundamentar o tratamento das falhas de qualidade e de segurança.

4.2.3 Fase B

Inicialmente, a Fase B do Ciclo 4 reuniu a atualização do *framework* geral, a execução das avaliações conversacionais com vinhetas clínicas na plataforma HarpIA, o acompanhamento da explicabilidade e da transparência, a atualização da matriz de transparência e a consolidação dos riscos e das medidas de mitigação em parecer técnico. Os instrumentos permitiram distinguir os avanços documentais e operacionais, os resultados observados pelos especialistas e as condicionantes ainda dependentes de implementação ou comprovação.

A) *Framework* geral baseado em recomendações da UNESCO

A Prevvine obteve 75 pontos de 100 no Ciclo 4, ante 63 pontos no Ciclo 3. O acréscimo de doze pontos ocorreu exclusivamente no Eixo 3 (Avaliação de impactos), que passou de três pontos, referentes ao plano de mitigação de riscos, para quinze pontos, pelo atendimento registrado dos critérios relativos ao Relatório de Impacto à Proteção de Dados, à Avaliação de Impacto em Direitos Humanos e à identificação de grupos vulneráveis afetados. As pontuações dos demais eixos permaneceram iguais: Eixo 1 (Centralidade da dignidade humana e dos direitos fundamentais), 20 pontos; Eixo 2 (Governança e responsabilidade clara), 7 pontos; Eixo 4 (Justiça, não discriminação e vieses algorítmicos), 7 pontos; Eixo 5 (Transparência, explicabilidade e comunicação), 11

[11] ANS — Agência Nacional de Saúde Suplementar

pontos; Eixo 6 (Segurança técnica e robustez), 5 pontos; e Eixo 7 (Proporcionalidade e necessidade), 10 pontos.

A pontuação do Ciclo 4 corresponde à faixa “Adequado, mas com pontos críticos a melhorar”. O avanço registra o atendimento dos critérios documentais de avaliação de impactos, sem comprovar, isoladamente, a efetividade das salvaguardas ou a superação das falhas observadas nos testes. Permaneceram lacunas relativas a:

- governança e definição de responsabilidades;
- avaliação e correção de vieses;
- acessibilidade da documentação e comunicação com os destinatários; e
- robustez técnica e segurança das respostas.

B) Avaliação conversacional do modelo

Nesta etapa, foi utilizado outro conjunto vinhetas que organizou os cenários clínicos e suas variações para o Ciclo 4, mantendo o acompanhamento das características demográficas e contextuais. O instrumento possui um total de trinta vinhetas que contemplam:

- idade;
- gestação e saúde materna;
- deficiência;
- gênero e identidade de gênero;
- raça e etnia;
- renda;
- orientação sexual;
- região do Brasil; e
- pertencimento a populações prioritárias do Sistema Único de Saúde (SUS).

A introdução dessas variações permite examinar a consistência das respostas e possíveis diferenças entre perfis de pacientes. A planilha, contudo, constitui instrumento de organização e cobertura dos casos; os resultados da execução devem ser examinados nos registros da HarpIA e nas avaliações dos especialistas.

O impedimento operacional registrado no ciclo anterior foi resolvido e o relatório da HarpIA apresenta avaliações efetivamente realizadas no Ciclo 4 por dois especialistas médicos, com quinze conversas válidas por avaliador. Uma conversa adicional do segundo avaliador, identificada como repetição equivocada de vinheta, foi excluída do processamento da avaliação do sistema. Os registros permitem examinar utilidade, compreensibilidade, precisão, satisfatoriedade, naturalidade, adequação dos exames e condutas, referências bibliográficas e atenção às particularidades dos casos.

Os resultados combinam percepções favoráveis sobre a compreensibilidade e a naturalidade com limitações de precisão e fundamentação. Entre os achados específicos do Ciclo 4, o segundo especialista atribuiu nota mínima de precisão a uma conversa que envolveu imprecisão temporal, hipótese diagnóstica considerada inadequada, referência incompatível com o quadro e recomendação de urgência considerada factualmente incorreta. Em outro caso, registrou discordância quanto à sugestão de exame especializado sem justificativa no contexto apresentado. Também foram apontadas

alucinações bibliográficas e referências desatualizadas. Esses registros constituem avaliações dos especialistas sobre as conversas examinadas e não autorizam generalizações automáticas sobre todas as respostas da solução.

Observa-se, ainda, que o relatório da HarpIA inclui uma meta-avaliação da qualidade metodológica das conversas. No Ciclo 4, essa análise considerou quinze conversas do primeiro especialista e dezesseis do segundo, preservando a interação adicional para avaliação do procedimento. Foram identificadas diferenças de profundidade e suficiência das interações. Como os especialistas avaliaram casos distintos, suas notas não constituem medida direta de concordância. A meta-avaliação tampouco verifica a correção clínica dos julgamentos, mas delimita a força das evidências e recomenda maior padronização, aprofundamento das conversas e preparação dos avaliadores.

C) Explicabilidade para transparência

No Ciclo 4, foram registrados avanços na documentação de governança, na elaboração do Manual do Médico, cuja disponibilização na área logada estava em andamento, e nas informações destinadas aos usuários. Também se reconheceu a introdução de respostas institucionais fixas no servidor para perguntas sobre a plataforma e o tratamento de dados, destinadas a reduzir respostas imprecisas do modelo sobre seu próprio funcionamento. Essas medidas representam evolução em relação às lacunas anteriores, mas a condicionante de explicabilidade permaneceu classificada como “em evolução”.

Na comunicação com médicos e demais usuários, a matriz registra a declaração da participante de que a interface apresenta aviso de suporte à decisão, alertas em cenários sugestivos de urgência, aviso de privacidade diante da detecção de dados pessoais e mecanismo de confirmação ou rejeição de sugestões. Entretanto, a mesma matriz informa que o manual estava sendo implementado na área logada e que as políticas ainda estavam em ajuste de linguagem. Esses registros devem ser distinguidos da conclusão mais abrangente do parecer sobre sua disponibilização, mediante verificação da versão efetivamente acessível aos destinatários.

Por fim, os materiais examinados não apresentam nova classificação item a item das três camadas de explicabilidade para o Ciclo 4. Assim, os avanços registrados não permitem afirmar aprovação integral do instrumento. Permanecem necessárias a consolidação dessa avaliação e a demonstração, na interface e na documentação, dos limites de uso, das fontes, das incertezas, dos canais de intervenção humana e das informações adequadas a cada destinatário.

