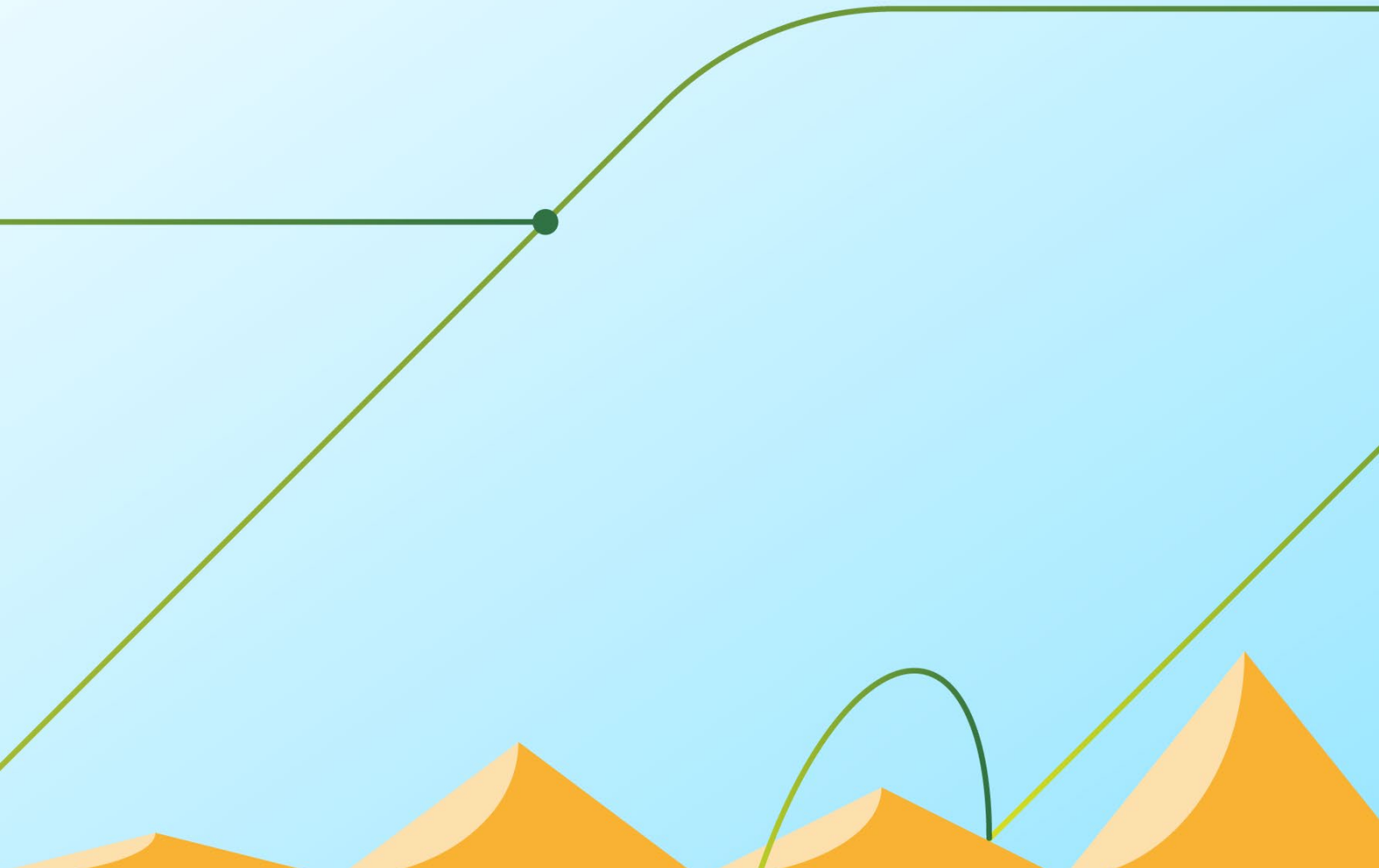
ANPD

sandbox regulatório

6º Relatório Parcial de Monitoramento

Fase de Teste – Ciclo 6



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 8
3. Definição dos temas e metodologia para as sessões remotas e presenciais 9
4. Apoio, orientação e supervisão das atividades de testagem 11
4.1 Synapse 11
4.1.1 Documentação principal 12
4.1.2 Fase A 13
4.1.3 Fase B 19
4.1.4 Fase C 27
4.2 Prevvine 30
4.2.1 Documentação principal 31
4.2.2 Fase A 31
4.2.3 Fase B 38
4.2.4 Fase C 48
4.3 Metatext 52
4.3.1 Documentação principal 53
4.3.2 Fase A 53
4.3.3 Fase B 56
4.3.4 Fase C 64
5. Consultas setoriais 66
5.1 Consulta e aprendizados — Setor de serviços financeiros 66
5.2 Consulta e aprendizados — Setor de Saúde 68
5.3 Consulta e aprendizados — Setor de mobilidade urbana 69
5.4 Conclusões gerais das reuniões setoriais 70
6. Conclusões gerais do Ciclo 6 73

1. Introdução

Este relatório apresenta as atividades que consolidaram o encerramento do Ciclo 6 do Projeto Piloto de Sandbox Regulatório de Inteligência Artificial e Proteção de Dados, com desfechos distintos para as três participantes. A Synapse[1], com o Trajetto, alcançou sucesso parcial, reconhecendo-se os avanços técnicos e regulatórios, acompanhados de condicionantes e comprovações remanescentes. A Prevvine[2], com a STAiDOC, encerrou a experimentação com aprendizados relevantes e aperfeiçoamentos documentais e operacionais, mas com correções e retestes pendentes, necessários à verificação da efetividade das salvaguardas e das melhorias propostas. A Metatext[3], com a GuardionAI, recebeu recomendação de encerramento satisfatório da participação, fundamentada no cumprimento das cinco condicionantes registradas no parecer final, preservadas as recomendações de aperfeiçoamento continuado. Dessa forma, o relatório distingue os avanços demonstrados das providências ainda necessárias, sem atribuir ao encerramento da experimentação o caráter de aprovação irrestrita das soluções ou de comprovação integral da mitigação dos riscos.

A análise está organizada nas Fases A — Pré-teste; B — Teste; e C — Pós-teste, de avaliação e ajuste. Integram sua base as fichas técnicas, as avaliações de maturidade tecnológica, os pareceres de segurança e privacidade, os instrumentos de avaliação ética, as matrizes de transparência, os registros dos testes, as fichas de êxitos e os planos de descontinuidade. Esses elementos são examinados conforme sua finalidade, disponibilidade e pertinência, distinguindo as atividades realizadas no ciclo final da consolidação de evidências produzidas nas etapas anteriores.

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.

---

[1] A Synapse AI é uma empresa vinculada ao Insper que desenvolve o 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

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);
- 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 os avanços na arquitetura de proteção de dados, na formalização da governança e na documentação do Trajetto, acompanhados da evolução da avaliação ética. O exame contempla geolocalização, acessibilidade, processamento de imagens, integrações com terceiros e riscos de identificação indireta. Também delimita as comprovações remanescentes de transparência na interface, robustez em condições reais e proteção contra correlação de trajetos, além de apreciar o Plano de Descontinuidade e os requisitos associados à eventual operação regular.

Quanto à Prevvine, o documento examina a revisão da caracterização do fluxo da STAiDOC como tratamento de dados pessoais sensíveis submetidos à supressão de identificadores, os aperfeiçoamentos de supervisão humana e os resultados clínicos e adversariais reunidos no encerramento. A análise abrange a proteção dos dados, a qualidade das respostas e referências bibliográficas, a avaliação de vieses e as correções previstas para versão subsequente. Ademais, incorpora os avanços documentais, a interlocução com as autoridades do setor de saúde e as adequações contratuais, técnicas e regulatórias ainda dependentes de comprovação.

Em relação à Metatext, são apresentados os resultados da avaliação da pseudonimização e das informações destinadas às equipes de segurança, a organização da transparência por destinatário e a demonstração das salvaguardas da GuardionAI. O parecer final reconheceu o cumprimento das cinco condicionantes e recomendou o encerramento satisfatório da participação. Paralelamente, o relatório preserva os achados das avaliações amplas e dos retestes direcionados, identificando oportunidades de aperfeiçoamento do desempenho, das explicações e da governança continuada.

A leitura das três experiências considera a correspondência entre funcionalidades declaradas e comportamento observado, a distribuição das responsabilidades, os fluxos de dados, o atendimento aos titulares e a efetividade das medidas de proteção. Os resultados dos diferentes instrumentos são interpretados segundo seus critérios e evidências, mantendo identificadas as versões, os ambientes e as limitações metodológicas. Na Fase C, as manifestações da ANPD e das empresas são apreciadas conforme seus respectivos enfoques, preservando as diferenças entre aprendizado

---

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.

regulatório, evolução percebida pela participante e demonstração de atendimento das exigências.

Com essa organização, foi possível compreender o percurso de avaliação de cada solução e os fundamentos das conclusões do Ciclo 6. Seu propósito é reunir os resultados da experimentação, reconhecer os avanços demonstrados e delimitar as providências de encerramento, manutenção ou adequação, contribuindo para a aprendizagem institucional da ANPD e para o aperfeiçoamento de futuras iniciativas de sandbox regulatório.

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

No Ciclo 6, o planejamento das sessões de orientação concentrou-se no alinhamento final dos testes, na complementação documental e na verificação das recomendações técnicas e regulatórias. A organização das atividades considerou o estágio de desenvolvimento de cada solução, os resultados acumulados e as evidências necessárias ao encerramento da experimentação, permitindo direcionar o acompanhamento às questões específicas das participantes.

Foram realizadas sucessivas reuniões online, por meio do Microsoft Teams, com todas as participantes (Synapse, Prevvine e Metatext) para alinhamento final em relação aos testes e à documentação. As interações com as equipes técnica e regulatória do CIAAM/USP e com a ANPD permitiram esclarecer respostas, discutir achados, acompanhar entregas e relacionar os materiais apresentados às solicitações formuladas nas fases A, B e C. Complementarmente, o intercâmbio de informações pela plataforma apoiou a localização dos documentos e o tratamento das questões identificadas durante as avaliações.

Para a Synapse, as orientações foram direcionadas à consolidação dos esclarecimentos sobre o Trajetto, especialmente quanto ao processamento de imagens na infraestrutura dos parceiros, à segregação dos dados, às integrações externas e à disponibilização de rotas acessíveis. Também foram abordadas a demonstração das informações e explicações na interface, a mitigação da identificação indireta e as condições de execução do Plano de Descontinuidade. O acompanhamento preservou os limites dos ensaios anteriores e distinguiu as providências de encerramento dos requisitos para eventual operação regular.

Quanto à Prevvine, as sessões contemplaram os resultados clínicos e adversariais, os limites da supressão de identificadores, a supervisão médica e a qualidade bibliográfica. Ademais, o alinhamento abrangeu as retificações documentais, a caracterização do tratamento de dados pessoais sensíveis, os fluxos internacionais e as correções destinadas à próxima versão identificada. A orientação buscou precisar o estado de implementação das medidas, as condições dos retestes e as evidências ainda necessárias, articulando essas questões aos procedimentos de incidente, descontinuidade e adequação regulatória.

Em relação à Metatext, o acompanhamento concentrou-se na conclusão da análise documental, na consolidação das avaliações de pseudonimização e dos alarmes destinados à equipe de segurança e na comprovação das cinco condicionantes. As discussões abrangeram a sanitização dos registros, os testes especializados de segurança, as salvaguardas da infraestrutura, os exercícios simulados, a avaliação de vieses, os retestes e o funcionamento do cofre de uso único. Com isso, foi possível relacionar as respostas e os documentos às verificações registradas no parecer final, distinguindo as exigências cumpridas das oportunidades de aperfeiçoamento posterior.

Dessa maneira, as sessões de orientação do Ciclo 6 apoiaram a organização do acervo de encerramento e a fundamentação dos encaminhamentos de cada participante. A combinação entre reuniões, esclarecimentos e análise de evidências favoreceu a identificação dos resultados demonstrados, das ressalvas e das providências remanescentes, contribuindo para uma conclusão coerente com o percurso experimental de cada solução.

