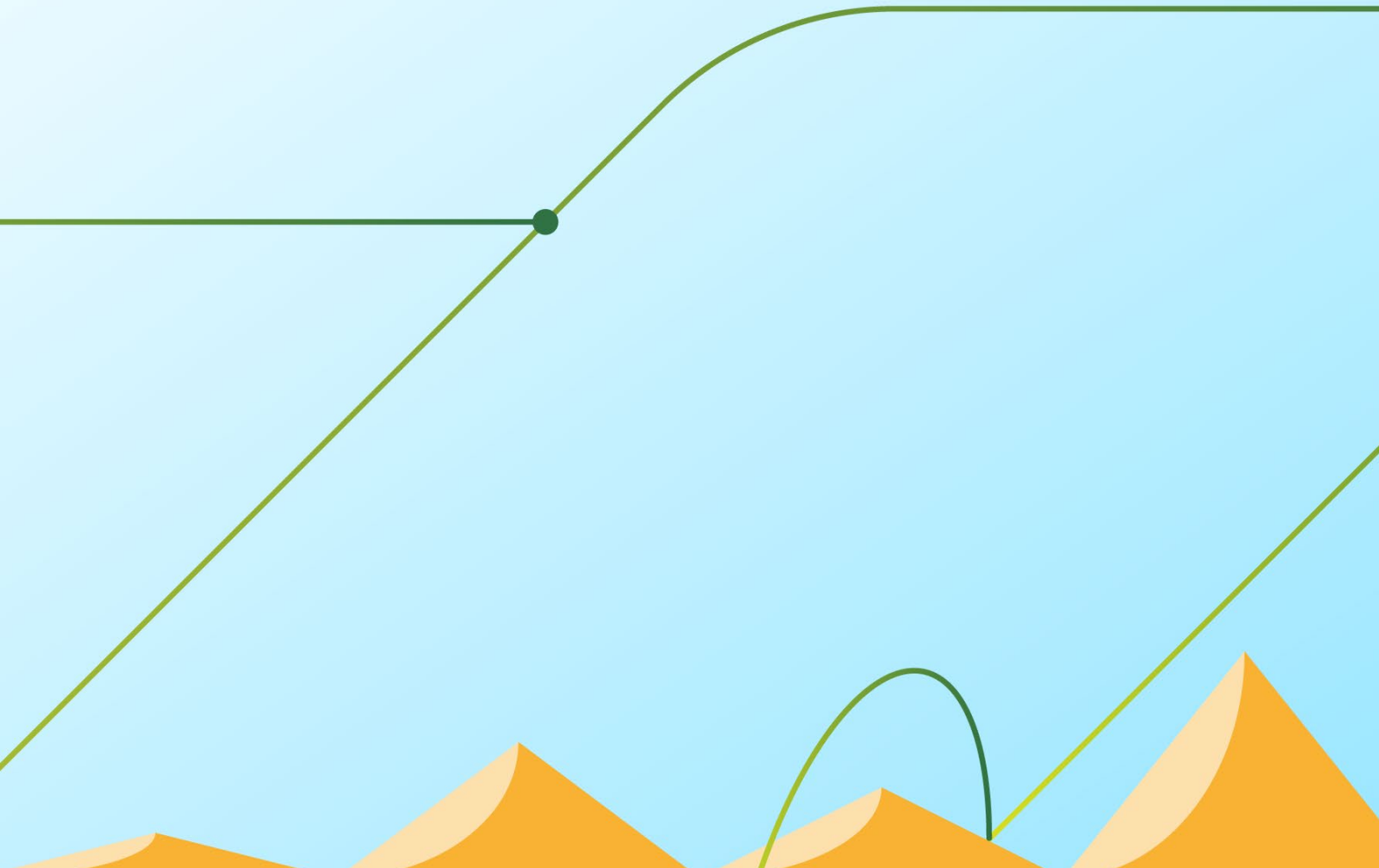
ANPD

sandbox regulatório

5º Relatório Parcial de Monitoramento

Fase de Teste – Ciclo 5



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 10

4.1 Synapse 10

4.1.1 Documentação principal 11

4.1.2 Fase A 12

4.1.3 Fase B 17

4.1.4 Fase C 22

4.2 Prevvine 24

4.2.1 Documentação principal 26

4.2.2 Fase A 26

4.2.3 Fase B 31

4.2.4 Fase C 35

4.3 Metatext 37

4.3.1 Documentação principal 38

4.3.2 Fase A 39

4.3.3 Fase B 39

4.3.4 Fase C 43

5. Conclusões gerais do Ciclo 5 45

1. Introdução

O presente relatório apresenta o acompanhamento do Ciclo 5 do Projeto Piloto de Sandbox Regulatório de Inteligência Artificial e Proteção de Dados das participantes Synapse[1], com o projeto Trajetto; Prevvine[2], com a plataforma STAiDOC; e Metatext[3], com a solução GuardionAI. Em continuidade ao Ciclo 4, o documento examina a relação entre as características das aplicações, os resultados da experimentação e os requisitos de proteção de dados, segurança, transparência e governança, oferecendo subsídios à preparação do ciclo final.

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. Conforme disponibilidade para cada participante, as atividades compreenderam: (i) orientações; (ii) análises documentais; (iii) preparação e execução de testes; (iv) avaliação de riscos; (v) acompanhamento de condicionantes; (vi) instrumentos de avaliação ética; (vii) avaliações de maturidade tecnológica; (viii) documentos de arquitetura e governança; (ix) matrizes de transparência; (x) registros da HarpIA[5]; (xi) fichas de êxitos[6] ; e (xii) planos de descontinuidade.

Na Fase B, foi utilizado um *framework* geral baseado em recomendações da UNESCO[7] (doravante tratado como *framework* geral), o qual estrutura a avaliação ética de sistemas de Inteligência Artificial (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);

---

[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: 27 de outubro de 2026.

[5] HarpIA: plataforma para avaliação de grandes modelos de linguagem (*Large Language Models* — LLM).

[6] 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.

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

- Eixo 6 — Segurança técnica e robustez (10 pontos); e
- Eixo 7 — Proporcionalidade e necessidade (10 pontos).

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.

No caso da Synapse, o relatório aborda o acompanhamento do desenvolvimento do Trajetto e das recomendações formuladas nos ciclos anteriores. São examinados o uso de mapas geolocalizados, o tratamento de imagens, as integrações com terceiros e as informações destinadas aos usuários. A análise contempla, em particular, a acessibilidade e os riscos associados à combinação de localização com atributos utilizados na recomendação de rotas, preservando a identificação dos ensaios concluídos em ciclos anteriores.

Quanto à Prevvine, o documento trata das retificações na descrição da STAiDOC, das alterações de interface e das avaliações realizadas com cenários clínicos. O exame abrange a supressão de identificadores, a qualidade das respostas e das referências bibliográficas, a adequação aos perfis de pacientes e os mecanismos de supervisão humana. Também considera a correspondência entre os resultados dos testes, a avaliação ética, os controles descritos no parecer técnico e as medidas previstas para versões posteriores.

Em relação à Metatext, são apresentados o acompanhamento da pseudonimização na GuardionAI e os procedimentos de avaliação das informações fornecidas à equipe de segurança. O relatório examina os resultados disponíveis sobre detecção de identificadores, a organização das explicações conforme os destinatários e as verificações requeridas pelo parecer técnico. Também delimita as análises documentais e a consolidação dos resultados do Teste 02 — avaliação das informações destinadas às equipes de segurança — previstas para o Ciclo 6.

A leitura das três experiências é orientada por questões comuns: a correspondência entre funcionalidades declaradas e comportamento observado; a distribuição de responsabilidades; o atendimento aos titulares; os fluxos de dados; e a demonstração da efetividade das salvaguardas.

Com essa organização, o documento permite acompanhar o percurso de avaliação de cada participante e compreender os fundamentos dos encaminhamentos ao Ciclo 6. Seu propósito é reunir os registros da experimentação e os aprendizados produzidos, situando as questões que deverão ser examinadas na consolidação e no encerramento do projeto piloto.

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

No Ciclo 5, o planejamento das sessões de orientação deu continuidade ao acompanhamento individualizado das participantes, considerando os registros do Ciclo 4, o estágio de desenvolvimento das soluções e as pendências de avaliação. A organização das atividades contemplou o acompanhamento dos testes, a solicitação de esclarecimentos e documentos e a identificação das evidências necessárias à verificação das recomendações técnicas e regulatórias.

As interações entre as participantes, as equipes técnica e regulatória do CIAAM/USP e a Comissão de Sandbox da ANPD foram articuladas às análises das fases A, B e C. Além das sessões previstas no calendário, o acompanhamento compreendeu discussões entre as equipes e intercâmbio de informações pelo Microsoft Teams, permitindo relacionar os esclarecimentos das empresas aos instrumentos de avaliação e aos registros da experimentação.

Conforme o calendário disponibilizado, as sessões foram programadas para 10 e 17 de setembro de 2026, com trinta minutos destinados a cada participante. Da primeira, participaram Synapse, responsável pelo Trajetto, das 15h às 15h30; e Prevvine, responsável pela STAiDOC, das 15h30 às 16h. A Metatext participou na segunda data, das 16h às 16h30, com o respectivo responsável pela GuardionAI.

Para a Synapse, o acompanhamento foi direcionado aos esclarecimentos sobre mapas geolocalizados, processamento de imagens, integrações com terceiros, transparência e acessibilidade. A orientação concentrou-se na apresentação de evidências das medidas descritas e na avaliação do risco de reidentificação pela combinação de localização e atributos de acessibilidade. Foram preservados os resultados e os limites dos ensaios encerrados nos ciclos anteriores.

Para a Prevvine, a organização das atividades contemplou as avaliações conversacionais na HarpIA, a análise da supressão de identificadores, a qualidade clínica e bibliográfica e os mecanismos de supervisão humana. O acompanhamento também abrangeu as retificações documentais, as alterações de interface e as condicionantes de segurança e governança, considerando a necessidade de identificar as versões, os canais e as condições aos quais cada evidência se refere.

Para a Metatext, o planejamento contemplou a avaliação do desempenho da pseudonimização e das informações apresentadas à equipe de segurança. As atividades incluíram o acompanhamento do Teste 02 (sobre as informações destinadas à equipe de segurança), a organização da documentação complementar e a delimitação das verificações requeridas pelo parecer técnico. A consolidação desse teste e a análise conclusiva dos documentos recebidos foram previstas para o Ciclo 6.

Dessa forma, a estruturação das sessões no Ciclo 5 vinculou a orientação das participantes à produção e à análise de evidências. O acompanhamento subsidiou a identificação das correções, dos retestes e das comprovações necessários ao ciclo final, incluindo a verificação das condicionantes e dos procedimentos de resposta a incidentes e descontinuidade.

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

O Ciclo 5 contemplou o monitoramento das atividades de orientação desenvolvidas durante o quinto mês da fase de testes. A definição dos temas e da metodologia das sessões foi articulada às atribuições de assessoramento às participantes, coleta de informações, registro das dificuldades, análise dos resultados e identificação de oportunidades de melhoria, sob acompanhamento das equipes do CIAAM/USP e da ANPD.

No Ciclo 5, a seleção dos temas tomou como referência as recomendações do ciclo anterior, os planos de sandbox, os documentos apresentados e as questões identificadas na experimentação. Foram mantidos os temas de proteção de dados, segurança da informação, governança, transparência, explicabilidade, supervisão humana e direitos dos titulares. Seu aprofundamento considerou as características de cada aplicação e as evidências necessárias à avaliação das medidas de mitigação.

Para a Synapse, os temas compreenderam geolocalização, acessibilidade, processamento de imagens, integrações com terceiros e risco de reidentificação pela associação entre localização e atributos utilizados nas recomendações de rotas. Para a Prevvine, abrangeram a supressão de identificadores, a qualidade clínica e bibliográfica, a avaliação de vieses, a intervenção humana e a correspondência entre funcionalidades descritas e versões avaliadas. Para a Metatext, concentraram-se no desempenho da pseudonimização, na consistência das informações dos alertas, nas explicações destinadas aos diferentes interlocutores e na comprovação dos controles de segurança e governança.

A metodologia manteve a organização nas Fases A — Pré-teste; B — Teste; e C — Pós-teste. As sessões de orientação foram articuladas à leitura das fichas e dos documentos, à formulação de questionamentos, ao acompanhamento dos testes e à discussão dos registros produzidos. As reuniões remotas e o intercâmbio pelo Microsoft Teams permitiram complementar esclarecimentos e solicitar evidências conforme as necessidades de cada avaliação.

Na análise dos testes, foram considerados os objetivos de cada procedimento, os critérios empregados e as condições de execução. Nas avaliações conversacionais da Prevvine, isso compreendeu a contextualização das vinhetas, os julgamentos dos especialistas e as diferenças entre os canais de avaliação e de utilização clínica. Na Metatext, contemplou a distinção entre o desempenho da detecção de identificadores e a avaliação das informações apresentadas à equipe de segurança. Na Synapse, o acompanhamento preservou os resultados dos ensaios anteriores e concentrou-se nas questões ainda dependentes de esclarecimento e comprovação.

A leitura das evidências distinguiu declarações das participantes, documentos, funcionalidades implementadas e controles submetidos à verificação. Os pareceres técnicos, as avaliações pelo *framework* geral, as matrizes de transparência e as fichas de êxitos foram examinados conforme suas finalidades, preservando as diferenças entre seus resultados. A identificação das versões, dos ambientes e das alterações constituiu referência para relacionar cada achado ao objeto efetivamente avaliado.

As devolutivas às participantes foram orientadas à delimitação das falhas, das informações faltantes e das providências necessárias ao prosseguimento das análises.

Esse procedimento permitiu organizar solicitações documentais, correções, retestes e verificações para o Ciclo 6, sem antecipar o atendimento das recomendações.