D) Matriz de transparência

A matriz apresenta, no Ciclo 4, registros para os cinco grupos de interlocutores: (i) usuários, (ii) médicos; (iii) hospitais e clínicas; (iv) ANPD e demais reguladores; e (v) equipes de tecnologia da informação e segurança. Foram identificadas alterações relativas ao ajuste das políticas para linguagem simples, à elaboração e implementação do manual médico e ao recebimento do plano de resposta a incidentes. O preenchimento integral dos campos não corresponde à completude documental, nem equivale à comprovação de conformidade. As fórmulas contabilizam vinte campos de conteúdo por ciclo, embora o rótulo da planilha e o parecer mencionem 24 itens, divergência que deverá ser harmonizada.

Para os usuários, permaneceram pendentes a definição do fluxo de revisão de decisões automatizadas, do momento de sua apresentação e da uniformização das informações entre canais. O termo de consentimento foi apresentado, mas a matriz manteve a necessidade de avaliar a adequação da base legal. Para os médicos, continuou indicada a revisão das avaliações sobre o uso da ferramenta, acrescida da declaração de elaboração do manual e dos avisos incorporados à interface.

Quanto aos hospitais e clínicas, a matriz registrou o envio do plano de resposta a incidentes, considerado satisfatório na avaliação apresentada, e manteve as necessidades de informação sobre arquitetura, segurança, políticas e métricas. Para a ANPD e os demais reguladores, continuaram previstos relatórios, registros, métricas de viés e demonstrações das mitigações, permanecendo necessário detalhar como essas evidências serão apresentadas. Para a equipe de tecnologia e segurança, manteve-se a necessidade de fixar um ciclo interno de revisão, apoiado por métricas, registros, alertas e documentação das integrações.

E) Parecer técnico: mapa de riscos e mitigações

O parecer técnico consolidado do Ciclo 4, elaborado com referências às normas ISO/IEC 27001 e 27002, manteve a classificação de risco inerente agregado crítico e de risco residual agregado médio, com conclusão favorável à continuidade dos testes sob condicionantes. A justificativa apresentada associa o risco residual aos avanços de governança e à execução sequencial do filtro de privacidade antes da inferência. Contudo, a matriz individual continua com pontuações na faixa alta e conserva valores anteriores para riscos descritos como resolvidos. Adicionalmente, cumpre realizar dois esclarecimentos que são relevantes para o resultado acima exposto:

- Regra de agregação sistêmica: o risco residual geral (médio) não resulta de uma média aritmética simples, nem é ditado isoladamente pelo maior risco individual. Trata-se da probabilidade combinada de falha do sistema como um todo. Como a arquitetura do STAiDOC opera em defesa em profundidade — com controles determinísticos na borda, execução em modo *fail-closed* e descarte de texto bruto antes de qualquer envio ao modelo de IA —, a exposição sistêmica global reduz-se para o nível médio.
- Ponderação de impacto e atualização dos riscos resolvidos: os riscos isolados foram mantidos na faixa 'Alta' (como a dependência da API do PubMed), pela gravidade de seu impacto, embora afetem apenas funcionalidades específicas e acionem gatilhos de pausa automática, impedindo que uma indisponibilidade isolada se transforme em um desastre de vazamento de dados.
- Riscos resolvidos foram marcados como 'Solucionados' ou 'Corrigidos por Código', ficando na faixa ‘Médio’, de acordo com a regra de agregação sistêmica e a tendência de risco após os encaminhamentos condicionantes que reduziram a exposição real do sistema.

Essa classificação deve, portanto, ser apresentada como conclusão do avaliador, sujeita ao esclarecimento de sua consolidação, sem pressupor redução uniforme dos riscos.

Neste Ciclo 4, o registro manteve seis riscos, com os seguintes estados:

- reidentificação de titulares: pendente, com classificação alta;
- indisponibilidade do PubMed e seus efeitos sobre a operação: pendente, com classificação alta;
- ataques à aplicação e *prompt injection*: pendente, com classificação alta;
- bloqueio da integração com a HarpIA: descrito como resolvido;
- alucinações relacionadas à autodescrição do sistema e às associações indevidas examinadas no parecer: descritas como corrigidas por código; e
- deficiências de transparência e explicabilidade: pendentes, com classificação alta.

Por um lado, o parecer reconheceu a execução do filtro de supressão de identificadores antes do envio ao modelo, o bloqueio preventivo em caso de falha, os registros de auditoria, a proteção criptográfica em trânsito e em repouso e a validação bibliográfica pelo NTAV. Também registrou a introdução de bloqueio determinístico de instruções adversariais e de respostas institucionais fixas. Por outro lado, manteve limitações relativas ao método criptográfico, à detecção de nomes fora dos padrões previstos, à comprovação do *Web Application Firewall* (WAF) e à contingência para indisponibilidade do PubMed. A verificação de identificadores bibliográficos existentes não comprova, por si só, a pertinência, a atualização ou a correta atribuição das referências, aspectos nos quais a HarpIA registrou falhas.

Também foram registradas como concluídas as condicionantes referentes à estabilização da integração com a HarpIA; à entrega e validação do conjunto documental de governança; e aos cronogramas e prazos dos avaliadores. Permaneceram em evolução ou andamento as condicionantes relacionadas à explicabilidade, ao método criptográfico, à comprovação do WAF, à contingência do PubMed e ao aperfeiçoamento da supressão por REGEX. O parecer caracteriza o mecanismo examinado como pseudonimização, diante das possibilidades de reidentificação que identifica. A futura implementação do método criptográfico deverá ser avaliada com os demais controles, sem ser tomada isoladamente como comprovação de anonimização.

Em conjunto, os resultados da Fase B demonstram avanço documental na avaliação de impactos, estabilização operacional da testagem e incorporação de medidas corretivas, acompanhados de falhas clínicas e bibliográficas e de condicionantes ainda abertas. A continuidade experimental foi considerada possível sob acompanhamento. Dessa forma, os resultados das vinhetas devem ser examinados separadamente da comprovação dos controles de privacidade, segurança, continuidade operacional, transparência e governança. Ademais, observa-se que, sobre o plano de descontinuidade, não houve alteração entre os ciclos 3 e 4.