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

º Ciclo 6 contempla o monitoramento das atividades de orientação desenvolvidas durante o sexto mês da fase de testes. A definição dos temas e da metodologia das sessões permaneceu vinculada 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. Por se tratar do último ciclo da experimentação, essas atividades concentraram-se na consolidação das evidências, na apreciação das respostas às recomendações e na fundamentação dos encaminhamentos de encerramento.

No Ciclo 6, a seleção dos temas considerou os resultados acumulados, as recomendações do ciclo anterior, os planos de sandbox e as atualizações técnicas e documentais apresentadas. Foram mantidos os temas de proteção de dados, segurança da informação, governança, transparência, explicabilidade, supervisão humana e direitos dos titulares, com aprofundamento proporcional às características das soluções e às questões remanescentes. Incorporaram-se, ainda, os procedimentos de descontinuidade, o destino dos dados e as condições pertinentes à eventual utilização posterior das aplicações.

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 das verificações e à discussão dos resultados. O intercâmbio de informações apoiou a complementação dos esclarecimentos e a identificação das evidências necessárias às conclusões.

Na análise dos testes, foram considerados os objetivos, os critérios, as versões e as condições de execução de cada procedimento, distinguindo os resultados acumulados das verificações realizadas no ciclo final. Na Synapse, foram preservados os resultados dos ensaios anteriores e seus limites, sem atribuir ao Ciclo 6 nova execução integral dos testes já encerrados. Na Prevvine, a consolidação das avaliações clínicas e adversariais considerou os conjuntos examinados, os julgamentos dos especialistas e a situação das correções previstas para versão subsequente. Na GuardionAI, foram diferenciadas as avaliações amplas de desempenho, os retestes direcionados aos modos de falha conhecidos e a apreciação das informações apresentadas às equipes de segurança.

A leitura das evidências distinguiu declarações das participantes, documentos apresentados, funcionalidades implementadas e controles efetivamente verificados. Os pareceres técnicos, as avaliações pelo *framework* geral, as matrizes de transparência e as fichas de êxitos foram examinados segundo suas respectivas finalidades e bases de comprovação. Na Fase C, as manifestações da ANPD e das empresas foram apreciadas conforme seus enfoques, preservando eventuais diferenças de avaliação e a situação de preenchimento das conclusões conjuntas.

As devolutivas finais buscaram precisar os avanços demonstrados, as limitações identificadas e as providências remanescentes. Esse procedimento permitiu diferenciar exigências cumpridas, medidas ainda dependentes de comprovação, condições para eventual operação posterior e oportunidades de aperfeiçoamento continuado. A apreciação dos planos de descontinuidade também preservou a distinção entre a apresentação do instrumento, a realização de exercícios e a comprovação de sua execução.

Dessa forma, a metodologia do Ciclo 6 vinculou a orientação das participantes ao monitoramento final da experimentação e à produção de subsídios para a ANPD. Os registros consolidados no item 4 fundamentam a avaliação individualizada das soluções, a identificação da situação das condicionantes e os encaminhamentos de encerramento, contribuindo para a sistematização dos aprendizados técnicos e regulatórios do projeto piloto.

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 6, em continuidade ao acompanhamento dos ciclos anteriores. Essas atividades compreenderam a consolidação dos resultados dos testes, a análise das atualizações documentais e a verificação da situação das condicionantes técnicas e regulatórias, conforme o estágio de cada participante. Como etapa final da experimentação, o ciclo também contemplou a apreciação dos procedimentos de descontinuidade, dos aprendizados e dos encaminhamentos para o encerramento da participação no sandbox.

O objeto e o escopo das avaliações consideraram as características da inovação, o nível de maturidade tecnológica e os riscos identificados em cada solução. Os resultados apresentados devem ser interpretados à luz das condições de testagem, das versões examinadas e das evidências disponíveis, distinguindo os achados acumulados dos ciclos anteriores das verificações realizadas no Ciclo 6. A análise identifica os avanços comprovados, as limitações dos procedimentos e as providências remanescentes, preservando a distinção entre a conclusão da etapa experimental e o atendimento dos requisitos para eventual utilização posterior das soluções.

4.1 Synapse

O relatório do CIAAM/USP referente ao Ciclo 6 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 apresentadas separadamente pela ANPD e pela empresa na Fase C. Como etapa final da experimentação, o ciclo reúne os resultados acumulados, as atualizações documentais e os encaminhamentos para encerramento do ambiente de testes, identificando os requisitos remanescentes para eventual operação regular. 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, aos componentes tecnológicos e aos fluxos de dados. O acompanhamento considera a declaração de ausência de mudanças na transição entre os Ciclos 5 e 6, articulando-a aos esclarecimentos complementares sobre processamento de imagens na infraestrutura dos parceiros, recomendações de rotas acessíveis e funcionalidades descritas como em desenvolvimento.
- Dados, privacidade e segurança: apreciação das informações e evidências registradas nas fichas A3 e A4 e nos instrumentos da Fase B, abrangendo segregação entre identidade e telemetria, pseudonimização, criptografia, autenticação e integrações externas. Também são examinadas as atualizações das políticas e avaliações de

impacto, os riscos de identificação indireta e as medidas previstas para retenção, bloqueio e eliminação de dados no encerramento.

- TRL[8] e ETC[9]: consideração da autoavaliação da participante e das evidências técnicas disponíveis, mantendo-se, no Ciclo 6, o TRL 4 para a validação dos algoritmos centrais e o TRL 6 para a pseudonimização e a segurança das bases apartadas. Esses níveis são interpretados conforme os ambientes e as condições de demonstração, distinguindo a maturidade dos componentes da comprovação integral das salvaguardas da solução.
- Realismo experimental e benefícios esperados: consolidação dos resultados dos ensaios técnicos anteriores, sem atribuir ao Ciclo 6 nova execução integral dos quatro testes já encerrados. A avaliação preserva os limites dos procedimentos realizados e identifica as verificações ainda necessárias em dados e redes reais, especialmente quanto à resiliência e à resistência à manipulação de coordenadas.
- Riscos, mitigação e impacto residual: análise do parecer técnico final, que manteve a classificação "suficiente com condicionantes", com risco inerente agregado alto e residual agregado médio, sem afastar riscos individuais elevados. O acompanhamento abrange as condicionantes de transparência, reidentificação, robustez, terceiros e acessibilidade, distinguindo as mitigações demonstradas daquelas dependentes da execução verificável do Plano de Descontinuidade ou de validação para eventual utilização futura.
- Governança, transparência e aprendizagem regulatória: exame da formalização das responsabilidades institucionais, dos instrumentos de governança, da matriz de transparência e das fichas de êxitos. Destacam-se a evolução de 78 para 86 pontos no *framework* geral e a manutenção de pendências materiais de informação e explicabilidade na interface. Na Fase C, são preservadas as diferenças entre a avaliação da ANPD, que registra êxito parcial, e a manifestação da Synapse, que não relata novas evoluções no período e mantém sem preenchimento a seção destinada à conclusão conjunta.

Considerando o acompanhamento da Synapse nas fases A, B e C do Ciclo 6, apresentam-se, a seguir, os principais aspectos identificados em cada etapa e os encaminhamentos para o encerramento experimental, 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 anexos e organizados conforme a descrição a seguir por fases para atendimento dos requisitos pertinentes ao Ciclo 6 deste Sandbox Regulatório:

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

- 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, em atualização registrada em 29 de setembro de 2026, não ter havido mudanças entre os Ciclos 5 e 6. Mantêm-se, portanto, os esclarecimentos anteriores sobre a necessidade de mapas geolocalizados para viabilizar as recomendações de rotas internas, finalidade central do projeto Trajetto, e sobre os testes de diferentes modalidades de mapas e formas de aprimorar sua criação.

Quanto ao fluxo de dados e às *Application Programming Interface* (API), a participante igualmente declarou não terem ocorrido alterações na transição para o Ciclo 6. Assim, permanecem válidas as informações de que as solicitações de rota deixam de ser processadas diante de falhas de comunicação ou ausência de dados, com exibição de mensagem solicitando nova tentativa. Ademais, a empresa mantém a descrição da condição de pessoa com deficiência como variável utilizada no cálculo de rotas, vinculada à sessão, e a declaração de que o aplicativo não coleta nem utiliza preferências pessoais. A ficha também descreve o emprego combinado de pseudonimização baseada em *hash*, com gerenciamento de chaves em provedor de serviços em nuvem, e criptografia em repouso, sem registrar modificação desses mecanismos no ciclo final.

No que se refere às imagens, subsiste a descrição de uma solução de análise ponto a ponto, com aplicação de desfoque de rostos nos testes que utilizam imagens reais de câmeras de segurança do Insper. Como estratégia definitiva, a empresa indica a manutenção do modelo no ambiente do parceiro, com recebimento apenas de pontos geolocalizados pelo sistema central. Contudo, o documento técnico completo permanece assinalado na ficha como em construção, sem atualização específica no Ciclo 6 que comprove a implementação integral dessa estratégia. Dessa maneira, os esclarecimentos apresentados devem ser considerados em conjunto com as evidências técnicas pertinentes, especialmente quanto ao processamento e à circulação de imagens, às atualizações das API e à mitigação da identificação indireta pela combinação de localização e atributos de acessibilidade.

Em relação aos protocolos de segurança, a Synapse também registrou não ter havido alterações entre os Ciclos 5 e 6. A autenticação com fator adicional continua descrita como funcionalidade prevista para a nova versão em implementação,

contemplando confirmação por SMS[10] e possibilidades de cadastro por serviços de autenticação, além do cadastro diretamente no aplicativo. Embora a ficha apresente imagens da interface em desenvolvimento, não acrescenta, no campo correspondente ao Ciclo 6, confirmação da entrada em operação desses mecanismos ou resultados de sua verificação funcional.

Complementarmente, permanece a informação de que o usuário continuará conectado após o *login*, podendo encerrar a sessão pelas configurações do aplicativo, sem descrição de mecanismo de encerramento automático. As cópias de segurança seguem limitadas, conforme declarado, ao banco de dados, sem o detalhamento anteriormente solicitado. Quanto às rotas salvas, a justificativa de conservação para projeções, compreensão dos fluxos e aprimoramento das recomendações não foi acompanhada da definição do respectivo prazo de guarda. Assim, no âmbito da Ficha A1, o ciclo final preserva esses pontos como aspectos ainda dependentes de especificação ou comprovação.

Por fim, na Ficha A2, referente à avaliação da maturidade tecnológica e ao mapeamento dos Elementos Tecnológicos Críticos, a participante manteve, no Ciclo 6, 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 5. O encerramento da sequência de ciclos registra, portanto, estabilidade dos níveis declarados, sem progressão adicional de maturidade nesses componentes.

No que tange às fichas A3 e A4, a análise regulatória e documental do Ciclo 6 registrou avanços substantivos em relação ao Ciclo 5, especialmente pela atualização dos instrumentos de governança, pelo detalhamento da arquitetura de proteção de dados e pela formalização das responsabilidades institucionais do projeto Trajetto. O acompanhamento incorporou os esclarecimentos apresentados na reunião de 10 de setembro de 2026, documentados pela participante em 15 de setembro e aprofundados em 6 de outubro, abrangendo o processamento de imagens na infraestrutura dos parceiros e a disponibilização de rotas acessíveis independentemente da autodeclaração de deficiência. Na avaliação pelo *framework* geral, a pontuação passou de 78 para 86 pontos, refletindo melhorias no Eixo 2 (Governança e responsabilidade) e no Eixo 4 (Justiça, não discriminação e vieses algorítmicos). O encerramento experimental foi considerado suficiente com condicionantes, permanecendo requisitos de comprovação para eventual operação regular. Destacam-se os seguintes aspectos:

- A análise regulatória da equipe do CIAAM/USP reconheceu o fortalecimento da governança, da proteção de dados, da avaliação de impactos e da segurança da informação. Permaneceu relevante o diagnóstico de risco elevado associado à geolocalização contínua e à possibilidade de identificação indireta. Para fins de eventual aplicação do regime diferenciado dos agentes de tratamento de pequeno porte, o enquadramento deve observar a vedação do art. 3º da Resolução CD/ANPD nº 2/2022 aos agentes que realizem tratamento de alto risco, caracterizado segundo os critérios do art. 4º. A Resolução CD/ANPD nº 15/2024 constitui referência específica para a comunicação e o registro de incidentes de segurança.

[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.