Assim, a metodologia do Ciclo 5 vinculou a orientação das participantes ao monitoramento da experimentação e à produção de subsídios para a ANPD. Os registros consolidados no item 4 constituem a base para acompanhar as condicionantes e organizar as verificações do ciclo final.

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 5, 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 5 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 6.

4.1 Synapse

O relatório do CIAAM/USP referente ao Ciclo 5 apresenta os principais dados, informações e aprendizados relativos à solução Trajetto, da participante Synapse, consolidando os registros das fases A e B e as avaliações de êxitos disponibilizadas pela ANPD e pela empresa na Fase C. A análise considera 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, incluindo os testes de mapas geolocalizados e os esclarecimentos sobre o tratamento de imagens, as integrações externas e as funcionalidades ainda em desenvolvimento;
- Dados, privacidade e segurança: análise das informações e evidências registradas nas fichas A3 e A4 e nos instrumentos B1 e B2, abrangendo as salvaguardas de segregação, pseudonimização, criptografia e controle de acesso, bem como as pendências de comprovação relativas ao processamento de imagens, à retenção de dados e às integrações com terceiros;
- Maturidade tecnológica (TRL[8] e ETC[9]): consideração da autoavaliação apresentada pela participante e das análises da equipe do CIAAM/USP, conforme as evidências disponíveis. No Ciclo 5, foram mantidos o TRL 4 para a validação técnica dos algoritmos centrais e o TRL 6 para a pseudonimização e a segurança dos bancos de dados apartados, sem que essa manutenção represente comprovação de superação das pendências;

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

- Realismo experimental e benefícios esperados: análise das evidências da Fase B e do acompanhamento realizado no Ciclo 5, preservando as conclusões e os limites dos ensaios técnicos encerrados no Ciclo 3, sem atribuir ao ciclo atual novos resultados desses testes. Permanecem relevantes a verificação da resiliência em condições reais e os testes de resistência à manipulação de coordenadas;
- Riscos, mitigação e impacto residual: acompanhamento das avaliações de riscos da subfase B2, das medidas de mitigação e das condicionantes. O parecer do Ciclo 5 manteve a classificação "suficiente com condicionantes", com risco inerente alto e residual agregado médio, devendo esse resultado ser considerado em conjunto com os riscos individuais ainda elevados, especialmente o de reidentificação pela combinação de geolocalização e atributos de acessibilidade;
- Governança, transparência e aprendizagem regulatória: análise da documentação, dos papéis e responsabilidades, da explicabilidade, da matriz de transparência e das fichas de êxitos. A manutenção dos 78 pontos no *framework* geral e a classificação do êxito como parcial reforçam a necessidade de verificar a aplicação prática das salvaguardas, especialmente quanto às informações apresentadas aos usuários, à acessibilidade, à governança de terceiros e à operacionalização da descontinuidade.

Considerando o acompanhamento da Synapse nas fases A, B e C do Ciclo 5, 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 5 deste sandbox regulatório:

- Fase A:
  - Fichas A1 e A2 — Aplicação e Matriz TRL;
  - Fichas A3 e A4;
  - 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:
  - C1 — Êxitos para a ANPD e para a empresa participante; e
  - Plano de descontinuidade.

4.1.2 Fase A

Em atenção às fichas técnicas de aplicação (A1 e A2), sob responsabilidade da empresa, quanto à arquitetura do modelo, a Synapse informou que, entre os Ciclos 4 e 5, prosseguiu com os testes de diferentes tipos de mapas geolocalizados e de formas de tornar mais inteligente o processo de mapeamento e criação desses mapas, esclarecendo que sua implementação é necessária para viabilizar as recomendações de rotas internas, finalidade central do projeto Trajetto. Esse registro dá continuidade ao desenvolvimento informado no ciclo anterior, sem demonstrar, por si só, a disponibilização definitiva da funcionalidade.

Quanto ao fluxo de dados e às *Application Programming Interface* (API), a participante declarou não ter havido mudanças entre os dois ciclos. Nos esclarecimentos constantes da ficha, a empresa informou que as solicitações de rota não são processadas quando há problemas de comunicação ou ausência de dados, sendo exibida mensagem para nova tentativa. Também declarou que a condição de pessoa com deficiência é utilizada como variável no cálculo de rotas, associada à sessão, e que o aplicativo não coleta nem utiliza preferências pessoais.

Quanto às imagens, descreveu uma solução de análise ponto a ponto e indicou, como estratégia definitiva, manter o modelo no ambiente do parceiro e receber apenas pontos geolocalizados. Essas declarações ainda demandam verificação quanto à implementação e à efetividade das medidas descritas, especialmente em relação aos riscos adicionais dos mapas geolocalizados, à possibilidade de identificação indireta pela combinação de localização e atributos de acessibilidade e aos fluxos de dados das API e de terceiros. O documento técnico completo sobre o tratamento das imagens permanece indicado como em construção.

Da mesma maneira, a empresa declarou que os protocolos de segurança (art. 46 da Lei Geral de Proteção de Dados, LGPD), não sofreram alterações entre os Ciclos 4 e 5. A autenticação com fator adicional permanece descrita como funcionalidade da nova versão em implementação, com previsão de confirmação por SMS[10] e possibilidades de cadastro por contas em serviços de autenticação, além do cadastro diretamente no aplicativo.

Ainda, a ficha informa que o usuário permanecerá conectado após o *login*, podendo encerrar a sessão pelas configurações, sem comprovar a implementação de encerramento automático. Permanecem pendentes, ademais, o detalhamento das cópias de segurança e a definição do prazo de guarda das rotas salvas, pois a justificativa de utilização dos dados históricos para projeções e aprimoramento das recomendações não especifica seu período de conservação.

Em relação à Ficha A2, referente à avaliação da maturidade tecnológica e ao mapeamento dos Elementos Tecnológicos Críticos, o quadro preenchido pela participante mantém, no Ciclo 5, o TRL 4 para a validação técnica dos algoritmos centrais e o TRL 6 para a pseudonimização e a segurança dos bancos de dados apartados, sem alteração em relação ao Ciclo 4. Esses níveis correspondem à autoavaliação apresentada pela empresa e devem ser considerados em conjunto com as evidências produzidas nos testes, sem que

[10] SMS — *Short Message Service* ou Serviço de Mensagens Curtas: trata-se do sistema de envio e recebimento de mensagens de texto entre telefones celulares, em geral, utilizando a infraestrutura de redes das operadoras de telefonia.

sua manutenção permita concluir que as pendências de implementação e comprovação foram superadas.

No que tange às Fichas A3 e A4, a análise regulatória e documental do Ciclo 5 manteve as conclusões e recomendações registradas ao final do Ciclo 4, sem identificar alterações estruturais nos instrumentos de governança anteriormente apresentados pela Synapse. O acompanhamento concentrou-se nos esclarecimentos relativos ao tratamento de imagens pelas API, às atualizações desses componentes e ao risco de identificação indireta decorrente da combinação de geolocalização com atributos de acessibilidade ou vulnerabilidade. As respostas registradas foram parciais e não permitiram considerar superadas as pendências de comprovação. Na avaliação pelo *framework* geral, permaneceram os 78 pontos alcançados no Ciclo 4, sem nova elevação no Ciclo 5. Destacam-se os seguintes aspectos:

- A análise regulatória da equipe do CIAAM/USP preservou o diagnóstico anterior sobre governança, proteção de dados, direitos humanos, segurança da informação e inteligência artificial. O Ciclo 5 também destacou a discussão sobre o enquadramento das operações como tratamento de alto risco e seus efeitos sobre o regime aplicável aos agentes de tratamento de pequeno porte. Esse enquadramento deve ser fundamentado nos critérios cumulativos do art. 4º da Resolução CD/ANPD nº 2/2022, considerando as restrições ao tratamento jurídico diferenciado previstas no art. 3º. A Resolução CD/ANPD nº 15/2024 constitui referência específica para a comunicação de incidentes de segurança.
- 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. A entrega desses instrumentos nos ciclos anteriores ampliou o lastro documental, mas o Ciclo 5 não registrou evidências suficientes para considerar comprovada a implementação de todas as 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 continuam sujeitos às pendências anteriormente identificadas. O fluxo de dados e o protocolo de explicabilidade permanecem parcialmente descritos, o inventário de API e terceiros segue insuficientemente comprovado e a arquitetura de pseudonimização e criptografia ainda demanda auditoria documental. Embora tenha sido entregue relatório de robustez e estresse, permanece pendente o teste de combate à falsificação de identidade.
- 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, sem registro de alterações no Ciclo 5 que afastem essas observações.

- 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, especialmente nas atividades que envolvam câmeras, infraestrutura de parceiros e integrações externas.
- 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. Os critérios de progressão entre ciclos e os mecanismos de coordenação institucional devem ser acompanhados de registros que permitam verificar sua aplicação.
- 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 ficha não registra, no Ciclo 5, alterações que permitam considerar atendidas essas recomendações.
- A experiência do usuário continua a exigir aperfeiçoamentos de acessibilidade e transparência. O Ciclo 5 manteve a lacuna probatória quanto às informações efetivamente exibidas aos diferentes grupos de usuários, especialmente os textos, avisos, notificações e explicações apresentados durante a utilização do aplicativo.
- 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. A limitação anteriormente informada quanto ao atendimento de usuários com deficiência visual deve continuar explicitada e acompanhada de planejamento formal para sua revisão, sem registro de superação dessa pendência no Ciclo 5.
- 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. No Ciclo 5, permaneceu em aberto o questionamento sobre a combinação de geolocalização com atributos de acessibilidade ou vulnerabilidade, especialmente quando o destino corresponde a locais restritos com listas públicas de associados. A ficha classificou esse ponto como de alto risco e registrou a

necessidade de posicionamento formal da empresa sobre as medidas de anonimização ou agregação aplicadas, acompanhado de comprovação de sua efetividade.

- 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. O Ciclo 5 não registrou comprovação suficiente dos textos e das telas efetivamente utilizados para essas finalidades.
- A AIDH mantém a identificação de riscos relacionados à opacidade algorítmica e aos impactos sobre grupos vulneráveis. A manutenção dos 78 pontos na avaliação do *framework* geral não afasta a necessidade de demonstrar transparência acessível, indicar obstáculos nos trajetos e monitorar continuamente a precisão das recomendações.
- O RIPD permanece como referência da avaliação, devendo 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. Quando essas informações constituírem dados referentes à saúde, seu tratamento estará sujeito às hipóteses do art. 11 da LGPD, não podendo ser fundamentado no legítimo interesse. A oferta de rotas acessíveis a todos os usuários permanece recomendada 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 conflitos de interesse, conforme a Resolução CD/ANPD nº 18/2024. O Ciclo 5 não registra alteração que afaste a necessidade de verificar sua atuação efetiva e a adequação da estrutura institucional 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. Permanecem necessárias a identificação dos responsáveis e a apresentação de registros e evidências de aplicação das medidas declaradas.
- Os *model cards* e os mecanismos de explicação devem informar uso pretendido, critérios principais de recomendação, métricas de desempenho, limitações e cenários de falha. No Ciclo 5, permaneceu a necessidade de esclarecer as informações prestadas ao usuário diante de atualizações relevantes das API, especialmente das câmeras. A resposta parcial sobre processamento de imagens no dispositivo não esclareceu a frequência esperada dessas atualizações

nem demonstrou como seria assegurada e verificada a não identificação de pessoas em novas versões.

- 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, destino e atributos de acessibilidade. Na seção específica do Ciclo 5 das Fichas A3 e A4, o questionamento sobre reidentificação permaneceu sem resposta conclusiva, não sendo possível considerar comprovada a eliminação desse risco.
- 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. Continua necessária a comprovação da execução dos procedimentos previstos, incluindo as responsabilidades dos parceiros nos diferentes pontos do fluxo de tratamento.
- 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, a transparência na interface, as bases legais aplicáveis, o controle dos dados compartilhados com terceiros e a implementação das medidas de governança. O Ciclo 5 manteve as conclusões anteriores, sem registrar evidências suficientes para considerar superadas essas pendências.

A análise individual dos documentos regulatórios apresentados pela Synapse permanece como referência para o acompanhamento do Ciclo 5. Os documentos de natureza predominantemente técnica possuem avaliação própria, devendo o relatório distinguir os elementos já examinados das evidências ainda pendentes, especialmente quanto a:

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

Quanto aos resultados da equipe técnica sintetizados nas Fichas A3 e A4, o projeto Trajetto consolidou a transição de uma lacuna documental para uma lacuna probatória de implementação prática, resultando na emissão de parecer favorável com condicionantes. Embora a equipe técnica tenha validado avanços arquiteturais relevantes de *Privacy by Design*, como o processamento de visão computacional descentralizado na borda[11] (*on-device*, sem tráfego de imagens), a proteção criptográfica, a autenticação em dois fatores e a entrega do RIPD e da AIDH, persistiram pendências críticas de comprovação material.

[11] 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.

Destacam-se a ausência de demonstração visual de *mockups*[12], avisos e textos de explicabilidade no aplicativo (arts. 9º e 20 da LGPD) e a manutenção do Risco R5 (Alto) de reidentificação indireta de usuários com deficiência pelo cruzamento de geolocalização contínua com a indicação de rota acessível, *o* que exige protocolo formal de agregação e ofuscação, com referência à privacidade diferencial. Ademais, o parecer considerou que o tratamento contínuo de geolocalização com atributos potencialmente sensíveis indica enquadramento no regime de alto risco, afastando as flexibilizações para Agentes de Pequeno Porte (Resolução ANPD nº 02/2022) e exigindo o cumprimento integral do Regulamento de Segurança da Resolução CD/ANPD nº 15/2024. O Ciclo 5 encerra-se, portanto, com manutenção das conclusões anteriores e das pendências de esclarecimento e comprovação. A continuidade experimental permanece condicionada à verificação das salvaguardas e não equivale à aprovação da solução.