4.2.4 Fase C

A Fase C consolidou os resultados e os aprendizados do Ciclo 4, mantendo a estrutura destinada às duas perspectivas complementares: os êxitos para a ANPD e os êxitos para a Prevvine. Até o encerramento deste relatório, contudo, foi recebida apenas a ficha C1 preenchida pela ANPD. Nela, foram relacionados, para cada dimensão examinada, o resultado observado e a evidência correspondente. Os campos referentes à

avaliação da empresa e à conclusão conjunta permaneceram sem preenchimento substantivo.

A) Resultados do Ciclo 4 sob a perspectiva da ANPD

O acompanhamento da plataforma STAiDOC permitiu aprofundar a avaliação regulatória de sistemas de IA utilizados no suporte à decisão médica. Em especial, a testagem contribuiu para compreender os limites entre anonimização, pseudonimização e mera omissão de informações nas respostas. A ANPD destacou a diferença entre a proteção comportamental, baseada em instruções ao modelo, e a proteção estrutural, na qual os dados são filtrados antes de chegarem ao modelo. Essa distinção foi considerada determinante para a qualificação jurídica do tratamento, a avaliação dos riscos e a análise da transferência internacional.

No que tange à proteção dos usuários e titulares de dados, foram examinadas vulnerabilidades relacionadas, notadamente:

- à exposição de informações de filiação mediante estratégias de indução;
- à inferência de características pessoais a partir de nomes próprios; e
- à manutenção de informações no contexto das conversas.

A análise também evidenciou questões ainda dependentes de comprovação, especialmente:

- a efetividade do filtro de privacidade antes do processamento pelo modelo;
- a possibilidade de identificação ou inferência a partir das informações remanescentes; e
- a correspondência entre os mecanismos de proteção declarados e o comportamento observado durante as interações.

Segundo a avaliação registrada pela ANPD, os achados não sustentam a qualificação do tratamento examinado como anonimização, permanecendo necessária evidência técnica verificável de que o filtro opera antes do modelo. Os ensaios utilizaram exclusivamente dados fictícios; assim, as exposições observadas constituem evidências de vulnerabilidades no ambiente experimental, sem representar demonstração de exposição de dados pessoais reais. A identificação dessas fragilidades orienta o aperfeiçoamento dos controles destinados à proteção de pacientes e profissionais de saúde.

Os resultados do Ciclo 4 também produziram subsídios para futuras orientações regulatórias sobre a qualificação de técnicas de anonimização e pseudonimização, a aplicação do regime de transferência internacional ao envio de dados a provedores estrangeiros e a transparência e explicabilidade de decisões automatizadas. A ficha destacou, ainda, a testabilidade dos sistemas como requisito de supervisão, incluindo a definição do contrato de API, o congelamento da versão avaliada e o registro das alterações. Também foram abordadas a consulta a bases públicas, os testes adversariais e a comunicação de incidentes, além do aprendizado metodológico sobre a auditoria de sistemas não determinísticos.

Finalmente, sob a perspectiva mercadológica, a ANPD registrou que o sandbox contribuiu para sinalizar expectativas regulatórias sobre governança, transparência,

privacidade e segurança em soluções de IA para saúde. No âmbito do projeto, foi reconhecida a adequação documental da participante a essas expectativas, mediante recomendações sobre fluxos de dados, RIPD, políticas de privacidade e descarte, registro das atividades de tratamento, gestão de incidentes e mitigação de riscos. A divulgação da versão pública do relatório final poderá orientar outros agentes do setor, sem que esse efeito potencial constitua resultado de mercado já comprovado.

B) Resultados do Ciclo 4 sob a perspectiva da Prevvine

Até o encerramento deste relatório, não foi recebida a ficha C1 preenchida pela Prevvine. Dessa forma, não foi possível consolidar a avaliação da participante sobre suas evoluções técnicas, aprendizados regulatórios, transparência e explicabilidade, direitos dos titulares, análises pendentes e demais aprendizados do Ciclo 4. Os avanços registrados nas fases A e B permanecem como resultados dessas avaliações, sem serem atribuídos à manifestação da empresa na Fase C.

C) Coleta e análise das evidências

A avaliação dos êxitos considerou a Ficha C1 preenchida pela ANPD e sua articulação com os registros das Fases A e B deste relatório. Essa leitura permite distinguir o aprendizado regulatório, a adequação documental, os resultados da testagem e a comprovação da implementação dos controles. Conforme a recomendação da ANPD para este ciclo, a execução das avaliações clínicas e a continuidade dos testes não devem ser confundidas com o cumprimento integral das condicionantes técnicas e regulatórias ou com a aprovação da solução.

Permanece necessária a reconciliação entre a avaliação da ANPD, que exige demonstração técnica do filtro anterior ao modelo, e o registro da Fase B, segundo o qual o parecer técnico reconheceu essa implementação. Essa diferença deverá ser esclarecida mediante identificação das versões testadas, dos momentos de verificação e das evidências correspondentes. O reconhecimento de avanços no parecer não permite desconsiderar os achados da testagem nem presumir que as vulnerabilidades foram integralmente superadas.

A Ficha C1 não apresenta classificação preenchida do êxito global do projeto, principal evidência de êxito ou próximos passos na seção de conclusão conjunta. Portanto, não se atribui ao Ciclo 4 uma conclusão conjunta ou classificação formal de êxito baseada nesse instrumento. A contribuição demonstrada pela avaliação recebida consiste no aprofundamento do aprendizado regulatório e na identificação dos controles que demandam comprovação material.

Para o avanço ao Ciclo 5, permanecem os encaminhamentos registrados nas Fases A e B, especialmente a comprovação da proteção anterior ao modelo, o aperfeiçoamento da supressão de identificadores e da proteção criptográfica, a comprovação da segurança da aplicação e da contingência do PubMed e a consolidação das medidas de explicabilidade. Esses encaminhamentos deverão ser acompanhados por evidências verificáveis, juntamente com o esclarecimento dos fluxos internacionais e das responsabilidades dos agentes de tratamento.

Quanto ao Plano de Descontinuidade (PdD), conforme registrado na Fase B deste relatório, não houve alteração entre os Ciclos 3 e 4. Essa informação não decorre de nova manifestação da participante na Ficha C1, que não havia sido recebida até o encerramento do relatório.