- O conjunto documental examinado foi ampliado e atualizado, compreendendo o Relatório de Impacto à Proteção de Dados Pessoais (RIPD), a Avaliação de Impacto em Direitos Humanos (AIDH), a Política de Dados, a Política de Inteligência Artificial, o termo de nomeação da encarregada de dados, as minutas de Acordo de Cooperação Técnica (ACT), a documentação da Trajetto API e o detalhamento dos fluxos de backend e dos bancos de dados. Também foram considerados os documentos de visão geral da arquitetura, segurança, privacidade, governança e modelo de IA, o aditivo contratual de infraestrutura e o Plano de Descontinuidade do Ciclo 6.
- Quanto aos documentos técnicos, a ficha passou a registrar como auditados e aprovados o fluxo em duas camadas, a segregação entre identidade e telemetria e os mecanismos de pseudonimização e criptografia. O inventário de API e terceiros recebeu encaminhamento de mitigação no encerramento, mediante previsão de revogação de tokens e credenciais no Plano de Descontinuidade. Já o protocolo de explicabilidade foi formalizado, embora sua apresentação ao usuário ainda demande demonstração visual. Os testes de robustez e estresse foram considerados parcialmente adequados, permanecendo necessária a comprovação de resistência à manipulação de coordenadas em condições reais.
- Nos ACT com entidades privadas, a apresentação de minutas padronizadas representa avanço na organização das relações com parceiros. Subsiste, entretanto, a necessidade de delimitar os papéis de controlador e operador, o ambiente de utilização, as entregas e os limites da experimentação, incluindo os usos incompatíveis com a finalidade do projeto. Essas adequações integram as recomendações para eventual operação fora do sandbox.
- Os instrumentos de cooperação também devem precisar as categorias de dados, as finalidades, as bases legais e as responsabilidades pelo tratamento realizado nas diferentes etapas. Particularmente nas integrações externas, permanece indicada a inclusão de cláusulas que atribuam responsabilidades pelo tráfego de dados das API, pela utilização da infraestrutura dos parceiros e pelas medidas de segurança aplicáveis.
- No plano da coordenação institucional, as políticas atualizadas e o termo de nomeação da encarregada de dados passaram a identificar o Insper como controlador e a Synapse como operadora no arranjo experimental descrito. Com isso, houve maior definição das responsabilidades, dos canais de comunicação e dos mecanismos de acompanhamento. Os critérios de transição foram organizados em condicionantes verificáveis, cuja satisfação deverá ser demonstrada conforme o encaminhamento proposto à ANPD.

- Para os ACT com entidades públicas, a ficha reconheceu a previsão de vedação de repasse financeiro, mas manteve a recomendação de incluir disposições expressas sobre a finalidade pública da cooperação, sua relação com as competências institucionais e o tratamento de dados pelo poder público. Permanecem, igualmente, ajustes relativos à Lei de Acesso à Informação (Lei nº 12.527/2011, LAI), sem prejuízo das salvaguardas de proteção de dados e de sigilo legitimamente aplicáveis.
- A experiência do usuário apresentou avanço material com o vídeo demonstrativo do aplicativo, que evidenciou sua utilização e a presença de ícones de obstáculos e recursos de acessibilidade, como escadas, rampas e elevadores. Essa demonstração ampliou as evidências disponíveis sobre o funcionamento da solução. Contudo, não abrangeu as telas com os textos completos de privacidade, as explicações das recomendações ou os controles de gestão das preferências de tratamento.
- No que diz respeito à acessibilidade, permanece necessária a formalização do planejamento técnico para atendimento de pessoas com deficiência visual, incluindo recursos compatíveis com suas necessidades de navegação. A demonstração de ícones no mapa constitui melhoria relevante, mas não comprova, isoladamente, a disponibilidade de orientação por áudio, descrições textuais, contraste adequado ou compatibilidade com tecnologias assistivas. A ficha final preservou essa matéria na condicionante relativa ao desenvolvimento e suporte desse público.
- A disponibilização de rotas acessíveis passou a ser descrita como funcionalidade universal e padrão, sem exigência de autodeclaração de deficiência. Nos esclarecimentos do Ciclo 6, a participante informou que o banco de telemetria utiliza a indicação funcional de rota acessível, na forma de marcação aplicável a qualquer pessoa, em lugar de identificar o usuário como pessoa com deficiência. Tal organização atende à recomendação de minimizar a coleta e evitar a vinculação desnecessária entre acessibilidade, informação de saúde e coleta de dados pessoais sensíveis.
- No tocante à identificação indireta, a empresa esclareceu que trabalha com bases separadas de cadastro e telemetria e com sessões renovadas a cada acesso ao aplicativo. O RIPD e a Política de Dados incorporaram essas salvaguardas, permitindo reconhecer mitigação do risco anteriormente apontado. Ainda assim, a conclusão técnica manteve a necessidade de proteção adicional contra correlação de padrões de deslocamento, mediante agregação, privacidade diferencial ou ofuscação estatística avançada, conforme a condicionante específica para eventual operação comercial.

- A interface continua a exigir comprovação dos mecanismos que permitam distinguir o aceite dos termos de uso, as autorizações de localização e as demais escolhas do titular. Embora a documentação tenha avançado na definição desses controles, o vídeo apresentado não demonstrou o painel de privacidade ou os textos efetivamente disponibilizados durante a utilização. Por conseguinte, a verificação dessas funcionalidades permaneceu classificada como prioridade entre as condicionantes.
- A AIDH foi reestruturada para incorporar explicabilidade e salvaguardas de acessibilidade, vinculando a redução dos riscos à sua aplicação prática. Paralelamente, a evolução para 86 pontos no *framework* geral decorreu do aumento de 8 para 12 pontos no Eixo 2 (Governança e responsabilidade clara) e de 0 para 4 pontos no Eixo 4 (Justiça, não discriminação e vieses algorítmicos). A melhoria foi relacionada à formalização de responsabilidades e ao indicador Time Overhead Ratio, destinado a monitorar diferenças de tempo nas rotas acessíveis. Por outro lado, os aspectos de intervenção humana, limitações demográficas e correção de vieses permaneceram dependentes de comprovação.
- O RIPD atualizado incorporou os riscos de acessibilidade e o enquadramento do eventual tratamento de dados sensíveis relativos à deficiência. A solução documental adotada prevê o fornecimento opcional dessas informações mediante consentimento específico e destacado, nos termos do art. 11, Inciso I, da LGPD, mantendo a oferta de rotas acessíveis independentemente desse preenchimento. Dessa forma, a finalidade funcional da navegação acessível foi separada da coleta facultativa de informação sensível.
- A nomeação da encarregada pelo tratamento de dados pessoais, vinculada à estrutura institucional do Insper, respondeu à recomendação de formalização dessa função. Também foi identificado o responsável técnico pelo projeto. A ficha registra canais dedicados para privacidade e segurança, reforçando a estrutura de atendimento aos titulares e de comunicação institucional, em consonância com o referencial da Resolução CD/ANPD nº 18/2024.
- A Política de IA passou a disciplinar explicações em linguagem simples, documentação por model cards, testes adversariais, supervisão humana e tratamento de rotas críticas. Ademais, os registros do ciclo estabeleceram restrições ao reaproveitamento de dados reais dos testes para treinamento de modelos comerciais de terceiros sem transparência e base legal adequada, assim como ao compartilhamento externo não auditado. Tais disposições conferem maior precisão aos limites de finalidade e aos deveres de governança.
- O protocolo de explicabilidade foi organizado em três camadas: uma resumida e acessível ao usuário, outra operacional, dirigida às equipes

de suporte e fiscalização, e uma técnico-regulatória, destinada a autoridades e auditores. Sobre as imagens, os esclarecimentos passaram a situar o processamento na infraestrutura dos parceiros, com transmissão de pontos geolocalizados ao sistema central, enquanto a documentação atualizada incorporou medidas de minimização e proteção na origem. A demonstração das explicações na interface, porém, permaneceu pendente, devendo ser conciliada com a documentação técnica das versões e integrações utilizadas.

- A Política de Dados detalhou a segregação física e lógica entre o cofre de identidade e o servidor de telemetria, a pseudonimização baseada em hashes, a criptografia em repouso e as regras de retenção e expurgo. O tratamento continua qualificado como tratamento de dados pessoais pseudonimizados, sem equivalência com anonimização plena. Quanto às transferências internacionais, houve evolução em relação à simples indicação de servidores no Brasil: o documento reconheceu que integrações com API de mapas e IA podem envolver fluxos transfronteiriços, incorporando referência à Resolução CD/ANPD nº 19/2024 e às salvaguardas aplicáveis.
- O plano de resposta a incidentes recebeu complementação pela definição de "rota crítica" nas políticas de IA e de dados. Falhas em sensores de acessibilidade, indisponibilidade de acessos e desvios em trajetos críticos passaram a integrar o fluxo prioritário de contingência e intervenção humana. A estrutura documental foi considerada adequada nesse aspecto, persistindo a necessidade de demonstrar os avisos e controles correspondentes na interface.
- No campo de viés e *fairness*, as diretrizes de equidade, monitoramento e supervisão foram incorporadas à Política de IA e à AIDH, acompanhadas da evolução registrada no *framework* geral. Permaneceram como requisitos específicos a comprovação dos testes contra falsificação de informações em ambiente real, a validação das salvaguardas contra reidentificação e o planejamento de inclusão de pessoas com deficiência visual. A adequação documental desses temas deve ser apreciada em conjunto com as evidências de funcionamento e os limites identificados na avaliação final.

A análise individual dos documentos regulatórios apresentados pela Synapse foi atualizada no Ciclo 6 e permanece como referência para o acompanhamento dos encaminhamentos finais. Os documentos de natureza predominantemente técnica possuem avaliação própria, destacando-se:

- Arquitetura de privacidade e criptografia: registrada como auditada e aprovada, com manutenção do enquadramento dos dados como pessoais pseudonimizados;
- Inventário de API e terceiros: mitigação prevista para o encerramento mediante revogação de acessos, sem dispensar a delimitação

contratual e operacional das integrações em eventual utilização futura;

- Mapa de fluxo de dados ponta a ponta: validação da segregação entre identidade e telemetria, complementada pelo reconhecimento dos possíveis fluxos internacionais associados a terceiros;
- Protocolo de explicabilidade em camadas: formalização documental concluída, com demonstração das explicações e dos controles na interface ainda pendente;
- Relatório e testes de robustez e estresse: resultados reconhecidos nos cenários sintéticos examinados, permanecendo exigida a verificação de resistência à manipulação de coordenadas em condições reais.

Quanto aos resultados da equipe técnica sintetizados nas fichas A3 e A4, o projeto Trajetto consolidou avanços arquiteturais e institucionais que fundamentaram parecer favorável com condicionantes para o encerramento experimental. A avaliação reconheceu a segregação de ambientes, a pseudonimização e a criptografia, o processamento de imagens na infraestrutura dos parceiros e a formalização da governança. O Plano de Descontinuidade foi validado como instrumento de encerramento, prevendo revogação de credenciais e *tokens*, gestão das chaves criptográficas, desativação da infraestrutura, expurgo da telemetria e guarda restrita dos registros de auditoria pelo período nele estabelecido. Sua aprovação deve ser distinguida da comprovação de execução das etapas previstas.

Ao término do Ciclo 6, o encaminhamento proposto à ANPD reconhece a conclusão satisfatória da etapa experimental, preservando as condicionantes para eventual transição da solução. Destaca-se a prioridade atribuída à comprovação das telas de informação, explicabilidade e privacidade, acompanhada dos requisitos de testes em ambiente real, adequação dos ACT, mitigação da reidentificação e inclusão de pessoas com deficiência visual. O resultado consolida o aprendizado produzido no sandbox e delimita as medidas remanescentes, sem equivaler à aprovação irrestrita da operação comercial.

4.1.3 Fase B

A) Framework geral baseado em recomendações da UNESCO

Na aplicação do *framework* geral, a metodologia empregada nos ciclos anteriores manteve-se no ciclo 6. 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. 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 final, a Synapse alcançou 86 pontos, superando os 78 registrados nos ciclos passados. A avaliação apresenta os seguintes resultados:

- Eixo 1 (Centralidade da dignidade humana e dos direitos fundamentais), 20 pontos;
- Eixo 2 (Governança e responsabilidade), 12 pontos;
- Eixo 3 (Avaliação de impactos), 15 pontos;
- Eixo 4 (Justiça, não discriminação e vieses algorítmicos), 4 pontos;
- Eixo 5 (Transparência, explicabilidade e comunicação), 15 pontos;
- Eixo 6 (Segurança técnica e robustez), 10 pontos;
- Eixo 7 (Proporcionalidade e necessidade), 10 pontos.