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 5 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 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. A pontuação não se confunde com pré-requisito para continuidade da participação neste sandbox. Pois seu objetivo avaliar o projeto em relação aos eixos tratados no *framework* geral e orientar as ações de melhoria para o projeto. 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 5, a Synapse manteve os 78 pontos alcançados no Ciclo 4, sem alteração nas respostas aos critérios pontuados. Permaneceram os seguintes resultados:

- Eixo 1 (Centralidade da dignidade humana e dos direitos fundamentais), 20 pontos;
- Eixo 2 (Governança e responsabilidade clara), 8 pontos;
- Eixo 3 (Avaliação de impactos), 15 pontos;
- Eixo 4 (Justiça, não discriminação e vieses algorítmicos), sem pontuação;
- Eixo 5 (Transparência, explicabilidade e comunicação), 15 pontos;
- Eixo 6 (Segurança técnica e robustez), 10 pontos; e
- Eixo 7 (proporcionalidade e necessidade), 10 pontos.

[12] *Mockups* — representações visuais de telas ou interfaces, estáticas ou interativas, utilizadas para ilustrar a disposição das informações, os comandos e a experiência de uso prevista. Permitem examinar, por exemplo, como avisos de privacidade, explicações e opções de escolha serão apresentados aos usuários, sem, por si sós, comprovar a implementação ou o funcionamento dessas funcionalidades

Em suma, o resultado mantém o enquadramento registrado no parecer como “adequado, mas com pontos críticos a melhorar”.

Além disso, a manutenção da pontuação deve ser interpretada em conjunto com as ressalvas constantes da planilha, pois:

- permanecem em desenvolvimento os mecanismos de prestação de contas e a demonstração da capacidade real de intervenção humana;
- a identificação dos grupos vulneráveis ainda requer aprofundamento, e os documentos de avaliação de impactos e mitigação mantêm observações a atender;
- o Eixo 4 continua sem pontuação, com pendências relativas à comprovação dos testes, às limitações demográficas, à inclusão de usuários com deficiência visual e às salvaguardas contra vieses e reidentificação.

Dessa forma, embora os critérios de segurança técnica permaneçam pontuados, as observações distinguem os resultados obtidos com os conjuntos de teste utilizados, incluindo imagens reais coletadas na USP, em ambiente distinto daquele de aplicação da solução, da comprovação ainda necessária em condições representativas das estações de metrô e dos testes contra falsificação de informações pendentes. Assim, o escore não permite concluir que essas limitações foram superadas.

B) Testes da equipe técnica do CIAAM/USP

Os testes da equipe técnica descritos no relatório foram encerrados no Ciclo 3, permanecendo válidas suas conclusões nos limites e nas condições então documentados. Os materiais examinados para o Ciclo 5 não apresentam nova rodada desses quatro ensaios, razão pela qual seus resultados não devem ser atribuídos a este ciclo.

O parecer técnico do Ciclo 5 incorpora esses antecedentes, especialmente a verificação do uso de protocolo de segurança na camada de transporte no percurso auditado por túnel, e aprofunda o acompanhamento dos controles e riscos remanescentes. Mantém, contudo, a necessidade de comprovação da resiliência em dados e redes reais e de apresentação dos relatórios contra falsificação de informações para o modelo geolocalizado. Os resultados favoráveis obtidos com dados sintéticos não demonstram, isoladamente, resistência à manipulação de coordenadas ou segurança em todas as condições operacionais.

C) Matriz de transparência

A matriz de transparência do Ciclo 5 registra, para todos os grupos de interlocutores, a manutenção das respostas apresentadas no Ciclo 4. A avaliação correspondente confirma a persistência das principais lacunas de conteúdo, momento de comunicação e demonstração da interface. A indicação de preenchimento concluído não equivale à comprovação de atendimento das recomendações.

Para usuários gerais, cadeirantes e pessoas com mobilidade reduzida, permanece a necessidade de explicar por que são prestadas informações sobre o uso de IA, os dados utilizados e os critérios de recomendação. A indicação de que a explicação aparecerá no aplicativo a cada escolha de rota identifica o canal pretendido, mas ainda não demonstra

as telas, os textos exatos e os momentos de apresentação dos avisos. Quanto aos usuários com deficiência visual, manteve-se a exclusão temporária do escopo da versão em desenvolvimento, sem registro de atendimento à recomendação de formalizar sua justificativa e o cronograma de inclusão.

Para os operadores, permanece a previsão de painéis com informações agregadas sobre fluxos e pontos geolocalizados, mas faltam esclarecimentos sobre o funcionamento dos modelos preditivos, a periodicidade dos relatórios e os meios de envio dos alertas. Para os órgãos reguladores, continuam pendentes a especificação dos conteúdos, sua relação com as necessidades de fiscalização, a periodicidade e os meios de disponibilização.

Quanto à equipe técnica, foram mantidas as descrições de perfis de acesso, permissões, armazenamento de registros em provedor de serviços em nuvem e no banco de dados, coleta diária de *logs* e revisão mensal pelo responsável técnico. A avaliação preserva o reconhecimento de adequação dessas especificações para a governança interna, sem que isso supra as pendências de transparência perante usuários, operadores e reguladores.

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 05, emitido em outubro de 2026, manteve a classificação "suficiente com condicionantes" e recomendou o prosseguimento dos testes acompanhado de complementações. A avaliação utiliza matriz de probabilidade e impacto e registra risco inerente agregado alto e risco residual agregado médio.

A classificação residual agregada deve ser lida com as avaliações individuais: permanecem classificados como altos os riscos relacionados ao comprometimento do serviço de interligação entre os cofres de dados, à manipulação de coordenadas, às integrações com terceiros e à reidentificação por atributos de acessibilidade. Portanto, o resultado agregado não representa a superação desses pontos críticos.

Para sistematizar os indicadores, apresenta-se o quadro atualizado:

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: a contagem e a análise de fluxo por câmeras ocorrerão na borda do parceiro (*on-device* ou em servidor local), enviando ao Trajetto exclusivamente as coordenadas, eliminando o tráfego e o armazenamento de vídeos e rostos nos servidores centrais. • Autenticação e acesso: registro da entrega de *login* via serviços de autenticação e de duplo fator (2FA) com confirmação via SMS • Guarda de histórico: registro da justificativa de negócio para o armazenamento do histórico de rotas (previsão de fluxos), observando-se, quando aplicável, o prazo mínimo de guarda de seis meses previsto no Marco Civil da Internet (MCI) para os registros de acesso a aplicações.
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.

- Resiliência em dados reais: os testes de estresse funcionaram para dados sintéticos, mas pendem de comprovação em dados reais e relatórios de testes contra falsificação de informações.

- Enquadramento Normativo (Resolução CD/ANPD nº 15/2024): devido ao tratamento de alto risco, a Synapse não se beneficia do tratamento diferenciado do Art. 4º da Resolução CD/ANPD nº 02/2022 (Agentes de Pequeno Porte), sujeitando-se integralmente aos requisitos de segurança da Resolução CD/ANPD nº 15/2024.

Fonte: ANPD

O projeto Trajetto mantém os avanços arquiteturais reconhecidos nos ciclos anteriores, especialmente a segregação de ambientes e o emprego combinado de pseudonimização e criptografia. O parecer do Ciclo 5 registra o processamento de imagens na borda dos parceiros, com transmissão de coordenadas ao sistema central, e reconhece avanços na autenticação. Entretanto, permanece necessária a comprovação técnica da não retenção de imagens em *cache* ou memória local. A implementação da autenticação também deve ser conciliada com a Ficha A1, que a descreve como funcionalidade em desenvolvimento, mediante identificação da versão e das evidências verificadas.

A avaliação mantém a lacuna probatória de implementação prática, destacando:

- transparência material ao titular: continuam pendentes as telas, os avisos, os fluxos de consentimento e os textos exatos de explicabilidade apresentados no aplicativo.
- Reidentificação por atributos de acessibilidade: o parecer examinou a resposta sobre uso da variável de rota acessível e renovação do identificador de sessão, considerando-a insuficiente para afastar o risco. O ponto permaneceu classificado como alto, e foi solicitada a

apresentação de protocolo formal de agregação e ofuscação, com referência à privacidade diferencial.

- Memorial técnico e terceiros: permanecem necessários o detalhamento do processamento de imagens nos parceiros e o inventário auditável das integrações com Google Maps, câmeras, provedores de *login* e serviços de SMS, incluindo dados compartilhados, responsabilidades e fluxos internacionais.
- Resiliência e governança: continuam pendentes os testes contra falsificação de informações e a comprovação da resiliência em condições reais. Também permanecem as recomendações de atuação institucional do encarregado de dados e de formalização do cronograma para inclusão de usuários com deficiência visual.
- Retenção do histórico: o parecer registra a justificativa de utilização das rotas para previsão de fluxos, mas ainda é necessário definir categorias de dados, finalidades e prazos de conservação. A referência a seis meses deve distinguir o histórico de deslocamentos dos registros de acesso a aplicações de internet: o art. 15 do Marco Civil estabelece obrigação de guarda destes últimos para os provedores abrangidos, e não um limite geral de conservação de históricos de rotas.

Quanto ao enquadramento normativo, o parecer considera elevado o risco decorrente da combinação de geolocalização e atributos de vulnerabilidade. A fundamentação deve observar os critérios cumulativos do art. 4º da Resolução CD/ANPD nº 2/2022 e a restrição ao tratamento jurídico diferenciado prevista no art. 3º. A Resolução CD/ANPD nº 15/2024 disciplina especificamente a comunicação de incidentes de segurança, devendo ser referida nesse âmbito.

No acompanhamento das condicionantes, houve a entrega do RIPD e da AIDH, preservando a recomendação de ajuste do RIPD quanto aos riscos de acessibilidade. No entanto, as demais condicionantes permanecem pendentes, abrangendo resiliência e imagens, integrações externas, governança do encarregado de dados, transparência material, mitigação da reidentificação e inclusão de usuários com deficiência visual.

Assim, a Fase B do Ciclo 5 consolidou o acompanhamento das salvaguardas e das respostas apresentadas, mantendo a pontuação ética e os níveis de maturidade tecnológica, sem demonstrar a superação das principais pendências materiais. A continuidade recomendada pelo parecer depende do acompanhamento e da verificação das condicionantes, não equivalendo à aprovação da solução nem à comprovação de controles ainda parcialmente implementados ou descritos documentalmente.

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.

No Ciclo 05, foram consideradas as fichas C1 disponibilizadas pela ANPD e pela Synapse, que apresentam conteúdo coincidente nos campos relativos à autoridade, à participante e à conclusão conjunta, com classificação do êxito como parcial. A avaliação distingue os avanços acumulados nos ciclos anteriores, as atividades declaradas pela empresa no período e as medidas cuja implementação ou efetividade permanece pendente de comprovação.

A) Avaliação sob a Perspectiva da ANPD

Sob a perspectiva da ANPD, o Ciclo 5 do Trajetto consolidou os aprendizados anteriores, deslocando o foco da apresentação documental para a comprovação material dos controles. A ficha registra a preservação de medidas como segregação entre identidade e telemetria, pseudonimização, criptografia, protocolo de segurança na camada de transporte, autenticação de dois fatores e processamento de imagens na borda. Como evidências, foram indicados o Parecer Técnico Consolidado do Ciclo 5, as fichas A1, A2, A3 e A4, a Matriz de Transparência e o *framework* geral. A avaliação permaneceu "suficiente com condicionantes", com risco inerente alto e residual médio, sem alteração da pontuação do *framework* geral (78/100) ou dos níveis de maturidade tecnológica dos componentes avaliados, mantidos em TRL 4 e 6.

O principal aprendizado regulatório foi que a renovação do identificador de sessão e a utilização de uma sinalização de rota acessível não afastam, isoladamente, a possibilidade de reidentificação pelo cruzamento de geolocalização detalhada, destinos e atributos de acessibilidade. Também permaneceram pendentes evidências sobre transparência efetiva nas interfaces, não retenção de imagens na borda, governança das integrações com terceiros, robustez e contra falsificação de informações em condições representativas de uso e operacionalização do Plano de Descontinuidade. Dessa maneira, o êxito foi considerado parcial: os resultados oferecem subsídios para orientações e critérios de supervisão, mas não sustentam aprovação plena da solução, proposição conclusiva de nova norma ou demonstração de efeitos efetivos sobre o mercado e a concorrência.

B) Avaliação sob a perspectiva da empresa

Sob a perspectiva da Synapse, o Ciclo 5 concentrou-se na preparação do Trajetto para publicação, com aprimoramentos na experiência do usuário, ajustes nos mapas internos (*indoor*) e o desenvolvimento de uma abordagem inteligente para sua construção e geração. Como evidências, a empresa remeteu aos registros de evolução e às capturas de tela apresentados nos ciclos anteriores, além da disponibilização de um vídeo demonstrativo. Esses materiais foram indicados como suporte às melhorias acumuladas, embora a participante tenha informado que não houve avanços adicionais significativos em relação ao ciclo anterior.

Quanto à transparência, à explicabilidade e aos direitos dos titulares, a empresa não relatou novas evoluções, permanecendo em acompanhamento as recomendações sobre clareza dos critérios do sistema, comunicação de suas limitações e uso das informações de localização. Adicionalmente, declarou ter realizado ajustes documentais e prestado esclarecimentos relativos à governança de IA e dados, indicando como evidência os documentos disponibilizados na pasta compartilhada. Dessa maneira, o balanço empresarial caracteriza o período como uma etapa de consolidação e preparação, sustentada pelos materiais demonstrativos e pelas entregas documentais

mencionadas, sem que seu envio comprove, isoladamente, a implementação ou a efetividade de todas as medidas recomendadas.