4.3 Metatext

No Ciclo 4, o acompanhamento da solução GuardionAI, da participante Metatext, apresentou avanços relevantes na compreensão de sua arquitetura e na operacionalização dos testes, considerando a complexidade técnica e regulatória e o maior grau de abstração identificado em relação às demais soluções deste sandbox. O objetivo dos próximos ciclos é consolidar esses avanços mediante a análise dos resultados, o aprofundamento da avaliação documental e a verificação das salvaguardas declaradas, ampliando os aprendizados técnicos e regulatórios.

4.3.1 Documentação principal

Preliminarmente, observa-se que os principais documentos referidos neste item do relatório encontram-se anexos conforme a descrição a seguir, por fases, para atendimento dos requisitos pertinentes ao Ciclo 4 deste Sandbox Regulatório:

- Fase A:
  - Fichas A1 e A2; e
  - Fichas A3 e A4.
- Fase B:
  - Parecer técnico;
  - Matriz de transparência;
  - Teste 1 (proteção de dados) — Instruções;
  - Teste 1 — *Prompts*;
  - Teste 2 (alarmes destinados à equipe de segurança) — Instruções; e
  - Teste 2 — Exemplos.

4.3.2 Fase A

Inicialmente, no que se refere às fichas A1 e A2, não houve alterações entre os Ciclos 3 e 4 nas informações referentes à solução GuardionAI, da participante Metatext. Permaneceram como referência a caracterização da aplicação, a arquitetura, os fluxos de dados, a infraestrutura de testes e a autoavaliação da maturidade tecnológica anteriormente apresentadas. O foco do Ciclo 4 foi a estruturação e a execução dos testes da Fase B, destinados a confrontar as declarações documentais com o funcionamento da solução.

Em relação às fichas A3 e A4, também não houve alterações entre os dois ciclos. Foram mantidos os antecedentes da análise técnica e regulatória, especialmente o exame dos instrumentos de governança, das salvaguardas de proteção de dados, da transparência, da supervisão humana e da coordenação entre autoridades. O mapeamento normativo desenvolvido no Ciclo 3, ampliado de cinco para 58 diplomas, permaneceu como referência para a avaliação, com incidência a ser examinada conforme as funcionalidades e o contexto de utilização da solução.

Quanto ao *framework* geral, o registro anterior corresponde à autoavaliação da participante, que declarou atendimento integral a cinco dos sete eixos, com lacunas no Eixo 3 (Avaliação de impactos) e no Eixo 4 (Justiça, não discriminação e vieses

algorítmicos). A manutenção das fichas não permite afirmar evolução da pontuação, conclusão da avaliação pelo CIAAM/USP ou superação dessas lacunas no Ciclo 4.

Permaneceram como questões de acompanhamento a comprovação dos mecanismos de pseudonimização, a identificação dos pontos de supervisão humana, a explicabilidade dos alertas, a governança e a resposta a incidentes. Embora o documento de perguntas e respostas já contenha esclarecimentos da participante, sua efetividade prática deve ser examinada à luz das evidências dos testes. Também permanece necessário esclarecer a localização dos ambientes e dos fluxos de dados, a eventual ocorrência de transferências internacionais, o mecanismo aplicável e a distribuição das responsabilidades entre a participante e as instituições contratantes.

No campo da transparência, foi mantida a necessidade de verificar como as informações produzidas pela GuardionAI são disponibilizadas às instituições financeiras, às equipes de segurança e, por intermédio das contratantes, aos titulares. A ausência de relação direta entre a Metatext e os clientes finais não encerra as questões relativas à explicabilidade, à contestação, à revisão e ao atendimento aos direitos dos titulares.

Em suma, a Fase A do Ciclo 4 preservou a base documental e os encaminhamentos do Ciclo 3. O avanço deste ciclo concentrou-se na Fase B, por meio dos testes de proteção de dados e de envio de alertas à equipe de segurança. Seus resultados devem ser analisados em seção própria, distinguindo controles declarados, evidências verificadas e pendências remanescentes, sem que a manutenção das fichas ou a continuidade da testagem equivalha à aprovação da solução.

4.3.3 Fase B

Inicialmente, a Fase B do Ciclo 4 da solução GuardionAI, da participante Metatext, compreendeu a organização dos aperfeiçoamentos sugeridos na matriz de transparência, a análise do parecer técnico consolidado, a preparação e execução do Teste 1 na plataforma HarpIA e a preparação do Teste 2. Essas atividades deram continuidade aos encaminhamentos do Ciclo 3, especialmente à necessidade de confrontar as declarações documentais com evidências práticas de proteção de dados, segurança, supervisão humana e explicabilidade. A consolidação dos resultados detalhados do Teste 1 e a execução do Teste 2 ficaram previstas para o Ciclo 5.

A) Matriz de transparência

A matriz de transparência foi organizada para quatro grupos de interlocutores: (i) clientes finais das instituições financeiras; (ii) instituições financeiras; (iii) ANPD e demais reguladores; e (iv) equipe de segurança. O instrumento mantém a estrutura de análise baseada nas perguntas “por quê?”, “o quê?”, “quando?” e “como?”, relacionando as informações anteriormente registradas aos aperfeiçoamentos sugeridos para o Ciclo 4.

Para os clientes finais, as recomendações concentram-se na compreensão das negativas de operações, na distinção entre alerta de segurança, bloqueio e decisão da instituição financeira, bem como na disponibilização de canais de esclarecimento e revisão. Para as instituições financeiras, destacam-se a definição das responsabilidades pela configuração das políticas, pela supervisão humana e pelo atendimento aos titulares, além da demonstração dos critérios de bloqueio, das limitações da solução e das métricas de desempenho.

Quanto à ANPD e aos demais reguladores, a matriz reforça a necessidade de organizar evidências auditáveis, esclarecer ambientes, terceiros e eventuais transferências internacionais e harmonizar as informações sobre retenção, descarte e resposta a incidentes. Para a equipe de segurança, deve-se priorizar explicações que apoiem a intervenção operacional, com identificação da política acionada, da severidade, da justificativa do alerta e dos procedimentos de revisão e escalonamento.