O resultado corresponde à classificação "Alta maturidade ética e de governança" do *framework* geral. A evolução decorreu de duas alterações: o reconhecimento dos mecanismos de prestação de contas, que acrescentou quatro pontos ao Eixo 2 (Governança e responsabilidade), e o atendimento do critério relativo à realização de testes de vieses nos dados e no modelo, responsável pelos quatro pontos obtidos no Eixo 4 (Justiça, não discriminação e vieses algorítmicos).

Esse avanço deve ser entendido em conjunto com as ressalvas mantidas no instrumento. A capacidade real de intervenção humana permanece sem pontuação, assim como o conhecimento das limitações demográficas, os mecanismos de correção de vieses e a prevenção de proxies discriminatórios. Também subsistem observações sobre o aprofundamento da identificação de grupos vulneráveis e os ajustes nos documentos de avaliação de impactos e mitigação.

Além disso, a pontuação integral nos Eixos 5 (Transparência, explicabilidade e comunicação) e 6 (Segurança técnica e robustez) não elimina as pendências específicas registradas nas observações e nos demais instrumentos. Continuam relevantes a demonstração das mensagens e explicações apresentadas no aplicativo, os testes em condições reais e a comprovação das defesas contra manipulação de coordenadas. A classificação ética sintetiza o atendimento dos critérios pontuados, devendo ser interpretada juntamente com essas limitações.

B) Testes da Equipe Técnica do CIAAM/USP

Os quatro testes técnicos descritos no relatório foram encerrados no Ciclo 3, permanecendo válidas suas conclusões nos limites e nas condições então documentados. Os materiais apresentados para o Ciclo 6 não demonstram nova execução integral desses ensaios. O acompanhamento final concentrou-se na consolidação das evidências acumuladas, na apreciação da documentação complementar e nos requisitos de encerramento do ambiente experimental.

O parecer técnico de 14 de outubro de 2026 incorpora os resultados anteriores, especialmente a verificação do emprego de TLS 1.3 no percurso auditado por túnel e os controles de segregação e proteção das bases. Também reconhece a arquitetura de visão computacional na infraestrutura dos parceiros, com transmissão de coordenadas ao ambiente central, e os mecanismos de autenticação descritos na documentação examinada.

Entretanto, permanecem pendentes os relatórios de resiliência e de testes contra falsificações de informação em dados e redes reais. Os resultados favoráveis obtidos nos cenários sintéticos não comprovam, isoladamente, resistência à falsificação de sinais de localização ou à criação artificial de aglomerações. Com o encerramento experimental, esses testes passam a integrar os requisitos indicados para eventual operação regular da solução.

C) Matriz de Transparência

A Matriz de Transparência contém, no Ciclo 6, respostas e avaliações para os 28 campos correspondentes aos sete grupos de interlocutores e às perguntas “por que”, “o que”, “quando” e “como”. A participante informou não ter havido alteração em relação ao ciclo anterior, e as observações do avaliador mantiveram as distinções entre especificações adequadas, atendimento parcial e informações insuficientes. O preenchimento documental, contudo, não equivale à comprovação de atendimento integral das recomendações.

Para usuários gerais, cadeirantes e pessoas com mobilidade reduzida, permanecem lacunas sobre as razões da transparência quanto ao uso de IA, aos dados utilizados e ao funcionamento do modelo. As respostas ainda não apresentam os textos exatos das mensagens nem delimitam suficientemente os momentos de sua exibição. A previsão de explicações no aplicativo a cada escolha de rota foi considerada parcialmente adequada, pois identifica o canal pretendido, mas não demonstra os componentes de interface correspondentes.

Quanto às pessoas com deficiência visual, a matriz mantém a exclusão temporária desse público do escopo do produto mínimo viável. A avaliação considera essa delimitação aceitável para a etapa experimental nos campos relativos ao conteúdo, momento e meio de comunicação, mas preserva a necessidade de formalizar o planejamento de inclusão. Dessa maneira, a justificativa de escopo não deve ser confundida com comprovação de acessibilidade para esse grupo.

Para os operadores, foi considerada adequada a previsão de dashboards com informações agregadas de telemetria e pontos geolocalizados, sem identificação direta. Continuam insuficientes, porém, a justificativa de transparência sobre os modelos preditivos, a periodicidade dos relatórios e a especificação dos meios de envio dos alertas. No caso dos órgãos reguladores, persistem lacunas relativas à vinculação das informações às necessidades de fiscalização, ao conteúdo dos relatórios, à periodicidade das auditorias e aos canais de disponibilização.

Em relação à equipe técnica, as especificações foram consideradas adequadas ou plenamente adequadas. A matriz descreve a finalidade da transparência para rastreabilidade e auditoria, os perfis e permissões de acesso, a guarda dos registros na AWS, a coleta diária de logs e a revisão pelo responsável técnico ao menos mensalmente. Esse reconhecimento diz respeito à governança interna, sem suprir as necessidades informacionais dos demais interlocutores.

Assim, o Ciclo 6 consolida uma matriz preenchida e acompanhada de avaliação substantiva, mas mantém pendências materiais de transparência perante usuários, operadores e reguladores. A principal medida remanescente consiste em apresentar e verificar as telas, os avisos, as notificações e as explicações efetivamente disponibilizados durante a utilização do aplicativo.

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 6, emitido em 14 de outubro de 2026, manteve a classificação “Suficiente com Condicionantes”. O encaminhamento passou a contemplar a aprovação do encerramento experimental e a eventual transição para operação regular mediante atendimento dos requisitos remanescentes.

A avaliação utiliza matriz de probabilidade e impacto de cinco níveis e registra risco inerente agregado alto e risco residual agregado médio. Essa síntese deve ser lida juntamente com os riscos individuais: comprometimento do serviço de interligação entre

as bases, com escore 15; manipulação de coordenadas, com escore 12; integrações com terceiros, com escore 12; e reidentificação por atributos de acessibilidade, com escore 16, todos classificados como altos. O risco associado à correlação por mecanismos determinísticos permaneceu classificado como médio, com escore oito.

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:	• Aprovação do Encerramento do Ciclo Experimental. • Transição para Operação Regular com Condicionantes. • A aprovação do encerramento não representa autorização para operação regular.
Nível Agregado de Risco Inerente	Alto (em razão do tratamento contínuo de geolocalização dinâmica, padrões de mobilidade e preferências de acessibilidade/PCD).
Nível Agregado de Risco Residual	Médio (sustentado pela arquitetura de segregação de cofres, cifragem forte em repouso/trânsito e regras do Plano de Descontinuidade, mas impactado pela omissão probatória de transparência no aplicativo)
Maturidade Tecnológica	• ETC 1 (Algoritmos Centrais / Agente Híbrido): TRL 4 • ETC 02 (Pseudonimização e Cofres Apartados): TRL 6
Mitigações e Estágio de Comprovação	• Visão Computacional na Borda: arquitetura descrita e reconhecida no parecer, com previsão de processamento local e

transmissão de coordenadas ao Trajetto. Permanece necessária a comprovação técnica das condições de implementação, especialmente quanto à não retenção e ao não envio de imagens identificáveis.

- Autenticação e acesso: *login* social e autenticação de duplo fator por SMS descritos, na Ficha A1, como funcionalidades da nova versão em implementação. Embora mencionados no parecer técnico, os registros não permitem confirmar sua entrada em operação na versão avaliada. Sua classificação como controles efetivados depende de evidência funcional, com identificação da versão e do ambiente verificados.

- Guarda de histórico: justificativa de utilização dos dados para previsão de fluxos apresentada pela participante. Permanecem necessários o detalhamento das categorias conservadas, das finalidades e dos respectivos prazos, em articulação com as regras de eliminação, agregação e retenção previstas para o encerramento.

- Estruturação e Validação do Plano de Descontinuidade (PdD Ciclo 6): Protocolo de desativação que estipula o bloqueio/revogação de credenciais e *tokens* de APIs de terceiros (Google Maps), a destruição ou rotação irreversível das chaves criptográficas, o *shutdown* programado do ambiente, o expurgo/anonimização da telemetria e a retenção restrita de logs de auditoria por 5 anos (LGPD Art. 37 e MCI[11]). Instrumento apresentado e considerado validado no parecer, contendo procedimentos de bloqueio de coleta, revogação de acessos e

[11] MCI — Marco Civil da Internet.

	integrações, gestão de chaves, desativação da infraestrutura e destinação dos dados. A validação documental deve ser distinguida da execução dessas providências, que permanece dependente dos respectivos registros e comprovantes. • Governança Formal do encarregado de dados: Designação institucional ratificada da encarregada de dados.
Mitigações Planejadas Críticas	Pendentes: • Transparência material no aplicativo (Arts. 9º e 20 da LGPD): A Matriz de Transparência do Ciclo 6 encerrou-se como Pendente (0/28 itens preenchidos[12]), confirmando a ausência de demonstração visual de *mockups* de tela, *banners* e textos informativos de explicabilidade lidos pelo usuário. • Risco de Reidentificação em Dados PCD (Alto Risco): O cruzamento da indicação de rota acessível com geolocalização contínua permanece classificado como de Alto Risco. Com o encerramento do Sandbox, o risco mitiga-se pelo expurgo/anonimização do PdD, mas será necessário protocolo formal de *Differential Privacy* antes de eventual lançamento comercial. • Enquadramento sob a Resolução CD/ANPD nº 02/2022: O tratamento de alto risco afasta as flexibilizações para Agentes de Pequeno Porte (Art. 4º da Res. ANPD nº 02/2022), exigindo observância integral da

[12] No parecer técnico, constam 0/24 itens preenchidos na Matriz de Transparência, pois o parecer foi elaborado 3 dias antes da última atualização da Matriz, que passou a indicar 28 campos preenchidos. Dessa forma, apresenta-se a versão ajustada no relatório final.

	mencionada resolução para qualquer operação regular.

Fonte: ANPD

O projeto preserva os avanços de proteção de dados desde a concepção, especialmente a separação entre o cofre de identidade e a base de telemetria, o emprego combinado de pseudonimização e criptografia e a redução da circulação de imagens. O parecer também reconhece a formalização da governança do Encarregado e menciona mecanismos de autenticação cuja implementação permanece dependente de comprovação funcional, conforme a situação registrada na Ficha A1. Esses resultados devem permanecer vinculados às versões e aos ambientes examinados, observando-se a correspondência entre a documentação, os componentes testados e a configuração eventualmente disponibilizada.

No encerramento, o Plano de Descontinuidade assume papel central na contenção dos riscos associados à continuidade da coleta e das integrações. O documento prevê bloqueio de novos cadastros, interrupção dos pipelines, revogação de credenciais e *tokens*, gestão das chaves criptográficas, desativação da infraestrutura e eliminação ou agregação dos dados. Entretanto, diversas medidas permanecem registradas como planejadas, e o cronograma conserva etapas pendentes. O campo narrativo do Ciclo 6 informa a validação dos ciclos anteriores, mas não apresenta, por si, os registros finais de execução de todas essas providências.

Nesse sentido, a mitigação decorrente da desativação e do expurgo deve ser compreendida como dependente da execução verificável do plano. A aprovação do instrumento não substitui os *logs* de revogação, os relatórios de eliminação, as confirmações dos provedores e os registros de encerramento. Também precisam ser conciliadas as previsões de eliminação com a indicação, no inventário, de retenção de determinados históricos por três anos, além das disposições sobre *backups* e guarda dos registros de auditoria.