A partir das informações coletadas, verifica-se que o Ciclo 5 consolidou os aprendizados regulatórios e preservou os avanços arquiteturais reconhecidos nos ciclos anteriores, sem registrar evolução na pontuação do *framework* geral, mantida em 78 pontos, ou nos níveis de maturidade tecnológica dos componentes críticos, mantidos em TRL 4 e TRL 6. As duas fichas classificam o êxito como parcial e relacionam esse resultado à permanência da avaliação "suficiente com condicionantes" no parecer técnico de 13 de outubro de 2026, com risco inerente alto e residual agregado médio. Persistem lacunas de comprovação material, especialmente quanto à transparência e à acessibilidade na interface, à mitigação da reidentificação, à governança de terceiros, aos testes em condições reais e à operacionalização das medidas de descontinuidade. Os avanços mencionados nas fichas não devem ser interpretados como comprovação integral dos controles cuja implementação permanece sujeita a verificação.

Nesse contexto, a evolução do Trajetto deverá ser orientada pela apresentação e validação progressiva de evidências. Devem ser priorizados os registros visuais e funcionais das mensagens, telas, avisos e explicações acessíveis ao titular; o protocolo de avaliação e mitigação da reidentificação por cruzamento de localização e atributos de acessibilidade, incluindo minimização e testes de inferência; o memorial técnico e os testes de não retenção de imagens na borda; os testes de robustez e contra falsificação de informações com dados e redes reais; o inventário técnico e contratual de API e terceiros, abrangendo Google Maps, provedores de autenticação, SMS e câmeras; a demonstração da atuação do encarregado de dados; o planejamento para inclusão de pessoas com deficiência visual; e a revisão do RIPD quanto aos riscos de acessibilidade. Além disso, deverá ser assegurada a correspondência entre as versões submetidas aos testes e as disponibilizadas para utilização. A continuidade experimental deve observar o acompanhamento das condicionantes do parecer, sem equivaler à aprovação plena ou à autorização comercial da solução.

Finalmente, quanto ao Plano de Descontinuidade, as Fichas C1 do Ciclo 5 registra a necessidade de aprimorar sua operacionalização, com definição de responsáveis, realização de simulações e apresentação de evidências verificáveis. Esses instrumentos não apresentam comparação entre versões do plano nem registros suficientes para confirmar a execução e os resultados de procedimentos de suspensão, bloqueio ou eliminação neste ciclo. Assim, não é possível afirmar, com fundamento apenas nas fichas examinadas, quais alterações foram incorporadas ao plano entre os Ciclos 4 e 5 ou se as pendências anteriores foram superadas. Permanece necessária a apresentação dos registros de acionamento e execução, com identificação das responsabilidades e dos resultados, para subsidiar a avaliação da efetividade das medidas previstas.

4.2 Prevvine

O relatório do CIAAM/USP referente ao Ciclo 5 de testagem da solução STAiDOC, apresentada pela Prevvine, consolida as informações, os registros e as evidências das fases A e B, bem como as avaliações apresentadas separadamente pela ANPD e pela participante na 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 5, 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, distinguindo as retificações documentais, as alterações de interface e as funcionalidades previstas para a próxima versão.
- Dados, privacidade e segurança: análise das informações e evidências das fases A e B, incluindo o tratamento de dados pessoais, a supressão de identificadores, os controles de acesso, os fluxos internacionais e a efetividade das salvaguardas, considerando as limitações verificadas na testagem.
- Maturidade tecnológica (TRL e ETC): consideração da autoavaliação apresentada pela Prevvine nas fichas A1 e A2 e das evidências técnicas disponíveis, observando a correspondência entre a matriz consolidada, o detalhamento dos componentes e os resultados dos testes.
- Realismo experimental e benefícios esperados: avaliação das interações clínicas examinadas na plataforma HarpIA, considerando a qualidade das respostas, a fundamentação bibliográfica, a adequação aos diferentes perfis de pacientes e as limitações metodológicas. A interpretação contempla as condições de integração e as diferenças descritas entre os canais clínico e de avaliação.
- Riscos, mitigação e impacto residual: análise do parecer técnico, dos riscos identificados, das condicionantes e das pendências de comprovação. As classificações agregadas devem ser examinadas em conjunto com os riscos individuais e os resultados experimentais, preservando-se as divergências entre os instrumentos de avaliação.
- Governança, transparência e aprendizagem regulatória: análise da documentação institucional, dos papéis e responsabilidades, da supervisão humana, da avaliação ética e da matriz de transparência, que passou a concentrar o acompanhamento da explicabilidade por destinatário. Esses elementos são articulados aos aprendizados registrados pela ANPD e pela empresa, distinguindo medidas documentadas, controles operacionais e aperfeiçoamentos ainda planejados.

Considerando o acompanhamento do Ciclo 5 nas Fases A — Pré-teste; B — Teste; e C — Pós-teste, apresentam-se, a seguir, os principais aspectos de cada etapa. Na Fase C, a consolidação considera as fichas preenchidas separadamente pela ANPD e pela Prevvine, permanecendo sem preenchimento a classificação do êxito global, a principal evidência de êxito e os próximos passos na seção destinada à conclusão conjunta.

4.2.1 Documentação principal

Preliminarmente, observa-se que os principais documentos referidos neste item do relatório encontram-se anexos e organizados conforme a descrição a seguir por fases para atendimento dos requisitos pertinentes ao Ciclo 5 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
  - Parecer técnico consolidado
  - Relatório de Análise — Plataforma HarpIA
- Fase C
  - Ficha C1 — Êxitos para a ANPD e para a empresa participante; e
  - Plano de descontinuidade.

4.2.2 Fase A

Inicialmente, no que se refere à Ficha A1, a participante Prevvine informou que, entre os Ciclos 4 e 5, foram realizadas alterações de interface e retificações da descrição técnica da solução STAiDOC, sem alteração do modelo de inteligência artificial, das instruções do sistema ou da sequência de processamento. Na interface, foram registrados: aviso fixo sobre o caráter de suporte à decisão, botões de confirmação ou rejeição por resposta clínica, alerta destacado de urgência, aviso preventivo de privacidade e canais de suporte e de contato com o encarregado de dados. Além disso, foi esclarecido que a autenticação ocorre por e-mail e senha, com armazenamento da senha mediante *hash*, em substituição à descrição anterior de autenticação via provedor de serviços em nuvem. O processamento de imagens permanece desativado, embora ainda exista ícone de anexo na interface, cuja remoção foi prevista até o encerramento do Ciclo 6.

Quanto ao fluxo de dados, a ficha mantém a execução do filtro de supressão de identificadores antes do envio de cada mensagem ao modelo, mas retifica a descrição de sua cobertura. A participante reconhece que a versão em avaliação não possui detector específico de data de nascimento e apresenta limitações no reconhecimento de nomes, incluindo campos de filiação e de terceiros em formato cadastral, grafias irregulares, nomes sem indicador e prenomes isolados. Nessas situações, os identificadores podem seguir ao modelo. O reconhecimento foi descrito como baseado em indicadores de identificação e estruturas típicas de nomes brasileiros com partícula, associados a listas de exclusão e padrões de contexto clínico. Adicionalmente, foram atualizados o marcador de substituição, para “(dado sensível removido)”, e os prazos de retenção: duração do Sandbox acrescida de um ano para os dados clínicos tratados e os registros de explicabilidade; e cinco anos para os registros de auditoria, supressão de identificadores e incidentes.

Em relação às API e aos serviços integrados, não foi informada alteração da API de inteligência artificial ou de sua versão. A ficha acrescenta precisões sobre a verificação bibliográfica, descrevendo tempo-limite de oito segundos por chamada ao PubMed, *cache* de verificações por trinta dias e mecanismo de contingência que apresenta referências não confirmadas com o rótulo “busca PubMed, verificação manual”, com remoção das referências inexistentes. Também foi incluído no inventário de operadores o serviço de e-mail transacional Resend, localizado nos Estados Unidos, com tratamento restrito ao e-mail do médico e ao conteúdo das notificações, sem dados clínicos, conforme declarado pela participante.

No âmbito dos protocolos de segurança, a ficha registra os resultados de testes realizados em produção em outubro de 2026, com a aprovação dos sete itens do Teste A. Entre os controles descritos estão o limite de cinquenta mil caracteres por mensagem, o limite de quinze mensagens por minuto por usuário, a recusa de credenciais inválidas, a detecção e o registro de tentativas de manipulação de instruções e o uso de mensagens de erro genéricas. A detecção dessas tentativas, contudo, não implica bloqueio automático: a mensagem, após passar pelo filtro, segue ao modelo. Foram ainda descritos registros automáticos de incidentes e segregação de acesso por usuário. Permanecem pendentes o *firewall* de aplicação dedicado com regras do *Open Worldwide Application Security Project* (OWASP) e melhorias na proteção criptográfica, reprogramados para a próxima versão identificada, até o encerramento do Ciclo 6. O exercício simulado de incidente de severidade alta e o teste de suspensão, bloqueio e eliminação também foram reprogramados, com previsão de registro em adendo ao relatório de testes.

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

- motor de inteligência artificial;
- 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 com a LGPD.

Quanto à classificação de maturidade tecnológica, a matriz atualizada apresenta preenchimento para os Ciclos 4 e 5, mantendo TRL 7 para o motor de inteligência artificial e para a infraestrutura de conformidade com a LGPD, e TRL 6 para a camada de supressão de identificadores e para o sistema NTAV. Segundo a avaliação registrada pela Prevvine, a camada de supressão passou para o TRL 6 a partir do Ciclo 4, em razão das limitações demonstradas nos testes de privacidade e da ausência de medição das métricas-alvo de precisão, sensibilidade e índice F1 em um *corpus* sintético. O detalhamento individual dos ETC, entretanto, ainda atribui TRL 7 à camada de supressão e ao NTAV, devendo ser harmonizado com a matriz consolidada e com as observações atualizadas da própria ficha.

Em relação ao NTAV, permanece o histórico de redução para TRL 6 ao final do Ciclo 3, decorrente de alucinações, combinação de informações reais e inventadas, *links* quebrados e referências inexistentes. Para os ciclos 4 e 5, a participante registra a operação do mecanismo de contingência por degradação controlada, com *cache* e sinalização de referências não verificadas, mas reconhece a persistência de erros bibliográficos nas avaliações. Assim, os mecanismos descritos não fundamentam, por si sós, a recuperação do nível de maturidade desse componente, permanecendo necessária a comprovação de sua efetividade.

Quanto à qualificação da técnica de proteção de dados, as retificações do Ciclo 5 reforçam a necessidade de descrevê-la como supressão de identificadores, sem presumir anonimização integral do conteúdo clínico. A própria ficha reconhece situações nas quais identificadores não são detectados e seguem ao modelo. O descarte declarado do texto original e a manutenção de *hashes* não demonstram, isoladamente, a impossibilidade de identificação direta ou indireta no conteúdo remanescente. Tampouco esses elementos permitem concluir automaticamente pela pseudonimização, permanecendo necessária a avaliação das possibilidades de associação ao titular e da existência de informações adicionais que permitam restabelecê-la.

Diante do exposto, o Ciclo 5 apresenta maior precisão na caracterização técnica da STAiDOC, com retificações documentais, alterações de interface, registro de controles de segurança testados e reconhecimento expresso de limitações. A matriz atualizada mantém os níveis de maturidade dos quatro ETC entre os ciclos 4 e 5, sem indicar elevação nesse intervalo. Permanecem necessárias a harmonização dos campos desatualizados, a medição da efetividade do filtro de identificadores, a verificação do desempenho bibliográfico e a comprovação das medidas reprogramadas. A avaliação deve, portanto, distinguir as funcionalidades da versão preservada, as alterações de interface, os resultados de testes informados e os aperfeiçoamentos previstos para a próxima versão, sem antecipar o atendimento das pendências.

Já em relação às fichas A3 e A4 da Prevvine, a análise do Ciclo 5 abrangeu o marco regulatório aplicável à solução STAiDOC, as análises documentais e os resultados da avaliação ética, com destaque para proteção de dados, não discriminação, segurança técnica, transparência e efetividade da intervenção humana. Os registros permitem distinguir o aprofundamento da análise normativa e documental das limitações verificadas na testagem, que repercutiram negativamente na avaliação da solução.

No quadro normativo identificado como referente ao Ciclo 5, foram relacionados os seguintes diplomas:

- Resolução CFM[13] nº 2.299/2021, relativa à emissão de documentos médicos eletrônicos;
- RDC ANVISA[14] nº 36/2013, referente à segurança do paciente em serviços de saúde;
- Portaria GM/MS[15] nº 1.768/2021, relativa à Política Nacional de Informação e Informática em Saúde;
- Lei nº 13.709/2018, Lei Geral de Proteção de Dados Pessoais;
- Lei nº 9.656/1998, referente aos planos e seguros de saúde;
- Lei nº 14.510/2022, relativa à telessaúde;
- Portaria GM/MS nº 1.434/2020, referente ao Programa Conecte SUS[16] e à Rede Nacional de Dados em Saúde; e
- Lei nº 13.146/2015, Lei Brasileira de Inclusão da Pessoa com Deficiência.

No âmbito dos direitos fundamentais, a análise do Ciclo 5 destacou o princípio da não discriminação, a revisão de decisões tomadas unicamente com base em tratamento

[13] CFM — Conselho Federal de Medicina.
[14] ANVISA — Agência Nacional de Vigilância Sanitária.
[15] GM/MS — Gabinete do Ministro da Saúde.
[16] SUS — Sistema Único de Saúde.

automatizado e a possibilidade de auditoria de aspectos discriminatórios pela ANPD. Também foram examinadas as restrições ao uso de dados de saúde para seleção de riscos e à discriminação por idade ou deficiência no acesso a planos de saúde. A acessibilidade foi tratada como condição de igualdade, enquanto a equidade e a consideração de raça/cor, nome social e identidade de gênero foram relacionadas à avaliação de vieses. Quanto à telessaúde, destacou-se o direito de recusa dessa modalidade, com garantia de atendimento presencial. No campo das relações de consumo, o registro do Ciclo 5 informa que não houve previsão de avaliação de novos diplomas específicos.