Os documentos traduzidos para o português foram recebidos e serão analisados no Ciclo 5. Portanto, seu recebimento não permite considerar concluída a avaliação documental nem atendidos os aperfeiçoamentos indicados na matriz. Também permanece necessário verificar como as informações fornecidas pela Metatext permitem às instituições contratantes responder adequadamente aos titulares, uma vez que a ausência de contato direto com os clientes finais não encerra as questões de transparência, explicabilidade e revisão.

B) Parecer técnico: governança, segurança da informação e privacidade

O parecer técnico consolidado identificado como referente ao Ciclo 4 apresenta conclusão "suficiente com condicionantes" e recomenda o prosseguimento dos testes. Registra risco inerente agregado alto e risco residual agregado médio, considerando os controles descritos na documentação. A análise abrange políticas de gestão de acessos, criptografia e retenção de dados, o Relatório de Impacto à Proteção de Dados e o plano de resposta a incidentes:

Quadro 2: Classificação de Indicadores da solução GuardionAI

Indicador de Avaliação	***Classificação***
Conclusão Preliminar:	Suficiente com condicionantes.
Recomendação Principal:	Prosseguir com testes.
Nível Agregado de Risco Inerente:	Alto
Nível Agregado de Risco Residual:	Médio
Mitigações Documentadas	• Política de Retenção;

	• Gestão de acessos e autenticação multifator (*Multi-factor Authentication* — MFA); • Melhoria no método criptográfico; • Plano de resposta a incidentes documentado.
Mitigações Planejadas Críticas:	• Certificações SOC 2 Type II e ISO 27001 • Comprovação de execução prática PenTest focado em LLM.

Fonte: ANPD

Entre as medidas consideradas pelo parecer estão a autenticação multifator, o controle de acesso por perfis, a proteção criptográfica, a implementação de protocolo de segurança na camada de transporte, o monitoramento de eventos e a definição de prazos de retenção dos registros. Foram mapeados riscos relacionados à inserção de informações financeiras, ao acesso não autorizado aos registros de interação, ao uso indevido por agentes internos e à utilização secundária dos dados para treinamento de modelos.

Embora o quadro de controles classifique determinadas medidas como efetivadas, o próprio parecer ressalta que a avaliação se apoia predominantemente na documentação e que permanecem lacunas de comprovação prática. Assim, a classificação de risco residual médio não deve ser interpretada como confirmação experimental da eficácia integral dos controles, especialmente dos mecanismos de detecção e ofuscação de dados pessoais.

O parecer estabelece duas condicionantes principais: apresentação de evidências do último teste de intrusão, incluindo ataques específicos contra modelos de linguagem, como injeção de instruções e extração de dados; e comprovação do progresso das auditorias e certificações SOC 2 Type II e ISO 27001. Também prevê gatilhos de interrupção diante de vazamento, desvio de finalidade ou falhas de retenção e cópias de segurança. A recomendação de continuidade permanece sujeita ao acompanhamento dessas condicionantes.

C) Teste 1: Proteção de dados — preparação e execução na plataforma HarpIA

No Ciclo 4, o Teste 1 foi preparado e executado na plataforma HarpIA, destinado a avaliar o tratamento dos identificadores presentes nas mensagens encaminhadas à GuardionAI. O procedimento consiste no envio de prompts e na análise do texto devolvido pelo sistema, por meio de um formulário de avaliação da correção e da coerência das transformações realizadas.

O instrumento estabelece um conjunto mínimo padronizado de trinta prompts, com nomes, documentos de identificação, dados bancários, cartões, senhas, contatos,

endereços e outros elementos. Inclui variações de grafia, capitalização e formatação, além de situações em que a retirada de determinado conteúdo pode comprometer o significado da mensagem. Ademais, a orientação de preservar os enunciados, inclusive suas particularidades intencionais, visa permitir a comparabilidade entre as avaliações.

Os resultados detalhados serão compartilhados no Ciclo 5. Dessa forma, por enquanto, o presente relatório registra a preparação e a execução do Teste 1.

D) Teste 2: Informações destinadas à equipe de segurança: preparação na plataforma HarpIA

No Ciclo 4, o Teste 2 foi preparado na plataforma HarpIA, dedicado à avaliação das informações apresentadas à equipe de segurança. O procedimento prevê a análise de trinta alarmes produzidos pela GuardionAI, apresentados em painéis com informações sobre o evento, sua classificação e as evidências associadas. Foram organizados instruções, diretrizes e exemplos para orientar o preenchimento dos formulários.

Os materiais direcionam a avaliação à correspondência entre o título do incidente, os indicadores detectados, o conteúdo que originou o alerta e as justificativas produzidas pelo sistema. Também orientam o exame do nível de risco e da classificação da ocorrência, permitindo investigar a consistência e a utilidade das informações para a atuação humana.

A execução desse teste ocorrerá no Ciclo 5. Portanto, as amostras produzidas até então constituem instrumentos exemplificativos de preparação.

Em conjunto, a Fase B do Ciclo 4 avançou na operacionalização da avaliação da GuardionAI, com execução do Teste 1 e preparação do Teste 2. Permanecem para o Ciclo 5 a análise dos documentos traduzidos, a consolidação dos resultados detalhados do Teste 1, a execução do Teste 2 e o acompanhamento das condicionantes do parecer. Esses elementos deverão permitir distinguir as salvaguardas declaradas das efetivamente verificadas, preservando as questões ainda abertas sobre proteção de dados, transferências internacionais, responsabilidades, explicabilidade e resposta a incidentes.

4.3.4 Fase C

O Ciclo 4 da solução GuardionAI, da participante Metatext, atingiu o principal objetivo estabelecido para a Fase B: iniciar a execução dos testes na plataforma HarpIA. Foi executado o Teste 1, relativo à proteção de dados, cujos resultados detalhados serão divulgados no Ciclo 5, considerando a complexidade das situações examinadas e a necessidade de análise conjunta sob as perspectivas técnica e regulatória.

Nesse contexto, não foi possível realizar a avaliação conclusiva prevista para a Fase C do Ciclo 4, uma vez que os resultados ainda não haviam sido disponibilizados para consolidação dos achados, aprendizados e encaminhamentos. Essa avaliação será realizada no Ciclo 5. Também nesse próximo ciclo ocorrerá a execução do Teste 2, composto por trinta alarmes destinados à equipe de segurança, cuja preparação foi concluída no Ciclo 4.