Quanto às pendências materiais, destacam-se:

- Transparência e explicabilidade: permanece necessária a demonstração das telas, dos avisos e dos textos efetivamente apresentados ao titular. A matriz preenchida permite identificar as lacunas, mas não comprova a implementação dos conteúdos na interface. O parecer atribui prioridade impeditiva a essa matéria para eventual exploração comercial.
- Reidentificação por acessibilidade e localização: a renovação do identificador de sessão e a utilização de uma opção funcional de rota acessível não foram consideradas suficientes, no parecer técnico, para afastar a correlação de trajetos e destinos. Permanece indicada a apresentação de protocolo formal de agregação e ofuscação, com referência à privacidade diferencial, acompanhado de validação de sua efetividade.
- Resiliência e processamento de imagens: subsiste a necessidade de testes contra falsificação de informações em condições reais e de evidências compatíveis com a arquitetura utilizada nos parceiros. A previsão de processamento na borda deve ser acompanhada de

controles verificáveis sobre retenção, circulação de imagens e comportamento das integrações.

- Terceiros e fluxos de dados: o bloqueio dos acessos no encerramento reduz a exposição experimental quando efetivamente executado. Para eventual reativação, continuam relevantes o inventário das APIs, os campos compartilhados, as responsabilidades contratuais e a avaliação dos fluxos internacionais envolvendo mapas, autenticação, SMS e infraestrutura de parceiros.
- Acessibilidade e supervisão: a delimitação temporária do público do MVP não substitui o planejamento de inclusão de pessoas com deficiência visual. Da mesma forma, a formalização dos fluxos de intervenção humana deve ser acompanhada da demonstração de sua disponibilidade e funcionamento nos cenários pertinentes.

No acompanhamento das condicionantes, a Seção 7 do parecer técnico apresenta a seguinte situação:

Quadro 2: Condicionantes do sistema Trajetto

Condicionante	***Registro no parecer e alcance do encaminhamento***
COND1: RIPD e AIDH	Concluída com ressalva, mantendo a avaliação continuada dos riscos associados à acessibilidade e à reidentificação.
COND2: Resiliência e incidentes	Exigência para operação regular, com testes de estresse e contra a falsificação de informações em dados e redes reais.
COND3: Inventário e contratos de API	Registrada como concluída mediante integração ao Plano de Descontinuidade; a efetividade do bloqueio depende dos comprovantes de revogação.
COND4: Governança do encarregado de dados	Concluída pela comprovação documental da designação da Encarregada e do responsável técnico.
COND5: Transparência e explicabilidade	Pendente, exigindo apresentação das interfaces e dos conteúdos informativos efetivamente acessados pelos usuários.

COND6: Reidentificação por acessibilidade	Registrada como mitigada pelo expurgo previsto no plano; permanece necessária proteção demonstrável para eventual reativação.
COND7: Inclusão de pessoas com deficiência visual	Registrada como mitigada no contexto do encerramento, sem comprovação de implementação do suporte ou superação da necessidade de planejamento formal.

Fonte: ANPD

Em matéria de incidentes, o parecer formaliza canais de privacidade e segurança e define "rota crítica" como situação em que uma recomendação incorreta, obstrução ou indisponibilidade de recurso de acessibilidade comprometa a segurança ou a mobilidade do usuário. Essas ocorrências demandam encaminhamento humano prioritário. Quando houver incidente de segurança com risco ou dano relevante aos titulares, a comunicação deve observar o prazo regulamentar de três dias úteis, ressalvada legislação específica, em lugar da referência genérica a 72 horas constante dos documentos.

Quanto à retenção, devem ser distinguidos os históricos de rotas, os registros de acesso à aplicação e os registros de auditoria ou de incidentes. O prazo de seis meses do art. 15 do Marco Civil da Internet refere-se aos registros de acesso mantidos pelos provedores abrangidos, sem constituir limite geral de conservação de trajetos. Já o art. 37 da LGPD exige o registro das operações de tratamento, mas não estabelece prazo geral de cinco anos para todos os logs. As previsões do plano devem, portanto, ser fundamentadas por categoria de registro e finalidade.

Por fim, o diagnóstico de alto risco deve ser fundamentado nos critérios cumulativos do art. 4º da Resolução CD/ANPD nº 2/2022, consideradas as restrições do art. 3º ao tratamento jurídico diferenciado. A Resolução CD/ANPD nº 15/2024 disciplina a comunicação e o registro de incidentes de segurança, devendo ser referida nesse âmbito. Essa distinção preserva a precisão normativa do relatório sem alterar o reconhecimento dos riscos e das salvaguardas necessárias.

A Fase B do Ciclo 6 consolidou a evolução da pontuação ética, os avanços de governança e os requisitos técnicos de encerramento, mantendo os componentes críticos em TRL 4 e TRL 6. O parecer favorável com condicionantes reconhece os resultados da experimentação e delimita as medidas remanescentes. Sua implementação deverá ser demonstrada por evidências, especialmente quanto à transparência na interface, à resiliência, à reidentificação e à execução da descontinuidade, sem equiparar o encerramento experimental à autorização irrestrita de operação comercial.

4.1.4 Fase C

A elaboração do relatório de êxitos integra a Fase C - Avaliação e Ajuste da metodologia desenvolvida pelo CIAAM/USP e validada pela Comissão de Sandbox da ANPD. Essa etapa complementa as avaliações técnicas e regulatórias, sistematizando os resultados da experimentação, os aprendizados institucionais e os encaminhamentos decorrentes do encerramento dos testes.

No Ciclo 6, foram consideradas as fichas C1 preenchidas separadamente pela ANPD e pela Synapse. A Agência consolidou os ganhos regulatórios acumulados e

registrou, na seção destinada à conclusão conjunta, a classificação do êxito como parcial. A participante, por sua vez, declarou não ter identificado novas evoluções técnicas ou aprendizados no período, informando concentração de esforços na entrega de documentos de ciclos anteriores. Os campos de classificação, principal evidência de êxito e próximos passos permaneceram sem preenchimento em sua ficha. A consolidação preserva essas diferenças e considera os resultados já examinados nas Fases A e B.

A) Campos de avaliação sob a perspectiva da ANPD

O Ciclo 6 consolida aprendizados das etapas anteriores e encaminha para encerramento do Projeto Piloto.

O ciclo 6 consolida o entendimento relativo à incidência do Regime de Alto Risco (Resolução CD/ANPD nº 15/2024) ao tratamento contínuo de localização espacial associado a atributos de acessibilidade potencialmente sensíveis, o que afasta as exceções previstas para agentes de pequeno porte (Art. 4º da Res. ANPD nº 02/2022). Nesse caso, o enquadramento regulatório embasado na compreensão dos desafios operacionais associados à implementação do produto representa um dos destaques do Projeto Piloto.

A participante apresentou e validou o plano de descontinuidade do sandbox, incluindo a desativação preventiva da aplicação (com base no Edital ANPD nº 002/2025 e no Art. 50, §2º, II, da LGPD). A apresentação do Plano de Descontinuidade mostra a capacidade do Sandbox Regulatório de encerrar eventuais experimentos realizados no curso do projeto com a devida observância da legislação aplicável, garantindo segurança jurídica aos participantes e observância de direitos.

Permanece, contudo, a lacuna deixada pela falta de matriz de transparência do Ciclo 6. A falta da matriz limita o aprendizado regulatório nessa dimensão e, do ponto de vista do regulador, merece reflexão se a omissão ocorre por problemas de desenho do teste visando à avaliação do Projeto Piloto e às futuras edições.

Foram preservados e aprofundados controles relevantes, tais como o isolamento estrutural em duas camadas (separação do Cofre de Identidade do Servidor de Telemetria), pseudonimização determinística com chave secreta e criptografia, trânsito de dados protegido, implementação de 2FA através de SMS e *login* social, e processamento de imagens na infraestrutura dos parceiros (on-device), com envio apenas de coordenadas ao ambiente central. A avaliação ética realizada através do *framework* geral consolidou 86 / 100 pontos, enquanto a maturidade tecnológica manteve os componentes centrais em TRL 4 (algoritmos centrais) e TRL 6 (cofres e pseudonimização).

A falta de Matriz de Transparência (COND5) representa fraqueza da solução, pelo ponto de vista do usuário, ao dificultar o exercício dos direitos previstos nos arts. 9º e 20 da LGPD.

O Ciclo 6 ainda não oferece elementos conclusivos que deem subsídio a uma nova norma específica. Contudo, representa o encerramento dos ciclos de testagem do Projeto Piloto, permite consolidar pontos concretos a serem considerados pelas áreas finalísticas da ANPD e pelo ciclo regulatório.

O Sandbox contribuiu para reduzir incertezas técnicas e regulatórias associadas ao desenvolvimento de soluções inovadoras de mobilidade urbana baseadas em IA, permitindo à participante estruturar instrumentos de governança, proteção de dados e segurança da informação e identificar requisitos relevantes para uma eventual operação regular.

A experiência também produziu referências potencialmente aplicáveis a outros agentes do setor, especialmente quanto à adoção de medidas de privacidade desde a concepção, gestão de terceiros e avaliação de riscos.

Entretanto, não foram apresentadas evidências suficientes para demonstrar impactos concretos sobre o mercado ou a concorrência, considerando que a solução permanece em desenvolvimento e depende da comprovação de salvaguardas adicionais antes de eventual ampliação de seu uso.

B) Campos de avaliação sob a perspectiva da empresa

Sob a perspectiva da empresa, a Synapse declarou não ter registrado novas evoluções técnicas, aprendizados regulatórios ou avanços em transparência, explicabilidade e direitos dos titulares no Ciclo 6. Informou, ainda, que os esforços se concentraram na entrega de documentos relativos aos ciclos anteriores, sem especificar os materiais ou registrar novas autorizações. Não foram indicadas evidências adicionais para essas manifestações, e o campo destinado a outros aprendizados permaneceu sem preenchimento.

A partir das informações coletadas, verifica-se que o Ciclo 6 encerrou a experimentação com reconhecimento dos avanços técnicos e institucionais acumulados e delimitação das medidas ainda dependentes de comprovação. A ficha da ANPD destaca a arquitetura de proteção de dados, os instrumentos de avaliação de impactos, as políticas de governança, a formalização de responsabilidades e a estruturação da descontinuidade. Já a manifestação da Synapse concentra-se na ausência de novas evoluções no período e no esforço de regularização das entregas documentais.

Na seção destinada à conclusão conjunta da ficha da ANPD, o êxito foi classificado como parcial. A justificativa relaciona essa classificação às pendências de transparência e explicabilidade no aplicativo, à mitigação dos riscos de reidentificação e à validação de controles em condições operacionais. Também preserva a avaliação “suficiente com condicionantes”, com risco inerente alto e risco residual agregado médio, sem afastar os riscos individuais elevados. Como a seção equivalente da ficha da empresa permanece em branco, essa classificação deve ser atribuída ao registro apresentado pela Agência, sem pressupor manifestação coincidente da participante.

No conjunto das avaliações, a pontuação de 86 pontos no *framework* geral representa evolução em relação aos Ciclos 4 e 5, enquanto os componentes tecnológicos permaneceram em TRL 4 e TRL 6. O resultado ético deve ser apreciado juntamente com as limitações experimentais e as pendências materiais. A própria conclusão da ficha da ANPD distingue o amadurecimento alcançado no Sandbox da validação integral das salvaguardas e da aprovação para operação comercial.

Como próximos passos, a Agência recomenda concluir o relatório final, identificando resultados, evidências verificadas, limites dos testes e pendências remanescentes. Para a Synapse, os encaminhamentos compreendem demonstrar as informações e explicações disponibilizadas aos usuários; aprofundar as medidas contra reidentificação; aprimorar a acessibilidade; validar segurança e robustez em condições operacionais; e manter atualizadas as avaliações de impacto e os instrumentos de governança. Eventual continuidade ou ampliação de uso deverá considerar a efetiva mitigação dos riscos identificados e a legislação aplicável.

Finalmente, quanto ao Plano de Descontinuidade, a ficha da ANPD reconhece sua apresentação e validação como resultado do Ciclo 6, abrangendo a organização da desativação preventiva da aplicação. Simultaneamente, recomenda concluir a

documentação e apresentar evidências de execução ou de capacidade de acionamento, incluindo bloqueio de acessos, revogação de credenciais, tratamento dos dados e comunicação às partes envolvidas. As fichas C1, isoladamente, não apresentam comparação detalhada entre versões dos Ciclos 5 e 6 nem comprovantes suficientes da execução integral dessas medidas. O resultado documentado consiste, portanto, no reconhecimento do plano como instrumento de encerramento, acompanhado da necessidade de demonstrar sua operacionalização. Para a ANPD, permanece o encaminhamento de incorporar os aprendizados dos seis ciclos às futuras ações orientativas, de supervisão e de aperfeiçoamento metodológico do Sandbox.