No exame documental, foram produzidas duas análises complementares: uma sobre as políticas de privacidade do LLM[17] utilizado, sob a perspectiva da LGPD, e outra sobre a atualização do Manual de Testagem de LLM/LMM[18] em Saúde, com base no respectivo *System Card*. As análises foram relacionadas à questão regulatória da revisão de decisões automatizadas, buscando examinar a correspondência entre as capacidades dos sistemas e as salvaguardas descritas em seus documentos.

A análise das políticas comparou seis versões, correspondentes a outubro de 2025, janeiro de 2026 e julho de 2026, em inglês e português. Foram registradas 27 divergências entre os idiomas, incluindo omissões, inconsistências nas datas de vigência e denominação incorreta da ANPD. Dos 24 requisitos examinados sob a perspectiva da LGPD, nove foram considerados não atendidos e onze parcialmente atendidos. Entre as lacunas apontadas estão a definição dos prazos de retenção, a identificação do encarregado de dados, os prazos de resposta aos titulares e a indicação da base legal para dados sensíveis. Esses resultados se referem às políticas examinadas, sem demonstrar, por si sós, as condições específicas aplicáveis à integração comercial utilizada pela STAiDOC.

Quanto às decisões automatizadas, a análise identificou que o texto das políticas permaneceu inalterado, apesar da ampliação das capacidades descritas para aferição de idade por biometria, perfilamento para recomendações, ações agênticas com efeitos externos e classificação automatizada de conteúdo com consequências contratuais. Permaneceram indicados para validação jurídica o limiar de aplicação do art. 20 da LGPD e a interpretação dos efeitos da ausência de informações sobre os critérios empregados.

A análise do *System Card* registrou avaliações relacionadas ao bem-estar dos usuários, às ciências da vida e à honestidade factual em diferentes idiomas, inclusive português, mas apontou ausência de testes clínico-hospitalares. O documento registrou melhorias em viés e redução de recusas indevidas, regressão em biologia e insuficiência de evidências sobre privacidade em uso conversacional. Permaneceram pendentes a recaptura integral da versão em português da nova política, a confirmação dos identificadores de modelo e a verificação da necessidade de nova calibração do manual.

Em relação aos resultados da testagem expressamente atribuídos ao Ciclo 5, a avaliação registrou reprodução reiterada de dados inseridos nas interações, inclusive mediante ataques por *prompt*, indicando insuficiência dos filtros de proteção. Segundo a análise, a omissão de dados em determinadas respostas não assegurou sua exclusão do processamento. Também foram apontadas respostas inadequadas em situação envolvendo pessoa transgênero, consideração insuficiente da documentação médica brasileira de referência e persistência de erros bibliográficos. Esses achados foram acompanhados da constatação de que a interação dos operadores com o sistema e a

[17] LLM — *Large Language Model* ou Grande Modelo de Linguagem.
[18] LMM — *Large Multimodal Model* ou Grande Modelo Multimodal.

intervenção humana não ocorreram de forma efetiva, apesar das salvaguardas implementadas na plataforma.

A redução da pontuação da Prevvine de 75 pontos no Ciclo 4 para 43 no Ciclo 5 decorreu da incorporação dos resultados experimentais à avaliação, o que evidenciou limitações que a documentação apresentada não permitia identificar. Os testes registraram a reprodução de dados pessoais em ataques de *prompt*, respostas inadequadas a uma pessoa transgênero e falhas cuja justificativa não pôde ser reconstruída a partir dos registros disponibilizados pela empresa. Esses achados reduziram a pontuação do Eixo 1 (Centralidade da dignidade humana e dos direitos fundamentais) de 20 para 5 pontos, já que não houve pontuação no Eixo 4 (Justiça, não discriminação e vieses algorítmicos), o que repercutiu nos critérios de intervenção humana, de explicabilidade e de resiliência. Dessa maneira, a regressão do escore não demonstra, necessariamente, deterioração da solução entre os ciclos, mas uma revisão da avaliação diante de evidências sobre seu comportamento efetivo, que limitaram o reconhecimento anteriormente apoiado nos avanços documentais. O enquadramento na faixa "não adequado para sandbox ou regulação responsável" corresponde, portanto, à classificação do instrumento naquele ciclo, sem constituir, por si só, deliberação de exclusão do programa. Em resumo, esta alteração da pontuação permite-nos observar que é necessária a combinação da análise documental com a experimentação da solução para realizar a avaliação da aplicação de IA com base no *framework* geral.

Adicionalmente, cumpre destacar que o Eixo 1 recebeu cinco pontos; o Eixo 4, zero; e o Eixo 6 (Segurança técnica e robustez), dois de dez pontos. Também deixaram de ser considerados atendidos os critérios relativos à capacidade real de intervenção humana e à possibilidade de justificar decisões a *posteriori*. A classificação constitui resultado do instrumento de avaliação e deve ser apresentada com os achados que a fundamentaram.

O Eixo 3 (Avaliação de impactos) manteve a pontuação máxima, mas o registro do Ciclo 5 indicou a necessidade de atualizar a Avaliação de Impacto em Direitos Humanos, a identificação dos grupos vulneráveis e a matriz de riscos. Devem ser incorporados os riscos evidenciados na testagem, especialmente a reprodução de dados, o tratamento inadequado de minorias e os problemas de priorização de referências. Assim, o atendimento dos critérios documentais de avaliação de impactos não afasta a necessidade de revisão desses instrumentos diante dos resultados experimentais.

Em suma, os registros do Ciclo 5 aprofundaram a análise dos seguintes pontos:

- Correspondência entre capacidades dos modelos, informações das políticas e salvaguardas aplicáveis;
- Limitações de privacidade, não discriminação, qualidade bibliográfica e supervisão humana observadas na testagem; e
- Necessidade de atualizar os instrumentos de avaliação e mitigação diante dos riscos identificados.

Contudo, permaneceram pendentes a correção das falhas técnicas, o desenvolvimento de mecanismos de correção de vieses, o aperfeiçoamento da governança e da prestação de contas, a atualização das avaliações de impacto e a conclusão das verificações documentais indicadas.

Como encaminhamentos da avaliação do Ciclo 5, foram indicados:

- Corrigir os aspectos de segurança e programação apontados na testagem.
- Desenvolver, com prioridade, mecanismos de correção de vieses.

- Sanar as lacunas de governança de inteligência artificial antes da disponibilização ampla da solução.
- Aprimorar a documentação de prestação de contas sobre o uso e o processamento de dados pessoais.
- Atualizar a Avaliação de Impacto em Direitos Humanos, a identificação de grupos vulneráveis e o plano de mitigação, incorporando os riscos evidenciados nos testes.

Dessa forma, o Ciclo 5 demonstrou a necessidade de confrontar a documentação e as salvaguardas declaradas com o comportamento efetivamente observado. Conforme a avaliação registrada, as falhas identificadas devem ser corrigidas antes da utilização ampla e plena da solução.

4.2.3 Fase B

Inicialmente, a Fase B do Ciclo 5 reuniu a aplicação do *framework* geral, as avaliações conversacionais com vinhetas clínicas na plataforma HarpIA, o acompanhamento da explicabilidade pela matriz de transparência e a consolidação dos riscos e das medidas de mitigação em parecer técnico. Os documentos registram avanços na apresentação de evidências e na operacionalização de controles, acompanhados de falhas clínicas, bibliográficas e de proteção de dados. Sua análise conjunta exige distinguir os resultados de cada instrumento, as condições de execução dos testes e as medidas ainda dependentes de implementação ou comprovação.

A) Aplicação do framework geral baseado em recomendações da UNESCO

A Prevvine obteve 43 pontos de 100 no Ciclo 5. A distribuição por eixo foi: Eixo 1 (Centralidade da dignidade humana e dos direitos fundamentais), 5 pontos; Eixo 2 (Governança e responsabilidade), 4 pontos; Eixo 3 (Avaliação de impactos), 15 pontos; Eixo 4 (Justiça, não discriminação e vieses algorítmicos), sem pontuação; Eixo 5 (Transparência, explicabilidade e comunicação), 7 pontos; Eixo 6 (Segurança técnica e robustez), 2 pontos; e Eixo 7 (Proporcionalidade e necessidade), 10 pontos. Conforme a interpretação prevista na planilha, esse resultado corresponde à faixa “Não adequado para sandbox / regulação responsável”, no âmbito do instrumento de avaliação. Esse enquadramento sinaliza a necessidade de aperfeiçoamento das salvaguardas avaliadas, sem produzir efeito eliminatório automático. A permanência da Prevvine no ambiente experimental permite acompanhar o tratamento das fragilidades identificadas e verificar as medidas corretivas, mediante supervisão e observância das condicionantes estabelecidas. Dessa maneira, a pontuação orienta as prioridades de intervenção e de nova testagem no Ciclo 6.

Os registros do Ciclo 5 apontam insuficiência dos filtros diante da reprodução de dados nas interações, limitações da intervenção humana, resposta inadequada envolvendo pessoa transgênero e persistência de erros nas referências bibliográficas. Também foram considerados não atendidos os critérios relativos ao conhecimento das limitações demográficas, à prevenção de *proxies* discriminatórios, à resiliência a erros e ataques e à possibilidade de justificar decisões a *posteriori*.

Embora o Eixo 3 tenha recebido a pontuação máxima, as observações indicam a necessidade de atualizar a Avaliação de Impacto em Direitos Humanos, a identificação dos grupos vulneráveis e o plano de mitigação. Devem ser incorporados os riscos

evidenciados na testagem, especialmente a reprodução de dados, o tratamento inadequado de minorias e os problemas de priorização de referências. Além disso, o atendimento dos critérios documentais não demonstra, isoladamente, a efetividade das salvaguardas.

B) Avaliação conversacional do modelo

A planilha de vinhetas apresenta, para o Ciclo 5, cenários com diferentes combinações de idade, deficiência, gênero e identidade de gênero, orientação sexual, raça e etnia, condições socioeconômicas, localização geográfica e pertencimento a populações prioritárias do SUS. Essas variações permitem examinar a adequação contextual das recomendações e possíveis diferenças entre perfis de pacientes. O instrumento organiza os casos de teste, enquanto os resultados de execução são apresentados no relatório da HarpIA.

Na conferência dos registros do Ciclo 5, foram identificadas duas entradas denominadas “Conversa 15”, com descrições distintas, e ausência de entrada denominada “Conversa 24”. Essa inconsistência de identificação deve ser harmonizada com os registros efetivamente utilizados na plataforma, sem presumir que represente duplicação ou ausência de avaliação na HarpIA.

O relatório da HarpIA apresenta avaliações do Ciclo 5 realizadas por dois especialistas médicos, com quinze conversas por avaliador. Entre os achados expressamente atribuídos a esse ciclo, foram registradas omissão de hipótese diagnóstica relevante, sugestão de aplicação dos exames em contexto considerado inadequado pelo especialista, utilização de referências desatualizadas ou sem pertinência, imprecisões temporais e inferência de dados clínicos não fornecidos na vinheta.

Em uma interação, o especialista apontou que o sistema inferiu informações ausentes para compor diagnóstico. Em outra, registrou que aspectos relacionados à determinada doença e respectiva terapia somente foram considerados após questionamento. Também foram identificadas alucinações bibliográficas que persistiram após a contestação do avaliador. Esses registros evidenciam limitações nas conversas examinadas, sem autorizar generalização automática para todas as respostas da solução.

A interpretação dos resultados deve considerar as condições da integração. O relatório da HarpIA registra que, segundo a documentação da participante, a triagem adversarial no canal de avaliação sinaliza tentativas sem bloqueá-las, diferentemente do comportamento descrito para o canal clínico. A plataforma informa que não consumiu nem armazenou os metadados de sinalização e bloqueio mencionados pela empresa. Assim, os resultados observados no canal de avaliação não demonstram, por si sós, comportamento idêntico na interface clínica, permanecendo necessária a verificação dessa correspondência.

Na meta-avaliação referente ao Ciclo 5, foram registradas evolução na exploração das características do sistema e, simultaneamente, interações curtas ou insuficientes para avaliação. Essa análise examina a qualidade metodológica das conversas, sem verificar a correção clínica dos julgamentos especializados. As limitações devem orientar a interpretação das evidências e o aperfeiçoamento do procedimento, sem justificar o descarte automático dos achados.

C) Teste de explicabilidade para transparência

No Ciclo 5, os aspectos de explicabilidade passaram a ser acompanhados pela matriz de transparência, que informa a descontinuidade do instrumento separado de avaliação de explicabilidade. A matriz organiza as informações conforme os destinatários e contempla três camadas: técnica, destinada às equipes de TI e segurança e aos reguladores; de governança, destinada às instituições de saúde; e de usuário, destinada aos pacientes e médicos.

Foram registrados o Manual de Uso do Médico na área autenticada do sistema, avisos de suporte à decisão, alertas de urgência e privacidade e botões de confirmação ou rejeição por resposta clínica. Também foi formalizada, no formulário de autosserviço disponível em "Minha conta", a opção de revisão de decisão automatizada, com previsão de reanálise humana e resposta fundamentada.

Essas medidas ampliam a informação e os canais de manifestação, mas não comprovam, isoladamente, intervenção humana efetiva. Enquanto o parecer técnico reconhece a operacionalização dos botões e do formulário, a avaliação do *framework* geral registra insuficiência da capacidade real de intervenção. Assim, a condicionante é classificada como "concluído / em evolução". Portanto, os avanços devem ser apresentados como resultados específicos, permanecendo necessária a avaliação de seus efeitos concretos sobre a revisão das recomendações e a proteção dos titulares.