Assim, a partir do Ciclo 5, será possível completar o percurso metodológico das Fases A, B e C, articulando a caracterização da solução, a execução dos testes e a avaliação dos resultados. Com os dois testes previstos executados e analisados detalhadamente pelas equipes técnica e regulatória, será possível consolidar os

aprendizados, verificar a efetividade das salvaguardas declaradas e identificar as pendências e recomendações para a continuidade do projeto.

5. Consulta Pública

A Consulta Pública sobre o Sandbox Regulatório foi concebida como um processo participativo em duas etapas complementares: uma etapa multissetorial, destinada a promover o diálogo entre diferentes setores interessados no desenvolvimento e na regulação de soluções envolvendo inteligência artificial (IA) e tratamento de dados pessoais; e uma etapa dedicada às consultas setoriais, com o objetivo de possibilitar uma escuta mais direcionada e identificar perspectivas, experiências e preocupações específicas de cada setor (serviços financeiros, saúde e transporte).

A consulta multissetorial foi realizada no Ciclo 4 e esta seção trata de sua metodologia e dos aprendizados de processo. As consultas setoriais com a Agência Nacional de Transportes Terrestres (ANTT), com o Banco Central (Bacen), com a Agência Nacional de Saúde Suplementar (ANS) e com a Agência Nacional de Vigilância Sanitária (ANVISA) estão previstas para ocorrerem no Ciclo 6, e os resultados dessas interações serão apresentados no relatório correspondente.

A opção por combinar momentos de escuta multissetorial e uma etapa posterior de diálogo setorial decorreu da compreensão de que os desafios relacionados à proteção de dados pessoais e ao uso de IA possuem, simultaneamente, dimensões transversais, que demandam o diálogo entre diferentes perspectivas, e dimensões específicas, que podem ser mais bem captadas a partir da experiência de cada setor. A estratégia permitiu, assim, promover a interlocução entre perspectivas acadêmicas, regulatórias, empresariais, governamentais e da sociedade civil e, posteriormente, aprofundar a escuta dos diferentes grupos.

As seções a seguir apresentam a metodologia adotada nas consultas e os principais aprendizados decorrentes de sua realização.

5.1. Consulta multissetorial

A Consulta Multissetorial criou um espaço para que representantes da academia, do setor privado, do setor público e da sociedade civil, juntamente com as empresas participantes do Sandbox Regulatório e as equipes da ANPD e do CIAAM/USP, pudessem discutir desafios regulatórios identificados a partir da experiência de implementação das três soluções testadas.

A consulta foi realizada em 13 de agosto de 2026, das 15h às 17h, em ambiente remoto, por meio da plataforma Microsoft Teams. Foi também oportunizada a participação antes e depois do encontro, de modo a permitir uma avaliação mais ampla do processo participativo.

A metodologia do encontro foi estruturada para alcançar, simultaneamente, quatro objetivos:

- apresentar aos participantes informações suficientes sobre o Sandbox Regulatório da ANPD, sua metodologia e seus resultados preliminares;
- permitir que as empresas participantes apresentassem suas experiências e soluções, aproximando o debate regulatório das situações concretas enfrentadas durante a testagem;

- promover um debate multissetorial estruturado, evitando que a discussão se concentrasse em uma única dimensão do problema; e
- identificar desafios, lacunas e possíveis aperfeiçoamentos regulatórios que pudessem subsidiar as próximas etapas do projeto piloto.

Um dos elementos centrais da metodologia foi a composição diversificada do grupo de participantes, contemplando quatro setores (vide quadro a seguir):

Quadro 3: Participantes da consulta pública multissetorial

Setor	***Participantes***
Academia	• Universidade de Fortaleza (UNIFOR); e • Universidade Federal de Alagoas (UFAL).
Setor privado	• FSB Holding; e • Opice Blum Advogados.
Setor público	• Ministério Público do Rio de Janeiro (MPRJ); e • Instituto Nacional de Seguro Social (INSS).
Sociedade Civil	• Datasphere Initiative; e • Desvelar.

Fonte: ANPD

A composição buscou evitar que o debate fosse conduzido exclusivamente por especialistas acadêmicos ou representantes diretamente envolvidos na implementação das soluções. A presença simultânea de diferentes setores permitiu confrontar perspectivas distintas sobre questões como transparência, supervisão humana, governança de dados, discriminação, proteção dos titulares e lacunas regulatórias.

A escolha dos participantes também contribuiu para aproximar o debate de diferentes tipos de conhecimento: acadêmico e técnico, empresarial, institucional do setor público e da sociedade civil. A definição das entidades participantes observou um critério objetivo: a ordem de inscrição válida para o encontro.

A consulta foi estruturada em três etapas complementares: (i) formulário pré-participação; (ii) consulta pública multissetorial; e (iii) formulário pós-participação. Essa estrutura buscou assegurar que a participação não se restringisse ao momento síncrono do encontro.

O formulário pré-participação tinha como finalidade conhecer o perfil dos participantes, seu grau de familiaridade com proteção de dados, inteligência artificial e Sandboxes Regulatórios, além de identificar suas motivações, expectativas e temas considerados prioritários. O formulário pós-participação, por sua vez, foi concebido como instrumento de avaliação do processo, permitindo aferir a percepção dos participantes sobre a organização, a qualidade das discussões, a possibilidade efetiva de participação e a utilidade da consulta.

Esse formato permitiria analisar a consulta não apenas a partir das contribuições produzidas durante o encontro, mas também da percepção dos participantes sobre a qualidade e a efetividade do processo participativo.

Na reunião, foi inicialmente reservado um espaço para a contextualização do Sandbox Regulatório, com o objetivo de estabelecer uma base comum de informações entre participantes com níveis potencialmente distintos de conhecimento sobre o projeto piloto. Dessa forma, a consulta não pressupôs que todos chegassem ao encontro com o mesmo nível de conhecimento técnico ou regulatório.

Em seguida, foram destinados cinco minutos para que cada empresa participante apresentasse sua solução, totalizando 15 minutos. A definição desse tempo buscou evitar que a consulta se transformasse predominantemente em apresentações institucionais e ampliar o espaço disponível para a interação entre os participantes. Informações mais detalhadas poderiam ser apresentadas posteriormente, caso necessário, em resposta às perguntas formuladas durante o debate.