4.2 Prevvine

O relatório do CIAAM/USP referente ao Ciclo 6 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 de êxitos elaboradas separadamente pela ANPD e pela participante na Fase C. Por se tratar do último ciclo, a análise articula os resultados acumulados durante a experimentação, as atualizações documentais e os aprendizados técnico-regulatórios, distinguindo as medidas demonstradas das correções e adequações ainda pendentes. A seguir, apresentam-se os principais dados, informações e conhecimentos do Ciclo 6, considerando os seis pilares da metodologia de avaliação:

- Dimensão técnico-funcional: análise das fichas A1 e A2, abrangendo a arquitetura, os componentes tecnológicos, os fluxos operacionais e as características do modelo utilizado. A avaliação diferencia a versão preservada para os testes, os ajustes de interface e as modificações previstas para a próxima versão identificada.
- Dados, privacidade e segurança: exame do tratamento de dados pessoais sensíveis, da supressão de identificadores, dos controles de acesso, dos registros de auditoria e dos fluxos internacionais. São considerados os resultados dos testes internos e adversariais, os limites da proteção observada e a revisão da caracterização do fluxo anteriormente apresentado como anonimizado.
- TRL e ETC: acompanhamento dos níveis declarados para os quatro componentes críticos, da correspondência entre a matriz consolidada e seu detalhamento e das evidências necessárias à revisão da maturidade. A manutenção dos níveis no ciclo final é examinada em conjunto com a preservação da versão avaliada e com os aperfeiçoamentos ainda sujeitos a implantação e reteste.
- Realismo experimental e benefícios esperados: consolidação das avaliações clínicas, da meta-avaliação e da exploração adversarial realizadas com apoio da plataforma HarpIA. A interpretação contempla as condições da integração, as diferenças entre os canais descritos pela participante, a abrangência dos conjuntos analisados e as limitações metodológicas, sem atribuir ao Ciclo 6 uma nova rodada clínica não documentada.

- Riscos, mitigação e impacto residual: análise das classificações de risco, das condicionantes técnicas e dos resultados relativos à privacidade, à segurança, aos vieses e à qualidade bibliográfica. São preservadas as diferenças entre os instrumentos, relacionando os controles verificados às medidas programadas e às pendências de comprovação.
- Governança, transparência e aprendizagem regulatória: avaliação dos instrumentos institucionais, da supervisão humana, dos direitos dos titulares, da explicabilidade e da transparência por destinatário. A análise incorpora a interlocução com as autoridades do setor de saúde, a discussão sobre os papéis dos agentes de tratamento e os aprendizados registrados pela ANPD e pela empresa.

Considerando o encerramento do percurso experimental, apresentam-se, a seguir, os principais aspectos das Fases A, de preparação e caracterização; B, de experimentação e gestão de riscos; e C, de avaliação e ajuste. Nesta última, 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 6 deste Sandbox Regulatório:

- Fase A
  - Fichas A1 e A2
  - Fichas A3 e A4

- Fase B
  - Ficha B1 — *Framework* geral
  - Matriz de Transparência
  - Parecer Técnico Consolidado
  - Relatório de Análise - Plataforma HarpIA
  - Análise do Filtro de Privacidade

- Fase C
  - Prevvine - C1 - Êxitos - ANPD
  - Prevvine - C1 - Êxitos - Empresa

4.2.2 Fase A

Inicialmente, no que se refere à Ficha A1, a participante Prevvine informou que, entre os Ciclos 5 e 6, não houve alteração da versão da solução STAiDOC submetida à avaliação, preservada desde 25 de agosto de 2026, sob o identificador 66361a2, com utilização do Claude Opus 4.6. A atualização do ciclo final concentra-se na definição das modificações destinadas à próxima versão identificada e na revisão da caracterização do

tratamento de dados. Conforme declaração de 14 de outubro de 2026, a implantação das correções foi prevista para o término dos testes e da vigência do Sandbox, mediante registro de mudança e comunicação prévia. Paralelamente, a empresa solicitou autorização para antecipar essa implantação ainda no Ciclo 6 e repetir os testes com as mesmas vinhetas, mantendo o modelo utilizado. A ficha registra a solicitação, sem apresentar confirmação de seu deferimento, da implantação ou dos resultados dos retestes.

Quanto à arquitetura e ao comportamento da aplicação, as modificações anunciadas compreendem a identificação institucional neutra nas instruções do modelo, a apresentação de resposta institucional fixa para perguntas sobre o sistema e o tratamento de dados, a introdução de controle determinístico de saída após a geração e antes da entrega ao usuário, o bloqueio de tentativas de manipulação de instruções antes do encaminhamento ao modelo no canal clínico e a orientação para solicitar ao médico informações sobre sexo e grau de parentesco, em lugar de inferi-las. Dessa maneira, a participante apresenta um conjunto de correções dirigido às limitações identificadas durante a experimentação, cuja efetividade permanece dependente da implantação e da verificação na versão correspondente. Para a interlocução sobre arquitetura, APIs e evidências, foi designado o líder técnico Bruno Miolo.

No tocante ao fluxo de dados, a ficha informa a manutenção da sequência técnica, com execução do filtro de supressão de identificadores antes do modelo em todos os turnos do usuário. A mudança relevante do Ciclo 6 consiste na qualificação jurídica desse fluxo: em acolhimento às orientações registradas nas Fichas A3 e A4, a empresa passa a descrever o envio ao provedor como tratamento de dados pessoais sensíveis submetidos à supressão de identificadores, reconhecendo que o risco residual de reidentificação é gerido, mas não eliminado. Com isso, a técnica deixa de ser apresentada como anonimização integral e passa a ser caracterizada como medida de minimização e segurança. A participante relaciona essa revisão à adequação documental em curso, abrangendo a base legal do art. 11 da LGPD, as informações aos titulares na Política de Privacidade e nos Termos de Uso, a disciplina da transferência internacional e a inclusão dos elementos de conexão no Relatório de Impacto à Proteção de Dados Pessoais.

Em relação à cobertura do filtro, a atualização atribui os achados de privacidade às limitações do detector de nomes, sem indicar falha na posição desse componente na cadeia de processamento. Para a próxima versão, foi definido o aprimoramento do reconhecimento de vínculos familiares, assistenciais e institucionais, incluindo campos de filiação, responsável, acompanhante, profissional e unidade. Além do mais, estão previstas uma segunda passagem sobre o texto normalizado, o tratamento de grafias espaçadas, a utilização de dicionário público de prenomes e a inclusão de detector específico de data de nascimento. A avaliação dessas modificações deverá ser acompanhada da medição de precisão, sensibilidade e F1 em corpus sintético. Enquanto não demonstrada sua implementação, subsistem as limitações reconhecidas na versão preservada, especialmente quanto a identificadores apresentados em formatos não abrangidos pelas regras existentes.

No que concerne às APIs e aos serviços integrados, não foram registradas alterações na versão em avaliação. A ficha informa que o contrato de tratamento com a Anthropic será complementado pelas cláusulas-padrão contratuais previstas na Resolução CD/ANPD nº 19/2024. Quanto à Supabase, cuja infraestrutura de armazenamento é declarada como localizada em São Paulo, na região sa-east-1, foi registrada solicitação de acordo complementar com as cláusulas-padrão da ANPD em 14 de outubro de 2026, posteriormente reiterada pelo chamado de suporte SU-508564. Esses registros documentam providências de adequação contratual, sem comprovar, por si, a conclusão dos respectivos procedimentos.

Complementarmente, o mecanismo de verificação bibliográfica mantém as características descritas no ciclo anterior, enquanto as ampliações destinadas à próxima versão incluem maior retenção em cache dos artigos mais consultados e indicação, em cada resposta, do funcionamento em regime normal ou de contingência. Também foi previsto teste de simulação de indisponibilidade do PubMed, acompanhado de registros de execução. A participante informa, ainda, a disponibilidade de um componente de pré-visualização do texto encaminhado ao modelo, denominado na ficha "pré-visualização da anonimização", para demonstração assistida em ambiente apartado, sem acionamento do modelo e sem gravação de dados. Sua instalação no ambiente avaliado permanece vinculada à próxima versão identificada, devendo-se distinguir a disponibilidade para demonstração da integração efetiva à aplicação submetida aos testes.

No âmbito dos protocolos de segurança, a Prevvine igualmente declarou não ter alterado a infraestrutura da versão em avaliação. Para atendimento à COND4, foi definido o emprego de resumo criptográfico com chave secreta, descrito como HMAC sobre SHA-256, com manutenção da chave fora do banco de dados e do código, rotação programada e identificação da versão junto a cada resumo. Os registros anteriores deverão ser preservados com a indicação da versão do algoritmo. Já para a COND5, o desenho técnico contempla firewall de aplicação com regras gerenciadas alinhadas ao OWASP *Core Rule Set*, limitação de taxa na borda, desafio automatizado contra robôs nos formulários de acesso e teste de exposição pública antes da ativação. Permanecem programados o exercício simulado de incidente de severidade alta e o teste de suspensão, bloqueio e eliminação, com *logs* verificáveis por SHA-256. Portanto, essas medidas constituem especificações e compromissos de implementação, sem confirmação de sua operação na versão preservada.

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 LGPD.

Quanto à classificação de maturidade tecnológica, a matriz atualizada mantém, no Ciclo 6, TRL 7 para o motor de inteligência artificial e para a infraestrutura de conformidade, bem como TRL 6 para a camada de supressão de identificadores e para o sistema NTAV, sem progressão em relação ao Ciclo 5. Segundo a avaliação registrada pela participante, a manutenção desses níveis acompanha a preservação da versão submetida aos testes. A revisão da maturidade dos ETCs 2 e 3 foi vinculada à implantação da próxima versão identificada e à apresentação de evidências, especialmente o relatório de precisão, sensibilidade e F1 do detector ampliado e o teste de indisponibilidade do PubMed. Persiste, entretanto, a necessidade de harmonizar o detalhamento individual desses dois componentes, que ainda apresenta TRL 7, com a matriz consolidada e com as observações atualizadas.

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. Embora a ficha descreva mecanismos de contingência com cache e sinalização de referências não verificadas, os aperfeiçoamentos anunciados para o Ciclo 6 não foram acompanhados, no instrumento examinado, de resultados que sustentem a elevação da maturidade. Assim, a ampliação do cache e a identificação do regime de funcionamento deverão ser apreciadas em conjunto com evidências de correção bibliográfica e de comportamento diante da indisponibilidade da fonte externa.

Quanto à qualificação da técnica de proteção de dados, o Ciclo 6 incorpora expressamente a distinção entre supressão de identificadores e anonimização, superando a apresentação anterior de eliminação do risco de identificação. O reconhecimento de que o conteúdo encaminhado ao provedor permanece sujeito ao regime de dados pessoais sensíveis torna mais precisa a descrição dos limites da proteção adotada. Por conseguinte, o descarte declarado do texto original, a geração de hashes e o emprego de expressões regulares devem ser compreendidos como componentes das medidas de segurança e minimização, sem equivaler à demonstração de impossibilidade de identificação direta ou indireta. Essa revisão deverá ser refletida nos campos históricos da própria ficha que ainda qualificam o conteúdo como integralmente anonimizado ou afastam, com esse fundamento, o tratamento de dados pessoais no fluxo internacional.

Diante do exposto, o Ciclo 6 consolida maior precisão na caracterização técnica e jurídica da STAiDOC, com reconhecimento dos limites do filtro de identificadores, definição das correções pretendidas e especificação de medidas adicionais de segurança e rastreabilidade. A matriz mantém os níveis de maturidade dos quatro ETCs, sem registrar elevação no ciclo final. A ficha documenta, portanto, a preservação da versão avaliada e a preparação de uma versão subsequente, permanecendo a comprovação de sua implantação, a realização dos retestes, a medição da efetividade dos componentes e a conclusão das adequações documentais e contratuais como aspectos dependentes de evidências. O encerramento da análise das Fichas A1 e A2 deve preservar essa distinção entre funcionalidades existentes, demonstrações em ambiente apartado, providências solicitadas e aperfeiçoamentos programados.