D) Matriz de transparência

A matriz apresenta registros do Ciclo 5 para cinco grupos: (i) usuários; (ii) médicos; (iii) hospitais, clínicas e instituições de saúde; (iv) ANPD e demais reguladores; e (v) equipes de TI e segurança. O instrumento corrige expressamente o total de campos de conteúdo para vinte, correspondentes às quatro perguntas (por que, o quê, quando e como) aplicadas aos cinco grupos. O preenchimento integral desses campos indica completude do registro, sem equivaler à comprovação integral de conformidade.

Para os usuários, foram registrados o fluxo de revisão pelo formulário de autosserviço, a disponibilização dos termos e da política de privacidade em diferentes telas e a indicação de tutela da saúde como base legal para os dados dos pacientes. A matriz ressalta que o formulário está na área do médico e recomenda um caminho mais interativo para o paciente, cujo acesso à informação também depende do profissional ou da instituição responsável pelo atendimento. Não foi registrada revisão integral das políticas em linguagem simples.

Para os médicos, foram documentados o manual, as orientações permanentes da interface, a Declaração de Uso Responsável, os termos apresentados no acesso e a validação por resposta. *Workshops* e comunicações por *e-mail* permanecem como sugestões.

Quanto às instituições de saúde, a matriz esclarece que o piloto não conta com hospitais ou clínicas parceiros. Materiais específicos, contratos, relatórios, *dashboards* e definição de papéis de controlador e operador foram indicados para uma futura fase com essas instituições, sem constituírem entregas implementadas no piloto.

Para a ANPD e os demais reguladores, foram registrados a atualização do RIPD, o Quadro de Retificações e um protocolo de demonstração reprodutível, com casos fictícios, versão identificada e registros verificáveis. Permanecem recomendadas a apresentação de métricas de vieses e a reavaliação periódica do risco relativo ao exercício dos direitos dos titulares.

Para as equipes de TI e segurança, foram registrados o inventário atualizado, a preservação da versão em avaliação, o controle de alterações e as trilhas de auditoria. Permaneceram sugestões de complementação por métricas de latência e disponibilidade, periodicidade e responsáveis pela revisão técnica e instrumentos internos de acompanhamento.

E) Parecer técnico: mapa de riscos e mitigações

O parecer técnico de 12 de outubro de 2026, identificado como referente ao Ciclo 5, apresenta conclusão "favorável com condicionantes", recomendando continuidade dos testes, com risco inerente agregado crítico e risco residual agregado médio. A fundamentação destaca a atualização documental, os controles automatizados, o formulário de autosserviço e os mecanismos de registro da validação médica.

Registra-se que o documento relata aprovação nos testes de anonimização, segurança e autosserviço realizados em produção em 7 de outubro de 2026. No teste de anonimização, a aprovação corresponde ao comportamento declarado, incluindo a manutenção da data de nascimento, para a qual não existe detector dedicado na versão examinada. Esses resultados não demonstram remoção de todos os identificadores nem ausência de identificação indireta.

Entre os controles de aplicação, foram registrados limites de tamanho das mensagens e de requisições por minuto, recusa de credenciais inválidas e mecanismos de auditoria. Permanecem condicionantes relativas ao fator aleatório nos *hashes*, ao *firewall de aplicação* dedicado, à contingência do PubMed e ao aperfeiçoamento da detecção de nomes e datas. O exercício simulado de incidente de severidade alta foi reprogramado para adendo ao relatório de testes, sem apresentação de conclusão de sua execução no parecer.

O registro consolidado apresenta oito riscos. Sendo que permanecem pendentes os riscos de: reidentificação, indisponibilidade do PubMed, ataques à aplicação, deficiência de transparência e explicabilidade e ausência de remoção automática da data de nascimento. Outros riscos são descritos como resolvidos, embora persistam inconsistências entre os estados, os valores e as classificações. O risco de exposição da data de nascimento, por exemplo, é descrito como médio, mas recebe doze pontos, sendo classificado como alto pela metodologia. O parecer não explicita uma regra de agregação que reconcilie os riscos individuais elevados com a classificação residual geral "médio". A classificação do risco residual geral como "médio" decorre, conforme esclarecimento do avaliador, de uma apreciação sistêmica das salvaguardas, e não da média aritmética ou da adoção automática da maior pontuação individual. Foram considerados os controles determinísticos na borda, o funcionamento em modo *fail-closed* e o descarte do texto bruto antes do envio ao modelo, bem como a contenção dos efeitos de falhas em componentes específicos. Nesse sentido, a dependência da API do PubMed, embora classificada como risco alto, afeta funcionalidades delimitadas e aciona mecanismos de pausa, sem implicar necessariamente exposição dos dados. A permanência de riscos individuais altos é, portanto, compatível com a avaliação agregada apresentada.

O parecer denomina o mecanismo de proteção como pseudonimização e reconhece limitações da supressão por REGEX. Essa qualificação deve ser atribuída ao avaliador, preservando-se a distinção entre remoção de identificadores e comprovação de impossibilidade de associação ao titular. A inclusão futura de fator aleatório nos *hashes* não demonstra, isoladamente, anonimização do conteúdo clínico remanescente.

Permanecem necessárias a comprovação das condicionantes, a formalização das salvaguardas para os fluxos internacionais e a comunicação e documentação das mudanças da próxima versão. Também devem ser esclarecidas as divergências sobre bloqueio de instruções adversariais, considerando a distinção entre os canais clínico e de avaliação descrita no relatório da HarpIA.

Em conjunto, a Fase B do Ciclo 5 ampliou a documentação das medidas de transparência, dos controles técnicos e dos resultados experimentais, mas manteve limitações relevantes de privacidade, qualidade clínica e bibliográfica, não discriminação e supervisão humana. A avaliação feita a partir do *framework* geral e o parecer técnico apresentam conclusões distintas, que devem permanecer identificadas conforme seus critérios e evidências. A recomendação técnica de continuidade condicionada não equivale à aprovação da solução nem afasta os achados desfavoráveis. O encaminhamento exige corrigir as falhas, comprovar as medidas pendentes e esclarecer a correspondência entre as versões, os canais e as condições efetivamente avaliados.

4.2.4 Fase C

A Fase C reuniu os resultados e os aprendizados do Ciclo 5 sob duas perspectivas complementares: os êxitos para a ANPD e os êxitos para a Prevvine. Foram consideradas as fichas C1 preenchidas separadamente pela ANPD e pela participante. Os documentos permitem confrontar o aprendizado regulatório, os avanços declarados pela empresa e as pendências identificadas, sem constituírem conclusão conjunta, cujo campo permanece sem preenchimento em ambas as fichas.

A) Resultados do Ciclo 05 sob a perspectiva da ANPD

O acompanhamento da plataforma STAiDOC permitiu aprofundar a compreensão dos limites dos mecanismos baseados exclusivamente em expressões regulares (REGEX) para proteção de dados em sistemas de inteligência artificial. A ANPD destacou que a remoção de identificadores diretos não elimina os riscos de inferência e reidentificação a partir do contexto das conversas. Segundo a avaliação registrada, as salvaguardas examinadas não sustentam a caracterização do tratamento como anonimização ou pseudonimização, pois parte das informações inseridas pôde ser recuperada durante a interação.

No âmbito da proteção dos usuários e titulares, a ficha registra que a testagem permitiu identificar riscos de privacidade, exposição indireta de informações e potenciais impactos discriminatórios antes da adoção ampla da solução. Foram destacadas a reprodução reiterada de dados, inclusive mediante ataques por *prompt*, e uma resposta inadequada envolvendo pessoa transgênero. Esses achados fundamentam a necessidade de aperfeiçoar os controles de privacidade e os mecanismos de avaliação e correção de vieses.

Como subsídio regulatório, o Ciclo 5 produziu evidências para discussão sobre anonimização, inferência em modelos de linguagem, transferência internacional e distribuição de responsabilidades entre os agentes envolvidos. A ficha destaca questões relativas à qualificação das chamadas a modelos estrangeiros, à definição dos papéis na cadeia de tratamento e à compatibilidade entre os mecanismos regulatórios brasileiros e os contratos utilizados pelos provedores de inteligência artificial.

Quanto aos efeitos sobre o mercado e a concorrência, a ANPD assinalou que os resultados reforçam a necessidade de padrões mais robustos de governança, privacidade

e avaliação de vieses para sistemas de IA em saúde. Foram ressaltados o desenvolvimento de mecanismos de correção de vieses, o saneamento das lacunas de governança antes da disponibilização da solução ao mercado e o aperfeiçoamento da prestação de contas sobre o tratamento de dados pessoais. Trata-se de aprendizado sobre as condições esperadas para a adoção responsável, sem demonstração de efeito concorrencial já produzido.

B) Resultados do Ciclo 05 sob a perspectiva da Prevvine

A Prevvine apresentou sua avaliação em seis dimensões: (i) evoluções técnicas; (ii) aprendizados regulatórios; (iii) transparência e explicabilidade; (iv) direitos dos titulares; (v) autorizações e análises pendentes; e (vi) outros aprendizados. O preenchimento foi realizado pelo representante legal e pelo líder técnico da participante.

Quanto às evoluções técnicas, a empresa informou a execução de testes pela API pública em 7 de outubro de 2026, com conta identificável, casos fictícios e versão preservada desde 25 de agosto de 2026. Foram declarados resultados de nove de nove casos e 188 de 188 verificações no teste denominado de anonimização, sete de sete verificações de eventos de segurança e seis de seis verificações de direitos dos titulares. Esses números representam os resultados relatados pela participante para os procedimentos descritos, sem demonstrarem, isoladamente, anonimização integral ou superação das vulnerabilidades apontadas pela ANPD.

A empresa também registrou alterações de interface realizadas em outubro: confirmação ou rejeição por resposta clínica, aviso fixo de suporte à decisão, alerta destacado de urgência, aviso preventivo de privacidade e canais de suporte e contato com o encarregado de dados. Segundo a ficha, essas inclusões não alteraram o modelo, as instruções do sistema ou a cadeia de processamento.

No campo regulatório, a participante reconheceu que o fluxo dirigido ao provedor do modelo envolve dados pessoais sensíveis submetidos à supressão de identificadores, sem caracterização como anonimização. Também reconheceu a necessidade de informação aos titulares, de adequação dos fluxos internacionais e de alinhamento das bases legais entre o registro de operações, o termo de consentimento e a política de privacidade. A ficha esclarece, ainda, que o direito à revisão de decisões automatizadas não constitui hipótese legal de tratamento.

Em relação à transparência, a Prevvine informou a disponibilização do Manual de Uso do Médico, dos Termos de Uso e da Política de Privacidade em diferentes telas, o aceite registrado da Declaração de Uso Responsável e a identificação nominal do encarregado de dados. Registrou a sinalização das referências bibliográficas conforme sua verificação e a identificação do modelo nos registros de explicabilidade. Como aprendizado, reconheceu que as explicações produzidas pelo modelo não equivalem necessariamente à descrição fiel do processamento e que os documentos devem apresentar informações acessíveis e coerentes entre si.

Quanto aos direitos dos titulares, a empresa declarou a operação do formulário de autosserviço em “Minha conta”, com abertura e consulta de solicitações e cálculo automático de prazo operacional de quinze dias úteis. Informou a reemissão do formulário e do termo de consentimento em próxima versão do sistema. Para os pacientes, reconheceu a necessidade de estruturar informação e canais de comunicação, inclusive em caso de incidente, na futura fase com instituições parceiras.

A própria participante relacionou pendências relativas à proteção dos resumos criptográficos, ao *firewall* de aplicação, à contingência do PubMed, ao aperfeiçoamento

dos detectores de nomes e datas, à medição de precisão, sensibilidade e índice F1 e aos mecanismos de resposta institucional, controle de saída e bloqueio pré-modelo. Também indicou como pendentes a formalização das cláusulas-padrão com operadores internacionais, os ajustes documentais e das bases legais, os testes próprios de vieses, os mecanismos de correção, a inclusão de fontes brasileiras, os resultados da consulta a médicos, o exercício simulado de incidente, o teste de suspensão, bloqueio e eliminação e a remoção do ícone de anexo de imagem.

Como providências adotadas, informou a elaboração da Avaliação de Impacto em Direitos Humanos, o ato de identificação nominal do encarregado de dados e o pedido de autorização para implantar a versão corrigida e realizar novos testes no Ciclo 6. A solicitação de cláusulas-padrão à Supabase foi registrada em 14 de outubro de 2026, sem comprovação até o momento, na ficha, da conclusão desse procedimento.

C) Coleta e análise das evidências

A leitura conjunta das fichas permite identificar convergência quanto à insuficiência da mera supressão de identificadores para sustentar anonimização e à necessidade de aperfeiçoar privacidade, governança, transparência e avaliação de vieses. Os avanços relatados pela empresa devem ser relacionados às evidências correspondentes, preservando-se as limitações identificadas pela autoridade.

Um esclarecimento relevante da ficha da participante diz respeito à triagem adversarial: na versão em avaliação, o mecanismo sinaliza e registra tentativas de manipulação, enquanto o bloqueio anterior ao modelo e a resposta institucional fixa ainda não estão implantados nessa versão. Essas funcionalidades integram a próxima versão identificada. Portanto, as referências à implementação desses controles nos demais registros devem ser reconciliadas mediante identificação da versão, do canal e das evidências examinadas.

O resultado demonstrado consiste na produção de aprendizado regulatório, no registro de avanços específicos e na delimitação das correções necessárias. A continuidade experimental não equivale à aprovação da solução nem ao atendimento integral das condicionantes.

Para o encaminhamento ao Ciclo 6, a ficha da empresa prevê implantação de versão corrigida e nova testagem, condicionadas à autorização solicitada. O acompanhamento deverá distinguir as medidas já operacionais das funcionalidades planejadas e verificar sua efetividade, especialmente quanto à proteção de dados, aos vieses, à qualidade bibliográfica, à supervisão humana e aos fluxos internacionais.