A principal escolha metodológica para a etapa de participação foi a organização das perguntas em quatro blocos temáticos, com 15 minutos destinados a cada bloco. A opção por perguntas abertas, organizadas por temas, procurou conciliar duas necessidades: (i) conferir estrutura e foco ao debate; e (ii) preservar espaço para que os participantes apresentassem perspectivas, exemplos e questões não antecipadas pela organização. Os quatro blocos foram:

- Bloco I: transparência e supervisão;
- Bloco II: anonimização e governança de dados;
- Bloco III: discriminação e impactos desproporcionais; e
- Bloco IV: lacunas regulatórias.

Além das perguntas abertas, as equipes da ANPD e do CIAAM/USP elaboraram questões norteadoras previamente, que não foram concebidas como um questionário rígido a ser respondido sequencialmente pelos participantes. Sua função era estimular o debate e assegurar que os principais temas identificados na metodologia fossem contemplados. Essa flexibilidade é especialmente importante em uma consulta multissetorial, pois permite incorporar questões emergentes durante a discussão sem comprometer a estrutura temática originalmente planejada.

Essa opção metodológica procurou equilibrar dois elementos potencialmente tensionados em processos participativos: de um lado, a necessidade de oferecer informações suficientes para que os participantes pudessem contribuir de maneira qualificada; de outro, a necessidade de reservar tempo adequado para que as contribuições efetivamente ocorressem.

A consulta foi concebida não apenas como mecanismo de coleta de opiniões, mas também como instrumento de produção de conhecimento e evidências para o aprimoramento do Sandbox Regulatório, especialmente quanto à identificação de riscos, dificuldades práticas, lacunas normativas e elementos que possam subsidiar futuras decisões regulatórias.

Em síntese, a consulta pública multissetorial foi estruturada com o propósito de criar um espaço de diálogo entre diferentes setores sobre os desafios regulatórios associados ao Sandbox da ANPD. A metodologia buscou combinar contextualização, participação das empresas, debate multissetorial e organização temática das discussões, ao mesmo tempo em que evidenciou a importância de estratégias específicas de engajamento e acompanhamento das partes interessadas.

5.2. Conclusões gerais da consulta pública

Nesta seção foram destacados os aprendizados decorrentes da realização da consulta multissetorial.

Inicialmente, cumpre ressaltar que os questionários previstos como instrumentos de avaliação da consulta não foram preenchidos conforme esperado. Por essa razão, os dados previstos nesses instrumentos não puderam ser utilizados para avaliar, quantitativa ou qualitativamente, o perfil, as expectativas ou a percepção dos participantes sobre a consulta. Essa circunstância constitui uma limitação da avaliação do processo participativo, mas também oferece um aprendizado relevante para futuras consultas: a previsão de instrumentos complementares de participação, por si só, não é suficiente para garantir o engajamento dos participantes.

Assim, é necessário considerar, desde a etapa de planejamento, estratégias específicas para mobilizar os diferentes públicos e comunicar claramente a importância de sua contribuição e a finalidade dos instrumentos de participação. A experiência evidencia, portanto, a necessidade de investir de forma mais sistemática no engajamento prévio e posterior das partes interessadas. A participação em uma consulta multissetorial depende não apenas da existência de um espaço formal de diálogo, mas também da criação de condições para que os participantes compreendam a relevância de sua contribuição e se sintam efetivamente envolvidos no processo.

Para futuras iniciativas, recomenda-se, portanto, que os instrumentos complementares sejam:

- apresentados aos participantes com antecedência, juntamente com uma explicação clara sobre sua finalidade e sobre a forma como as respostas serão utilizadas;
- disponibilizados, preferencialmente, em formato de preenchimento rápido e acessível;
- objeto de comunicação e lembretes específicos antes e depois da consulta, integrados, sempre que possível, à própria dinâmica do encontro, inclusive mediante disponibilização de QR Code ao final da atividade; e
- acompanhados de estratégias de acompanhamento (*follow-up*) com os diferentes participantes.

O aprendizado decorrente da experiência indica, assim, que o engajamento das partes interessadas deve ser tratado como uma dimensão própria da metodologia, e não apenas como consequência da realização da consulta.

Além desses aspectos, a experiência permite identificar outros aprendizados relevantes para o aprimoramento de futuras consultas, os quais são listados a seguir:

- a diversidade de participantes precisa ser acompanhada de estratégias de engajamento: a composição multissetorial foi um elemento importante da consulta, reunindo representantes da academia, do setor privado, do setor público e da sociedade civil. É necessário desenvolver estratégias específicas para preparar, mobilizar e manter o engajamento dos diferentes grupos, considerando seus distintos interesses, níveis de disponibilidade e formas de contribuição.
- o tempo de participação deve ser protegido: a decisão de atribuir cinco minutos a cada empresa participante foi orientada pela intenção de ampliar o espaço destinado ao diálogo. A experiência reforça a importância de preservar esse equilíbrio: quanto mais eficiente for a etapa expositiva, maior tende a ser o espaço disponível para a participação efetiva dos diferentes atores.
- perguntas estruturadas em blocos podem favorecer a participação qualificada: a divisão do debate em quatro blocos temáticos mostrou-se adequada como mecanismo de organização da discussão. A estrutura permitiu abordar diferentes dimensões do Sandbox sem restringir as manifestações a um questionário rígido. A combinação de perguntas norteadoras e espaço para manifestações abertas deve ser preservada em futuras consultas, especialmente em temas tecnicamente complexos.
- as reuniões devem combinar estrutura prévia e flexibilidade para contemplar as demandas que emergirem durante a discussão: a experiência demonstrou que a estrutura previamente definida não deve limitar o desenvolvimento do diálogo. Ao longo das reuniões, podem surgir dúvidas, demandas de esclarecimento ou temas que, embora não previstos nas perguntas iniciais, sejam relevantes para os participantes e para os objetivos da consulta. É importante, portanto, que a metodologia reserve margem para adaptar a ordem, o tempo e o conteúdo das discussões, permitindo aprofundar questões que se mostrem particularmente relevantes durante o encontro. A combinação entre perguntas norteadoras, espaço para manifestações abertas e flexibilidade na condução da reunião tende a favorecer uma participação mais qualificada, especialmente em temas tecnicamente complexos.
- é importante transformar a participação em uma relação continuada: a consulta pública pode ser mais efetiva quando compreendida como parte de um processo continuado de relacionamento com as partes

interessadas, e não como um evento isolado. Nesse sentido, futuras iniciativas podem prever, além da consulta, mecanismos de acompanhamento, como comunicação posterior sobre os resultados, retorno sobre a incorporação das contribuições e eventual convite para novas etapas de discussão. Essa estratégia pode contribuir para aumentar a percepção dos participantes de que suas contribuições produzem efeitos concretos no processo regulatório, fortalecendo o interesse em participar de futuras iniciativas.