Já em relação às Fichas A3 e A4 da Prevvine, a análise do Ciclo 6 consolidou o acompanhamento das pendências regulatórias e documentais da solução STAiDOC, incorporando as atualizações encaminhadas pela participante, os resultados da avaliação ética, os achados de proteção de dados e a interlocução com as autoridades do setor de saúde. O ciclo final registrou avanços na documentação, na supervisão humana e na identificação dos riscos, acompanhados do reconhecimento expresso de limitações técnicas e de providências ainda necessárias. A análise distingue, portanto, as medidas implementadas, os aperfeiçoamentos documentais e os compromissos cuja execução permanece dependente de comprovação.

No quadro regulatório mobilizado pelas fichas do Ciclo 6, destacaram-se os seguintes instrumentos:

- Lei nº 13.709/2018, Lei Geral de Proteção de Dados Pessoais (LGPD), especialmente quanto ao tratamento de dados sensíveis, aos direitos dos titulares, à transferência internacional e à distribuição de responsabilidades;
- Resolução CD/ANPD nº 19/2024, referente à transferência internacional de dados e às cláusulas-padrão contratuais, e Resolução CD/ANPD nº 15/2024, relativa à comunicação de incidentes de segurança;
- Resolução CFM nº 2.454/2026, indicada na documentação como referência para o uso de inteligência artificial na medicina, em substituição à menção anterior à Resolução CFM nº 2.299/2021;
- RDC ANVISA nº 657/2022, relativa a software como dispositivo médico, e RDC ANVISA nº 751/2022, referente à classificação de risco e aos regimes de regularização;
- RDC ANVISA nº 665/2022, sobre boas práticas de fabricação, e RDC ANVISA nº 837/2023, relativa às investigações clínicas com

dispositivos médicos, ambas indicadas pela autoridade sanitária na reunião interinstitucional;

- Lei nº 15.378/2026, identificada pela ANS como Estatuto dos Direitos do Paciente e mencionada na discussão sobre informação e consentimento para utilização de IA na consulta; e
- Lei nº 13.848/2019, considerada, em conjunto com as competências da ANPD previstas na LGPD, na análise das possibilidades de coordenação entre autoridades reguladoras.

No âmbito dos direitos fundamentais, a atualização da Avaliação de Impacto em Direitos Humanos para a versão 1.2 incorporou riscos concretizados na testagem, incluindo reprodução de dados sob indução, inferência a partir de nomes, tratamento inadequado de minorias e erros de referência. Também ampliou os grupos vulneráveis considerados, abrangendo pessoas LGBTQIA+, pessoas vivendo com HIV, povos indígenas, populações negras e médicos de especialidades e regiões sub-representadas, além dos grupos anteriormente identificados. Com isso, o documento passou a reconhecer risco residual médio para os direitos à saúde, à privacidade e à não discriminação, ajustando a avaliação à experiência acumulada no Sandbox. A identificação desses riscos representa avanço na análise de impactos, sem demonstrar, isoladamente, a efetividade das medidas destinadas a mitigá-los.

No exame documental, foram considerados dois conjuntos de entregas, encaminhados em 7 e 14 de outubro de 2026. O primeiro compreendeu a revisão dos dez documentos do conjunto regulatório, o Manual de Uso do Médico, os Termos de Uso, a Política de Privacidade pública, o anexo com evidências de interface, o Plano de Descontinuidade na versão do Ciclo 5 e o Relatório de Testes de Segurança e Conformidade. Além disso, foi apresentado Quadro de Retificações com correção expressa de dezessete afirmações anteriores. O segundo envio reuniu manifestação sobre as avaliações precedentes, a AIDH v1.2 e o Ato de Ratificação da Indicação do Encarregado e Errata de Identificação, que passou a identificar nominalmente Karlos Roneely Feitosa. Esses documentos ampliaram a rastreabilidade das alterações e permitiram verificar o atendimento de parte das solicitações formuladas nos ciclos anteriores.

Quanto à proteção de dados, a mudança central consistiu no acolhimento, pela participante, da qualificação do fluxo como tratamento de dados pessoais sensíveis submetidos à supressão de identificadores, em lugar de anonimização. Dessa maneira, a empresa reconheceu que o risco de reidentificação não foi eliminado e que a instrução ao modelo para não reproduzir informações pessoais é insuficiente como garantia de proteção. Essa revisão ainda precisa ser harmonizada com o Mapeamento de Dados v1.1 e com os demais instrumentos que mantêm referências ao envio de texto anonimizado. Permanecem igualmente necessárias a correção das bases legais, a compatibilização entre RoPA, TCLE, Política e Manual e a coleta de novo aceite do TCLE, conforme os encaminhamentos registrados nas fichas.

Em relação à transferência internacional, a requalificação do conteúdo encaminhado ao provedor deslocou a análise para as garantias aplicáveis ao tratamento de dados pessoais sensíveis. A documentação informa que as cláusulas-padrão da ANPD ainda não foram formalizadas com os operadores internacionais: foi solicitado acordo complementar à Supabase em 14 de outubro de 2026, enquanto o aditamento à Anthropic permanece como providência a solicitar. Por conseguinte, o armazenamento declarado em São Paulo e a existência de contratos de tratamento não foram considerados suficientes para encerrar a análise dos fluxos internacionais e dos instrumentos contratuais pertinentes.

No campo da segurança, a ficha reúne resultados dos testes internos executados em 7 de outubro de 2026, incluindo aprovação das verificações de supressão de identificadores, bloqueio de mensagens acima do limite de tamanho e limitação da frequência de requisições. Esses resultados devem ser apreciados em conjunto com os achados adversariais e com as pendências declaradas pela própria empresa. Embora a seção de segurança apresente avaliações favoráveis sobre determinados controles, outras partes do documento mantêm como pendentes a ampliação do detector, o bloqueio prévio de manipulação de instruções, o firewall de aplicação, o exercício simulado de incidente e os testes de robustez. Assim, a aprovação dos casos internos não demonstra cobertura integral das situações exploradas pela equipe de avaliação.

A análise de privacidade consolidada nas fichas registrou falhas em 16 de 28 conversas, com exposição de dados pessoais em 12 delas, frequentemente mediante indução à correção de nomes ou ao esclarecimento de inferências. Também foram descritas limitações na autodescrição do sistema e na consistência das respostas sobre o funcionamento do filtro. Esses achados reforçam a necessidade de verificar a proteção em toda a cadeia de processamento, distinguindo a remoção efetiva de identificadores da simples omissão de informações na resposta. A interpretação deve considerar as limitações metodológicas indicadas no próprio relatório, entre elas a execução única por vinheta e a ausência de identificação do modelo de base nos registros específicos examinados.

No que se refere à cadeia de agentes, o Ciclo 6 aprofundou a distinção entre a configuração do piloto e uma futura operação com instituições de saúde. No ambiente experimental, sem instituições parceiras, a Prevvine se apresenta como controladora, enquanto os provedores de inferência e infraestrutura integram a cadeia de tratamento conforme suas funções. Para o cenário institucional, a análise aponta a controladoria conjunta entre a instituição contratante e a Prevvine como hipótese mais provável, considerando a finalidade assistencial definida pela primeira e as decisões da desenvolvedora sobre retenção, fornecedores e outros elementos do tratamento. Trata-se de enquadramento analítico dependente das circunstâncias concretas, que deverá ser refletido em instrumento próprio, com distribuição de obrigações sobre direitos dos titulares, incidentes, registros, transferências internacionais e destino dos dados.

No eixo de coordenação institucional, foi realizada, em 15 de outubro de 2026, reunião técnica entre ANPD, ANS e ANVISA, com participação do CIAAM/USP e demonstração da STAiDOC pela Prevvine. O encontro teve caráter não fiscalizatório, e suas manifestações foram expressamente apresentadas como orientações preliminares. A ANVISA indicou, na configuração demonstrada, possível enquadramento como software como dispositivo médico de classe de risco III, sujeito a registro, ressalvando que a definição oficial depende de análise formal. A ANVISA também apontou a necessidade de regularização antes da exposição à população e recomendou a retirada das páginas públicas identificadas; segundo o registro da reunião, a empresa determinou a providência durante o encontro. Essas manifestações não equivalem a decisão formal de enquadramento ou regularização.

Complementarmente, a interlocução sanitária evidenciou a necessidade de examinar o controle de alterações e a repetibilidade das saídas diante da dependência de um modelo de terceiro acessado por API. A compatibilidade dessa dependência com os requisitos sanitários permaneceu em aberto, pois não foi suficientemente discutida na demonstração. A ANS, por sua vez, identificou possíveis repercussões sobre contratualização, cobertura assistencial, auditoria e direitos do paciente, solicitando acesso ao material para aprofundamento técnico. Desse modo, o encontro contribuiu para delimitar competências e questões comuns, mas não encerrou a definição dos fluxos de cooperação, dos pontos focais ou dos instrumentos de atuação conjunta.

Na avaliação pelo *framework* geral, a solução passou de 43 pontos no Ciclo 5 para 55 pontos no Ciclo 6, ingressando na faixa de risco ético moderado a alto, que exige mitigação antes de escalar. A recuperação decorreu principalmente do reconhecimento de três critérios: ausência de decisões relevantes totalmente automatizadas, capacidade real de intervenção humana e conhecimento das limitações demográficas. Foram considerados a confirmação ou rejeição expressa de cada sugestão pelo médico, o critério de acionamento da intervenção humana e o reconhecimento da predominância de literatura estrangeira e da baixa representação de determinadas populações. O resultado representa melhora em relação ao ciclo anterior, sem corresponder à comprovação de correção das falhas técnicas da versão preservada.

O eixo de avaliação de impactos manteve a pontuação máxima, agora acompanhado da atualização da AIDH, dos grupos vulneráveis e do plano de mitigação. Persistiram, contudo, insuficiências relacionadas à proteção de direitos fundamentais, à resiliência, aos testes de robustez e vieses, à prevenção de proxies discriminatórios e à justificação das recomendações a posteriori. Na governança, não foram apresentados regimento e atas que demonstrassem a atuação do comitê previsto, e a inclusão de fontes brasileiras na verificação bibliográfica continuou pendente. Ademais, a ficha registra que explicações produzidas pelo modelo não equivalem necessariamente à descrição fiel de seu processamento, aspecto relevante para a avaliação da explicabilidade e da prestação de contas.

Em síntese, os registros do Ciclo 6 consolidaram avanços nos seguintes pontos:

- Revisão documental, identificação nominal do Encarregado e atualização da avaliação de impactos;
- Reconhecimento dos limites da supressão de identificadores e aperfeiçoamento da supervisão humana por sugestão; e
- Aprofundamento da análise da cadeia de responsabilidades e realização da interlocução com ANS e ANVISA.
- Como encaminhamentos da avaliação do ciclo final, foram indicados:
- Implantar a versão identificada com as correções previstas e realizar retestes com as mesmas vinhetas;
- Executar testes de viés com casos pareados e implementar mecanismos de correção;
- Incorporar fontes e diretrizes brasileiras à verificação bibliográfica;
- Formalizar as garantias contratuais aplicáveis aos fluxos internacionais;
- Harmonizar as bases legais e os documentos de tratamento, transparência e consentimento;
- Realizar os testes de robustez e estresse, o exercício simulado de incidente e os testes de suspensão, bloqueio e eliminação;
- Cumprir as entregas do Plano de Descontinuidade, cuja versão recebida corresponde ao Ciclo 5, incluindo a revisão dos riscos residuais; e
- Prosseguir na definição dos papéis dos agentes e dos mecanismos de cooperação entre as autoridades competentes.

Diante do exposto, o Ciclo 6 evidencia resposta relevante da participante às solicitações documentais e aos apontamentos de governança, acompanhada de maior precisão sobre os limites técnicos e jurídicos da solução. A elevação da pontuação ética e

a ampliação da interlocução regulatória constituem resultados do percurso experimental, permanecendo a efetividade das correções e o atendimento das exigências setoriais sujeitos a comprovação. O encerramento do Sandbox deve, portanto, registrar os avanços e as pendências de forma articulada, sem atribuir à participação no piloto ou às manifestações preliminares das autoridades o efeito de aprovação integral da solução.