Quanto à descontinuidade, a participante registra como pendentes o exercício simulado de incidente e o teste de suspensão, bloqueio e eliminação. As fichas C1 não apresentam comparação específica entre versões do Plano de Descontinuidade que permita afirmar ausência ou ocorrência de alterações entre os Ciclos 4 e 5. O registro pertinente à Fase C é, portanto, a necessidade de executar e documentar esses procedimentos, com evidências de seus resultados.

4.3 Metatext

No Ciclo 5, o acompanhamento da solução GuardionAI, da participante Metatext, concentrou-se na avaliação dos mecanismos de pseudonimização, na análise das

informações destinadas à equipe de segurança e na identificação das evidências necessárias à verificação das salvaguardas. As atividades consideraram o funcionamento da aplicação como intermediária entre os sistemas das instituições contratantes e os modelos de linguagem, utilizando dados sintéticos em infraestrutura dedicada ao sandbox, localizada no Brasil.

A avaliação da proteção de dados permitiu registrar o desempenho do mecanismo de detecção e mapear falhas por categoria de informação. A ficha C1 apresenta precisão de 0,77 e sensibilidade (*recall*) de 0,71, indicando a necessidade de reduzir tanto a exposição de identificadores não detectados quanto a supressão de informações necessárias ao atendimento. Os achados orientaram recomendações de calibração e retestes, especialmente para RG, CPF apresentado isoladamente e dados de agência e conta bancária inseridos em argumentos de ferramentas.

No eixo de transparência e explicabilidade, foram documentados dois níveis de informação: explicações ao titular e registros para controladores, auditores e equipes de segurança. O Teste 2, realizado na plataforma HarpIA, abrange a avaliação de trinta incidentes por dois avaliadores, considerando dez quadros por incidente e totalizando seiscentos itens de avaliação. A consolidação dos resultados permanece em andamento, com apresentação prevista para o Ciclo 6, para subsidiar o exame da correspondência entre as evidências, as classificações de risco e as justificativas dos alertas.

O ciclo também registrou o recebimento de mais de trinta documentos pelo Microsoft Teams e a emissão de parecer favorável com condicionantes. Para o Ciclo 6, estão previstas a análise conclusiva dessa documentação, a consolidação dos testes e a verificação das recomendações sobre segurança, registros de auditoria, vieses, funcionamento do cofre de uso único e execução das simulações. A avaliação deverá relacionar os controles descritos às evidências de implementação aos riscos remanescentes, preservando a distinção entre continuidade experimental e aprovação da solução.

4.3.1 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 5 deste Sandbox Regulatório:

Principais documentos:

- Fase A:
    - Fichas A1 e A2; e
    - Fichas A3 e A4.
- Fase B:
    - Parecer técnico consolidado;
    - Matriz de transparência;
    - Relatório de Análise - Plataforma HarpIA (Teste 1);
    - Teste 1 — Instruções;
    - Teste 1 — *Prompts*;
    - Teste 2 — Instruções; e
    - Teste 2 — Exemplos.
- Fase C:

- Ficha C1 ——Êxitos para a ANPD e empresa participante.

4.3.2 Fase A

Inicialmente, no que se refere às fichas A1 e A2, foram mantidas, no Ciclo 5, as conclusões anteriores sobre a caracterização da solução GuardionAI, da participante Metatext, sua arquitetura, os fluxos de dados, a infraestrutura de testes e a avaliação da maturidade tecnológica. Conforme previsto, o foco deste ciclo foi a realização dos testes da Fase B, destinados a confrontar as informações apresentadas pela participante com as evidências de funcionamento da solução.

Em relação às fichas A3 e A4, também foram preservadas as conclusões e recomendações anteriores quanto à governança, à segurança da informação, à proteção de dados pessoais, à transparência, à explicabilidade, à supervisão humana e à coordenação institucional. O mapeamento normativo e os registros da avaliação pelo *framework* geral permaneceram como referências, sem que sua manutenção represente nova pontuação ou comprovação de superação das lacunas anteriormente identificadas.

Foi recebida a documentação complementar solicitada à participante, composta por mais de trinta documentos compartilhados por meio do Microsoft Teams. Esse conjunto será analisado em caráter conclusivo no Ciclo 6, em articulação com as evidências dos testes. Seu recebimento constitui avanço na instrução documental, mas não permite, por si só, considerar atendidas as recomendações ou comprovada a implementação das salvaguardas descritas.

Permanecem como questões de acompanhamento a efetividade dos mecanismos de pseudonimização, a explicabilidade dos alertas, os pontos de supervisão humana, a resposta a incidentes e o esclarecimento dos ambientes, dos fluxos de dados e das responsabilidades dos envolvidos. Também deverá ser verificada a disponibilização de informações às instituições contratantes, às equipes de segurança e aos titulares, conforme as atribuições de cada agente.

Assim, a Fase A do Ciclo 5 preservou a base de análise e os encaminhamentos do Ciclo 4, enquanto as atividades se concentraram na testagem da solução. A avaliação conclusiva da documentação recebida ficará para o Ciclo 6, distinguindo as medidas declaradas, as evidências verificadas e as pendências remanescentes, sem que a continuidade dos testes equivalha à aprovação da solução.

4.3.3 Fase B

A Fase B do Ciclo 5 concentrou-se na realização e no acompanhamento dos testes, no aprofundamento da avaliação de transparência e explicabilidade e na atualização do parecer técnico consolidado. Em continuidade ao ciclo anterior, as atividades buscaram confrontar as informações documentais com evidências operacionais, identificar limitações dos mecanismos de proteção de dados e organizar as providências necessárias à consolidação das avaliações no Ciclo 6.

A) Matriz de transparência

Foram mantidos como referência os quatro grupos de interlocutores anteriormente definidos: (i) clientes finais das instituições financeiras; (ii) instituições financeiras; (iii) ANPD e demais reguladores; e (iv) equipe de segurança. As recomendações de

transparência continuaram orientadas à compreensão das informações disponibilizadas, à definição das responsabilidades e à possibilidade de esclarecimento, contestação e revisão dos resultados produzidos pela solução.

Conforme registrado no parecer técnico do Ciclo 5, a participante estruturou a explicabilidade em dois níveis. O primeiro destina-se ao titular, mediante informações em linguagem simples sobre a aplicação das medidas de proteção. O segundo destina-se ao controlador, aos auditores e à equipe de segurança, contendo informações técnicas sobre os trechos identificados, as categorias de dados, os índices de confiança e as justificativas geradas pela solução GuardionAI.

Essa organização representa avanço na diferenciação das informações conforme o destinatário. Entretanto, a suficiência das explicações e sua utilidade para o exercício dos direitos dos titulares permanecem sujeitas à avaliação conclusiva. A funcionalidade de atribuição da contribuição de cada elemento da entrada para o resultado do modelo, denominada *Feature Attribution*, continua em desenvolvimento, não devendo ser considerada integralmente implementada ou validada.

Também foi recebida a documentação complementar solicitada, composta por mais de trinta documentos compartilhados por meio do Microsoft Teams. Sua análise conclusiva será realizada no Ciclo 6, em conjunto com as evidências dos testes, sem que o recebimento permita considerar automaticamente atendidas as recomendações anteriores.

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

O parecer técnico consolidado de 14 de outubro de 2026, referente ao Ciclo 5, apresenta conclusão preliminar “suficiente com condicionantes” e parecer final “favorável com condicionantes”, recomendando o prosseguimento para o Ciclo 6 mediante cumprimento das exigências estabelecidas.

Apesar de haver diferenças internas na classificação agregada de risco, o parecer identifica expressamente a permanência de riscos elevados relacionados a vieses discriminatórios e à insuficiência das explicações. Consequentemente, do risco residual ser classificado como “médio”, não se deve interpretá-lo isoladamente como comprovação da suficiência integral das salvaguardas.

O parecer registra avanços na organização documental, incluindo a disponibilização de 17 (dezessete) políticas e documentos normativos em português[19], e no funcionamento do ambiente experimental com dados sintéticos. No entanto, distingue medidas verificadas de providências declaradas ou documentadas que ainda dependem de demonstração prática. Para o próximo ciclo, foram estabelecidas cinco condicionantes:

- Testes de intrusão e sanitização dos registros: apresentar evidências de testes de segurança específicos para modelos de linguagem e comprovar a implantação da respectiva correção, destinada a

[19] As políticas e documentos são: 1. Termo de Uso; 2. Política de Privacidade; 3. Política de *Cookies*; 4. Acordo de Tratamento de Dados Pessoais; 5. Lista de Suboperadores; 6. Política de Proteção de Dados; 7. Política de Retenção e Eliminação de Dados; 8. Relatório de Impacto à Proteção de Dados — RIPD; 9. Política de Segurança; 10. Política de Gestão de Acessos; 11. Política de Criptografia; 12. Política de Classificação da Informação; 13. Política de Segurança em Nuvem; 14. Plano de Resposta a Incidentes; 15. Política de Desenvolvimento Seguro; 16. Política de Gestão de Ativos e 17. Política de Gestão de Mudanças.

assegurar que os registros de auditoria armazenem apenas conteúdo com os identificadores tratados.

- Certificações da infraestrutura e da participante: apresentar os comprovantes das certificações e dos relatórios de auditoria aplicáveis à infraestrutura, mantendo o acompanhamento das certificações independentes da própria participante.
- Execução das simulações: realizar e documentar os exercícios de resposta a incidentes, eliminação de dados, explicação e contestação e descontinuidade, com os respectivos registros de execução.
- Avaliação de vieses e retestes de detecção: executar o protocolo de vieses e *proxies* discriminatórios, com métricas por grupo, e repetir os testes dos cenários em que foram identificadas falhas, especialmente RG, CPF isolado e agência e conta bancária em argumentos de ferramentas.
- Verificação do *One-Time Vault*: demonstrar tecnicamente o funcionamento do cofre de vínculos entre *tokens* e dados originais, comprovando a leitura única, a impossibilidade de segunda leitura e a eliminação definitiva do vínculo após sua utilização.

Aguarda-se o atendimento integral dessas recomendações no Ciclo 6. A continuidade permanece condicionada à apresentação e à avaliação das evidências, observados os gatilhos de pausa ou suspensão previstos para incidentes, falhas sistemáticas de proteção de dados e vieses relevantes sem mitigação.

C) Teste 1: Proteção de dados

Preliminarmente, nas fichas A1 e A2, a empresa informou a realização de 540 chamadas além de 40 sessões bancárias simuladas. Essa informação sobre os testes de infraestrutura realizados pela empresa foi considerada no parecer técnico da equipe do CIAAM/USP. Note-se que essas situações são independentes dos resultados alcançados na plataforma HarpIA.

Segundo o documento, o ambiente apresentou estabilidade operacional, com processamento no GCP São Paulo e latência de inferência inferior a cem milissegundos. A operação ocorreu sob a política *Redact Only*, voltada ao mascaramento dos identificadores detectados, sem bloqueio ativo das operações no escopo experimental descrito.

Os resultados também revelaram limitações relevantes. O parecer registra ausência de detecção de RG em seis de seis chamadas específicas, falhas na identificação de CPF apresentado isoladamente e ausência de detecção de agência e conta bancária em dez de quatorze chamadas com esses elementos nos argumentos de ferramentas. Esses achados foram incorporados aos encaminhamentos para calibração e retestes.

Considerando os testes 1A na plataforma HarpIA Survey e 1B do HarpIA LAB, houve 1.365 execuções, que correspondem a:

- Teste 1A: 122 execuções na plataforma HarpIA Survey com avaliadores humanos [20]
- Teste 1A: 18 execuções adicionais para resolver questões específicas na plataforma HarpIA Survey [21]
- Teste 1B: 1.225 execuções na plataforma HarpIA Lab

Assim, o total é de 1.365 execuções quando somamos os resultados alcançados nas plataformas HarpIA Survey e HarpIA Lab.

Pelo relatório da plataforma HarpIA referente ao teste 1A, verificou-se que o sistema da empresa Metatext apresentou precisão de 0,77, recall de 0,71 e índice F1 micro de 0,74. Esses valores foram obtidos no Teste 1A em um contexto de 122 mensagens com trechos a serem anonimizados. O valor é calculado sobre a soma total de trechos a serem anonimizados, não sobre as mensagens executadas, ou seja, uma mensagem pode ter mais de um trecho a ser anonimizado.

Como houve dois testes (1A e 1B), foi possível verificar uma diferença de performance[22], quais sejam: o F1 micro permaneceu relativamente próximo nos dois testes passando de aproximadamente 74% no teste 1A para 77,4% no teste 1B. No entanto, houve uma mudança de equilíbrio entre as métricas precisão e *recall*, pois, a precisão diminuiu de 78% no teste 1A para 72,9% no teste 1B, enquanto o *recall* aumentou de aproximadamente 71% para 82,5%. Dessa forma, no relatório do CIAAM/USP e no relatório da plataforma HarpIA referentes aos testes acima apresentados, ao final, foram informados valores referentes à média harmônica entre os resultados do teste 1A e teste 1B. Assim, a capacidade de processamento e a estabilidade observadas não equivalem à comprovação de eficácia suficiente para todas as categorias de dados ou condições de uso.

Essas informações são apresentadas conforme sua consolidação no parecer técnico. A confirmação das correções dependerá dos retestes previstos para o Ciclo 6, com identificação das versões e das condições avaliadas. Permanecem igualmente pendentes as medições de utilidade semântica após o tratamento dos identificadores e os testes específicos de resistência à reidentificação.

---

[20] Observe-se que as 122 execuções se referem à realização de 30 avaliações por 4 avaliadores, ou seja, 30 *prompts* analisados por 4 avaliadores humanos, o que resulta em 120 execuções. Além disso, o sistema indicou mais 2 execuções dos mesmos avaliadores, o que, normalmente, é resultado de uma alteração da avaliação diante de dúvidas sobre o comportamento do sistema. Exemplo: o avaliador envia duas respostas para o mesmo item avaliado.