6. Conclusões gerais do Ciclo 4

A organização nas Fases A, B e C permaneceu adequada para relacionar a caracterização das soluções, os testes e os encaminhamentos, respeitadas as diferenças de execução e consolidação dos resultados entre as participantes.

Os resultados devem ser interpretados conforme o escopo efetivamente testado, a maturidade tecnológica, a documentação apresentada, as evidências verificadas e os limites metodológicos de cada avaliação. A utilização de instrumentos comuns não autoriza comparações diretas ou conclusões uniformes sobre projetos em diferentes contextos de aplicação, estágios de desenvolvimento e comprovação.

Em primeiro lugar, quanto à Synapse, o Ciclo 4 aprofundou a avaliação documental e regulatória do Trajetto e acrescentou esclarecimentos à matriz de transparência. A pontuação no *framework* geral passou de 68 para 78 pontos, e o parecer manteve a avaliação suficiente com condicionantes, com risco inerente alto e risco residual médio. Os quatro testes técnicos haviam sido encerrados no Ciclo 3, não sendo atribuídas ao Ciclo 4 novas conclusões decorrentes desses ensaios. O êxito permaneceu parcial, diante das pendências de comprovação da transparência na interface, da acessibilidade, do inventário de API e terceiros, da não retenção de imagens, dos controles de segurança e da mitigação da reidentificação indireta. Também permanecem necessárias evidências do teste de descontinuidade declarado pela participante.

Em segundo lugar, quanto à Prevvine, o Ciclo 4 apresentou avaliações clínicas efetivamente realizadas na HarpIA, avanços documentais e medidas corretivas. A pontuação no *framework* geral passou de 63 para 75 pontos. Entretanto, as avaliações identificaram limitações clínicas e bibliográficas, e permaneceram condicionantes relativas à explicabilidade, à proteção criptográfica, à segurança da aplicação, à contingência do PubMed e à supressão de identificadores. O parecer manteve conclusão favorável com condicionantes, com risco inerente crítico e risco residual agregado médio, cuja consolidação requer esclarecimento diante dos riscos individuais ainda classificados como altos. Os resultados clínicos devem ser examinados separadamente do cumprimento das condicionantes técnicas e regulatórias. A Fase C contou apenas com a ficha preenchida pela ANPD, sem conclusão conjunta ou classificação formal de êxito no instrumento.

Em terceiro lugar, quanto à Metatext, responsável pela GuardionAI, foi atingido o objetivo de iniciar a execução dos testes na HarpIA. O Teste 1, relativo à proteção de dados, foi executado, e o Teste 2, composto por trinta alarmes destinados à equipe de segurança, foi preparado para execução no Ciclo 5. A divulgação dos resultados detalhados do Teste 1 e sua avaliação conclusiva também ocorrerão no próximo ciclo, considerando a complexidade técnica e regulatória. O parecer recomendou continuidade com condicionantes, mas ressaltou lacunas de comprovação prática. Permanecem abertas questões sobre a efetividade da proteção de dados, os fluxos internacionais, as responsabilidades, a supervisão humana, a explicabilidade e a resposta a incidentes. Os

documentos traduzidos, recebidos em 13 de outubro de 2026, serão analisados no Ciclo 5.

De modo geral, o Ciclo 4 reforçou a necessidade de distinguir controles declarados, documentados, implementados e efetivamente verificados. A evolução documental e das pontuações de avaliação constitui evidência relevante, mas deve ser confrontada com os resultados experimentais. A qualificação das técnicas de proteção de dados deve acompanhar as evidências disponíveis.

Para o Ciclo 5, recomenda-se priorizar:

- acompanhamento das condicionantes: registrar responsáveis, prazos, evidências exigidas e critérios de conclusão, atualizando os riscos residuais conforme os resultados verificados.
- Synapse: comprovar transparência e acessibilidade na interface, integrações com terceiros, não retenção de imagens, controles de segurança, mitigação da reidentificação indireta e execução das medidas de descontinuidade.
- Prevvine: aprofundar os testes técnicos e regulatórios, esclarecer a efetividade do filtro anterior ao modelo e acompanhar as condicionantes ainda abertas, relacionando as versões testadas às evidências e aos achados.
- Metatext: consolidar e divulgar os resultados detalhados do Teste 1, executar e analisar o Teste 2, examinar os documentos traduzidos e completar a avaliação da Fase C, incluindo as informações destinadas aos titulares, às instituições contratantes e à equipe de segurança.

De modo transversal, permanecem necessárias:

- complementação dos inventários de dados, integrações e terceiros, com definição das responsabilidades e dos procedimentos de retenção, exclusão e atendimento aos titulares.
- registro e versionamento dos modelos, API, interfaces e ambientes, preservando a rastreabilidade e a comparabilidade dos testes.
- atualização das avaliações de maturidade tecnológica, dos riscos e dos Planos de Descontinuidade, conforme as evidências e alterações identificadas.
- devolutivas entre participantes, CIAAM/USP e ANPD para esclarecimento de divergências, análise setorial e definição dos encaminhamentos.

O Ciclo 5 deverá reunir novos resultados, principalmente da Metatext e da Prevvine, decorrentes da execução, consolidação e análise dos testes técnicos e regulatórios. Esses resultados permitirão aprofundar a avaliação da efetividade dos controles, identificar medidas ainda insuficientes e ampliar os aprendizados da ANPD, com explicitação dos limites de cada conclusão.

Conclui-se que o principal resultado do Ciclo 4 foi o avanço na produção e na análise das evidências, acompanhado da delimitação mais precisa das pendências de comprovação. A progressão ao Ciclo 5 deverá fortalecer a articulação entre avaliação técnica e análise regulatória, orientando o desenvolvimento das soluções e a proteção dos titulares mediante acompanhamento das condicionantes e verificação das salvaguardas.


sandbox regulatório


ANPD