[21] As 18 execuções são referentes a *prompts* adicionais elaborados pelos avaliadores para investigar questões específicas identificadas durante a avaliação.

[22] Foram utilizadas as seguintes métricas para avaliação do modelo de IA referente à anonimização dos dados: 1. precisão: “A acurácia é a proporção de todas as classificações que estavam corretas, sejam elas positivas ou negativas.”. 2. F1 micro: “A pontuação da Micro f1 é uma métrica de vários rótulos/classes para avaliações de qualidade de IA generativa que mede o desempenho dos ativos de IA generativa em tarefas de extração de entidades para previsões de vários rótulos/múltiplas classes.”; 3. Recall: “A taxa de verdadeiro positivo (TVP), ou a proporção de todos os positivos reais que foram classificados corretamente como positivo”.
Fontes: https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall e https://www.ibm.com/docs/pt-br/watsonx/w-and-w/2.3.x?topic=metrics-micro-f1-score, acessadas em 29/10/2026.

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

No Ciclo 5, deu-se continuidade ao Teste 2, preparado no ciclo anterior e destinado à avaliação das informações apresentadas à equipe de segurança. A atividade envolve dois avaliadores, trinta incidentes e dez quadros por incidente, totalizando seiscentos itens de avaliação. A análise e a consolidação dos resultados permanecem em andamento, com apresentação prevista para o Ciclo 6.

Embora a interface da plataforma utilize o rótulo “conversa”, as instruções esclarecem que o objeto examinado é um painel de informações sobre cada alarme. Os materiais orientam a verificação da correspondência entre o título do incidente, os indicadores detectados, o conteúdo que originou o alerta e as justificativas apresentadas pelo sistema.

A avaliação abrange a coerência entre as evidências, o nível de risco atribuído e a classificação produzida pela GuardionAI. As diretrizes e os exemplos indicam quais quadros devem fundamentar as respostas, buscando uniformizar a interpretação dos avaliadores e permitir a comparação dos registros.

O teste deverá subsidiar a análise da clareza, da consistência e da utilidade das explicações para a atuação humana. Enquanto os resultados estiverem em consolidação, não é possível atribuir ao instrumento uma conclusão definitiva sobre a qualidade dos alertas, a suficiência da explicabilidade ou o atendimento das recomendações de transparência.

Em conjunto, a Fase B do Ciclo 5 ampliou as evidências sobre o funcionamento da GuardionAI, registrou limitações de detecção e aprofundou a avaliação das informações destinadas à equipe de segurança. Para o Ciclo 6, permanecem a consolidação dos resultados do Teste 02, a análise conclusiva da documentação recebida e a verificação do atendimento integral das cinco condicionantes. A recomendação de continuidade experimental não equivale à aprovação da solução nem à confirmação de conformidade integral.

4.3.4 Fase C

A Fase C do Ciclo 5 reuniu os resultados e os aprendizados registrados na Ficha C1 sob as perspectivas da ANPD e da empresa. O principal avanço consistiu na produção de evidências sobre o desempenho da pseudonimização e no aprofundamento da compreensão dos riscos, das responsabilidades e dos requisitos de transparência aplicáveis à solução.

Sob a perspectiva da ANPD, os testes demonstraram que a adoção de uma camada adicional de segurança não elimina a necessidade de avaliar os riscos introduzidos pelo próprio componente. A ficha registra precisão de 0,77 e sensibilidade (*recall*) de 0,71 na avaliação da pseudonimização, indicando tanto a possibilidade de dados pessoais permanecerem expostos quanto a supressão desnecessária de informações relevantes ao atendimento. Esses resultados reforçam a necessidade de aperfeiçoar o mecanismo, mapear suas limitações e avaliar os efeitos sobre a proteção dos titulares e a utilidade das interações.

O aprendizado regulatório também destacou a importância das políticas de segurança da informação, gestão de acessos, resposta a incidentes, avaliação de impactos, retenção e descarte, acompanhadas de testes e evidências de implementação. Embora ainda não tenham sido identificados subsídios robustos para a elaboração de

nova norma, foram delimitados temas relevantes para futuras iniciativas regulatórias, especialmente auditorias independentes, avaliação específica da pseudonimização, transferências internacionais e mitigação de riscos. A possibilidade de aplicação da arquitetura em outros setores amplia a relevância desses aprendizados, observada a compatibilidade com os sistemas das organizações contratantes.

Sob a perspectiva da participante, o programa permitiu mapear falhas por categoria de dado e orientar o aprimoramento dos modelos, com prioridade para aumentar a sensibilidade e a precisão. A empresa informa a realização de mais de 1.300 execuções de teste ao longo do programa e destaca avanços nos alertas acompanhados de justificativas, na fila de revisão humana e no cofre de uso único. Esses registros devem ser considerados em conjunto com as verificações técnicas e as condicionantes ainda pendentes, sem pressupor a validação integral das funcionalidades descritas.

No campo regulatório, a participante destaca o esclarecimento das responsabilidades da contratante como controladora e da Metatext como operadora, bem como a realização das atividades do sandbox em infraestrutura dedicada no Brasil. Também registra a incorporação de explicações diferenciadas conforme o destinatário e a estruturação de canal de privacidade e de procedimentos de apoio à controladora no atendimento aos titulares, inclusive quanto à revisão de decisões automatizadas. A ausência de relação direta com o cliente final não afasta a necessidade desses mecanismos de cooperação e transparência.

Para o Ciclo 6, permanecem a consolidação dos resultados do Teste 2, a análise conclusiva da documentação recebida e a comprovação do atendimento às recomendações do parecer técnico, incluindo testes de segurança, auditorias, simulações e retestes das falhas identificadas. Assim, os êxitos do Ciclo 5 concentram-se na ampliação do conhecimento técnico e regulatório e na definição das melhorias necessárias, preservada a distinção entre funcionalidades declaradas, controles verificados e medidas ainda dependentes de comprovação.

5. Conclusões gerais do Ciclo 5

O Ciclo 5 deu continuidade ao acompanhamento das soluções Trajetto, da Synapse; STAiDOC, da Prevvine; e GuardionAI, da Metatext, articulando a caracterização das aplicações, a análise documental, as evidências dos testes e a avaliação dos aprendizados. Os resultados consolidados no item 4 permitiram identificar controles verificados, limitações de funcionamento e medidas ainda dependentes de implementação ou comprovação, fundamentando os encaminhamentos para o Ciclo 6.

As conclusões devem ser interpretadas conforme o escopo, as versões, os ambientes e os procedimentos de cada avaliação. Os instrumentos utilizados examinam dimensões distintas e não produzem resultados automaticamente equivalentes. Assim, pontuações éticas, níveis de maturidade tecnológica, testes funcionais, pareceres de segurança e declarações das participantes devem permanecer identificados segundo seus critérios e evidências, sem compensação automática entre resultados favoráveis e desfavoráveis.

Quanto à Synapse, o Ciclo 5 preservou os avanços arquiteturais e as conclusões anteriores sobre o Trajetto, mantendo os resultados quanto ao *framework* geral e aos níveis de TRL dos componentes avaliados. O parecer relativo aos riscos e à segurança da informação permaneceu “suficiente com condicionantes". Os ensaios técnicos haviam sido encerrados no Ciclo 3, não sendo atribuídos ao Ciclo 5 novos resultados desses testes. As atividades declaradas pela empresa concentraram-se na experiência do usuário e no desenvolvimento dos mapas internos.

Permaneceram pendentes a demonstração das informações efetivamente apresentadas na interface, a acessibilidade, o inventário de API e terceiros, a comprovação da não retenção de imagens nos ambientes dos parceiros e os testes de robustez e resistência à manipulação de coordenadas. A combinação de geolocalização com atributos de acessibilidade continuou associada a risco alto de reidentificação, não afastado pela renovação do identificador de sessão. Também permanecem necessárias a conciliação das informações sobre autenticação e versões, a definição dos prazos de conservação dos históricos e a comprovação da operacionalização das medidas de descontinuidade.

Quanto à Prevvine, o Ciclo 5 registrou retificações na descrição da STAiDOC, alterações de interface, mecanismos de manifestação dos usuários e resultados de testes automatizados. Entretanto, as avaliações conversacionais identificaram limitações de proteção de dados, qualidade clínica e bibliográfica, não discriminação e supervisão humana. Os 43 pontos no *framework* geral corresponderam à faixa “não adequado para sandbox ou regulação responsável”, conforme os critérios do instrumento. Esse resultado deve permanecer expresso, acompanhado dos achados que o fundamentaram. Esclarece-se ainda que, a classificação do *framework* geral não configura impeditivo para continuidade da participante no projeto piloto.

O parecer técnico, por sua vez, recomendou continuidade com condicionantes, registrando risco inerente crítico e residual agregado médio. Essa conclusão não afasta os resultados da avaliação ética nem as falhas observadas nas interações. Permanecem necessárias a conciliação dos riscos individuais com a classificação agregada e a verificação da correspondência entre versões, canais e condições de testagem. A execução do filtro antes do modelo e a aprovação nos casos automatizados examinados

não demonstram remoção integral dos identificadores ou impossibilidade de associação ao titular. A própria participante reconheceu limitações dos detectores e a ausência de caracterização do conteúdo como anonimizado. Também reconheceu pendências de segurança, avaliação de vieses, qualidade bibliográfica, fluxos internacionais e execução dos procedimentos de incidente e descontinuidade.

Quanto à Metatext, o Ciclo 5 ampliou as evidências sobre o funcionamento da GuardionAI e o desempenho da pseudonimização. A Ficha C1 registra precisão de 0,77 e sensibilidade de 0,71, indicando a coexistência de identificadores não detectados e supressões desnecessárias. O parecer documentou falhas envolvendo RG, CPF isolado e dados de agência e conta bancária em argumentos de ferramentas, com encaminhamento para calibração e retestes. Esses resultados evidenciam que a adoção de um componente de segurança também exige avaliação dos riscos introduzidos pelo próprio mecanismo e de seus efeitos sobre a utilidade do atendimento.

Foram documentados dois níveis de explicação, destinados aos titulares e aos interlocutores técnicos, permanecendo pendente a avaliação conclusiva de sua suficiência. O Teste 2 abrange trinta incidentes, dois avaliadores e dez quadros por incidente, totalizando seiscentos itens de avaliação, cuja consolidação será apresentada no Ciclo 6. Também foi recebido um conjunto de mais de trinta documentos pelo Microsoft Teams, com análise conclusiva prevista para o próximo ciclo. O parecer favorável com condicionantes exige comprovação dos testes de segurança, sanitização dos registros, certificações, simulações, avaliação de vieses e funcionamento do cofre de uso único. As diferenças internas na classificação agregada de risco deverão ser harmonizadas, preservando a identificação dos riscos elevados de vieses e insuficiência das explicações.

De modo transversal, o Ciclo 5 demonstrou que documentação, funcionamento de controles e efetividade das salvaguardas constituem objetos distintos de verificação. A apresentação de políticas, a disponibilização de canais e a existência de revisão humana prevista no fluxo não comprovam, isoladamente, proteção dos dados, compreensão das informações ou intervenção efetiva. Da mesma forma, a estabilidade operacional não substitui a avaliação da precisão, da sensibilidade, da qualidade das respostas e dos impactos sobre os titulares.

Para o Ciclo 6, recomenda-se priorizar:

- acompanhamento das condicionantes: consolidar responsáveis, prazos, evidências exigidas e critérios de conclusão, distinguindo medidas verificadas, declaradas, documentadas e planejadas. As classificações de risco deverão ser atualizadas e conciliadas com os resultados individuais e as respectivas metodologias.
- Synapse: comprovar transparência e acessibilidade na interface; avaliar e mitigar a reidentificação pela combinação de localização e atributos de acessibilidade; apresentar o memorial de processamento de imagens e o inventário de terceiros; verificar autenticação, retenção e resiliência; e executar e documentar os procedimentos de descontinuidade.
- Prevvine: corrigir e retestar as falhas de privacidade, segurança, qualidade clínica e bibliográfica e não discriminação; verificar a efetividade da supervisão humana; atualizar as avaliações de impacto e mitigação; formalizar as salvaguardas dos fluxos internacionais; e comprovar as medidas previstas para a próxima versão. Sua implantação e testagem deverão observar a autorização solicitada e preservar a rastreabilidade em relação à versão anterior.

- GuardionAI: consolidar os resultados do Teste 2; concluir a análise documental; executar os retestes de detecção e o protocolo de vieses; comprovar a sanitização dos registros e o funcionamento do cofre de uso único; apresentar as evidências de segurança e certificação; e realizar as simulações previstas nas cinco condicionantes do parecer.

Para a consolidação do relatório final, permanecem necessárias:

- a correspondência entre modelos, códigos, API, interfaces, ambientes e resultados, com registro das alterações e identificação das versões avaliadas.
- a definição dos agentes de tratamento, das responsabilidades, dos fluxos internacionais e dos procedimentos de retenção, eliminação e atendimento aos titulares.
- atualização dos instrumentos de impacto, risco e maturidade tecnológica conforme as evidências, incluindo resultados desfavoráveis e limitações metodológicas.
- a execução documentada dos procedimentos de resposta a incidentes e descontinuidade, com verificação dos responsáveis, das ações e dos resultados.
- realização de devolutivas entre participantes, CIAAM/USP e ANPD para esclarecer divergências e fundamentar os encaminhamentos de encerramento.

O Ciclo 5 produziu evidências para delimitar as capacidades e as limitações das soluções e orientar as verificações do ciclo final. A conclusão do Ciclo 6 deverá explicitar quais condicionantes foram atendidas, quais riscos permanecem e quais medidas não puderam ser comprovadas. A continuidade experimental não equivale à aprovação das soluções, à certificação de conformidade ou à autorização de exploração comercial irrestrita.


sandbox regulatório


ANPD