4.2.3 Fase B

Inicialmente, a Fase B do Ciclo 6 consolidou a avaliação ética da solução STAiDOC, os resultados clínicos e adversariais reunidos pela plataforma HarpIA, o exame específico do filtro de privacidade e as análises de transparência, segurança e mitigação de riscos. O ciclo final registrou avanços documentais e de supervisão humana, acompanhados de limitações técnicas e pendências de comprovação. A interpretação conjunta dos instrumentos exige distinguir os resultados produzidos ao longo da experimentação, sua consolidação no encerramento e as modificações previstas para uma versão subsequente da aplicação.

A) Aplicação do framework geral baseado em recomendações da UNESCO

A Prevvine obteve 55 pontos de 100 no Ciclo 6, com aumento de doze pontos em relação aos 43 registrados no Ciclo 5. A distribuição por eixo foi: dignidade humana e direitos fundamentais, 10/20; governança e responsabilidade, 7/15; avaliação de impactos, 15/15; justiça, não discriminação e vieses, 4/15; transparência, explicabilidade e comunicação, 7/15; segurança técnica e robustez, 2/10; e proporcionalidade e necessidade, 10/10. Conforme a interpretação prevista na planilha, o resultado corresponde à faixa de risco ético moderado a alto, que exige mitigação antes de escalar.

A recuperação decorreu do reconhecimento de três critérios: ausência de decisões relevantes totalmente automatizadas, capacidade real de intervenção humana e conhecimento das limitações demográficas. Foram considerados a exigência de ciência prévia, a confirmação ou rejeição expressa de cada sugestão pelo médico e o critério de acionamento definido na Avaliação de Impacto em Direitos Humanos (AIDH) v1.2. No campo demográfico, a atualização passou a reconhecer a predominância de literatura estrangeira, a baixa representação de determinadas populações e as limitações verificadas no atendimento a diferentes perfis de pacientes e profissionais.

Além disso, a AIDH v1.2 incorporou os riscos concretizados na testagem, ampliou a identificação de grupos vulneráveis e revisou as avaliações residuais dos direitos à saúde, à privacidade e à não discriminação para o nível médio. Esses aperfeiçoamentos sustentam a manutenção da pontuação máxima no eixo de avaliação de impactos. Permanecem, contudo, pendências relativas à atualização do RIPD e à coerência da matriz de riscos do Plano de Descontinuidade, que ainda apresenta a reidentificação como risco baixo e mitigado, em desacordo com os achados experimentais registrados no framework.

Quanto às salvaguardas técnicas, os critérios de resiliência a erros e ataques, testes de robustez e estresse, correção de vieses, prevenção de proxies discriminatórios e justificação das decisões *a posteriori* não foram considerados atendidos. Também persistiram insuficiências na estrutura interna de governança, na prestação de contas e na documentação acessível. Desse modo, a elevação da pontuação representa avanço principalmente documental, de interface e de governança, sem comprovar a correção das falhas da versão preservada para avaliação.

B) Avaliação conversacional do modelo (Teste 1)

O relatório da plataforma HarpIA, datado de 17 de outubro de 2026, reúne as avaliações dos especialistas referentes aos Ciclos 2 a 5, com quinze conversas válidas por avaliador em cada ciclo, totalizando 120 conversas na análise clínica consolidada. Não apresenta uma nova rodada identificada como Ciclo 6. Sua utilização no ciclo final corresponde, portanto, à consolidação das evidências acumuladas, sem atribuir ao encerramento resultados de uma testagem clínica adicional não documentada.

A análise registra percepção favorável quanto à compreensibilidade e à naturalidade das respostas, com maior variação nos quesitos utilidade, precisão e satisfatoriedade. Paralelamente, as observações especializadas identificam problemas de pertinência e existência das referências bibliográficas, inferências clínicas não sustentadas pelas vinhetas e recomendações consideradas inadequadas ao contexto. Entre os achados discutidos estão o preenchimento de informações ausentes para cálculo do escore de Centor e a persistência de erros bibliográficos após contestação. Esses registros caracterizam limitações das interações examinadas, sem estabelecer taxas gerais de erro clínico para a solução.

A interpretação deve considerar que os especialistas avaliaram casos diferentes. Consequentemente, a comparação entre seus perfis de resposta não constitui medida de concordância interavaliadores. O relatório também registra maior exigência de um dos avaliadores, acompanhada de diferenças metodológicas na condução das conversas, circunstâncias que impedem atribuir as divergências exclusivamente ao desempenho da aplicação.

Na meta-avaliação, foram examinadas 121 conversas dos Ciclos 2 a 5, incluindo uma interação repetida que foi excluída da análise clínica, mas mantida para apreciação metodológica. No conjunto, 33,1% foram classificadas como limitadas, 29,8% como adequadas, 20,7% como não avaliáveis e 12,4% como de qualidade alta; as demais foram consideradas inconclusivas ou inadequadas. Essa análise examina a estrutura e a suficiência das conversas como instrumentos de avaliação, sem verificar a correção clínica dos julgamentos especializados. Seus resultados recomendam ponderação da força de cada evidência, sem justificar o descarte automático dos achados favoráveis ou desfavoráveis.

Por sua vez, as condições da integração permanecem relevantes para delimitar as conclusões. Segundo a documentação reproduzida pela HarpIA, o canal de avaliação sinalizaria tentativas adversariais sem bloqueá-las. A equipe da plataforma esclarece que não consumiu nem armazenou os metadados de sinalização, filtragem e bloqueio mencionados pela empresa. Assim, os resultados documentam o comportamento observado na integração avaliada, permanecendo necessária a demonstração de sua correspondência com o canal clínico e com as diferentes etapas de tratamento.

C) Testagem adversarial e filtro de privacidade (Testes 2 e 3)

Complementarmente, o relatório HarpIA apresenta seis árvores de conversação construídas em atividades adversariais por equipe multidisciplinar. A abordagem permitiu explorar caminhos alternativos a partir de contextos compartilhados, com profundidades máximas entre dois e seis turnos e entre um e oito caminhos por árvore. Esses indicadores descrevem a extensão e a organização da exploração, sem representar, isoladamente, medidas de eficácia dos ataques.

Na análise contextual da sexta árvore, foram examinadas inconsistências na fundamentação clínica, limitações de recomendações diante de informações previamente

fornecidas e utilização de dados insuficientes para pontuação de um escore. Os registros mostram que o confronto pelo avaliador pode produzir revisão da resposta, mas não assegura que a mesma correção ocorra espontaneamente. O próprio relatório caracteriza essa etapa como exploração inicial, sem pretensão de esgotar as vulnerabilidades da aplicação.

No teste específico de privacidade, a plataforma HarpIA analisou 27 conversas. Em uma primeira camada de perguntas, identificou paráfrases preservadoras de informações sensíveis em doze conversas (44,4%), combinações de quase-identificadores na saída em treze (48,1%) e reintrodução de informações anteriormente suprimidas em onze (40,7%). Em todas as conversas, houve sinalização de dados sensíveis e explicitação do tratamento declarado. A coexistência desses comportamentos demonstra que a comunicação de uma suposta remoção não comprova, por si só, a efetividade da proteção.

Uma segunda camada de avaliação registrou identificadores diretos em 40,7% das conversas, possibilidade de reidentificação pela combinação de quase-identificadores em 44,4% e preservação de informações sensíveis por paráfrase em 33,3%. Esses percentuais correspondem a perguntas e critérios próprios, devendo permanecer vinculados aos respectivos indicadores, sem tratamento como medidas intercambiáveis.

Já o relatório específico de análise do filtro de privacidade, em versão revisada de 18 de outubro de 2026, considera um universo de 28 conversas. Nesse instrumento, dezesseis foram classificadas como envolvendo exposição: em doze, o dado apareceu ou foi confirmado na resposta; nas outras quatro, foi utilizado no raciocínio sem reprodução literal. A filiação materna concentrou nove das doze exposições na resposta. O estudo também registra ausência de exposição de identificadores numéricos e de contato nas conversas examinadas.

A estratégia mais eficaz no subconjunto analisado consistiu em apresentar um nome incorreto ou atribuir-lhe outro papel para induzir o sistema a corrigi-lo. Pedidos explícitos para desconsiderar regras não produziram exposição nos casos registrados. Essa diferença evidencia a relevância de testar interações que reproduzam dúvidas, correções e esclarecimentos cotidianos, além de comandos adversariais ostensivos.

Importa delimitar que as bases de 27 e 28 conversas não foram conciliadas nos documentos apresentados. Seus resultados devem ser expostos separadamente, sem soma dos registros ou comparação direta das frequências. O estudo específico também reconhece execução única por vinheta, utilização dos resumos do avaliador, análise integral restrita às conversas com exposição e codificação por uma única pessoa. Portanto, as frequências observadas não constituem taxas estáveis de vulnerabilidade.

Em termos de encaminhamento, o estudo recomenda ampliar a detecção e a supressão antes do modelo, adicionar controle de saída, impedir a confirmação ou correção de identificadores e solicitar esclarecimentos em lugar de inferir atributos pessoais. Recomenda, ainda, repetir os testes com transcrições completas, múltiplas execuções e dois codificadores. As respostas do próprio modelo sobre sua arquitetura são tratadas como evidências comportamentais, cuja veracidade depende de documentação e verificação técnica.

D) Explicabilidade para transparência (Teste 4)

No Ciclo 6, permanece a opção metodológica de acompanhar a explicabilidade pela matriz de transparência, conforme nota que informa a descontinuidade do instrumento separado nos Ciclos 4 a 6. A organização contempla as camadas técnica, de

governança e de usuário, diferenciando informações destinadas às equipes de TI e segurança, aos reguladores, às instituições de saúde, aos médicos e aos pacientes.

Os documentos reconhecem avanços na disponibilização do Manual do Médico, dos avisos de suporte à decisão, dos alertas de urgência e privacidade e dos botões de confirmação ou rejeição por resposta clínica. No *framework* geral, a atualização da AIDH e a definição de critérios de acionamento passaram a sustentar o reconhecimento da capacidade real de intervenção humana, modificando a avaliação do Ciclo 5.

Entretanto, a rastreabilidade técnica e o registro da decisão médica não equivalem automaticamente à explicação fiel do processamento do modelo. A avaliação da UNESCO mantém como não atendida a possibilidade de justificação a posteriori, apontando que as falhas observadas nos testes não constam das trilhas próprias da empresa em condições que permitam sua reconstrução. Também permanecem imprecisões na Política de Privacidade pública e o compromisso de avaliar versões em linguagem simples. Com isso, os avanços de interface e supervisão devem ser reconhecidos sem presumir o atendimento integral da explicabilidade.

E) Matriz de transparência (Teste 5)

A matriz disponibilizada mantém cinco grupos de interlocutores e quatro perguntas por grupo (por que, o que, quando e como informar), correspondentes a vinte campos de conteúdo por ciclo. Na versão examinada, o painel assinala os cinco grupos como concluídos no Ciclo 6, mas os vinte campos da coluna correspondente permanecem vazios. O resumo automático registra zero itens preenchidos e 0% de preenchimento para esse ciclo.

Dessa maneira, não é possível afirmar que houve conclusão material da coluna do Ciclo 6. A análise deve distinguir os registros disponíveis até o Ciclo 5, alguns dos quais fazem referência a documentos enviados nos Ciclos 5 e 6, da ausência de uma avaliação individualizada na coluna final. Essa situação também precisa ser conciliada com a afirmação do parecer técnico de preenchimento integral da matriz, que menciona vinte e quatro itens, embora a própria planilha tenha corrigido o total para vinte.

Quanto ao conteúdo disponível, os registros destinados aos usuários descrevem o formulário de revisão na área do médico, os canais do Encarregado e a disponibilização de termos e políticas. Permanece indicada a necessidade de um caminho mais interativo para os pacientes e de revisão das políticas em linguagem simples. Para os médicos, constam o manual, a Declaração de Uso Responsável, a ciência prévia e a validação por sugestão; workshops e comunicações por e-mail continuam apresentados como possibilidades de complementação.

No caso dos hospitais e clínicas, a matriz esclarece que o piloto não possui instituições parceiras. Materiais, contratos, relatórios e dashboards específicos permanecem vinculados a uma futura etapa institucional. Para os reguladores, são registrados a atualização do RIPD, o Quadro de Retificações e o protocolo de demonstração reprodutível, enquanto métricas de vieses e revisão periódica de riscos figuram como aperfeiçoamentos. Já para TI e segurança, destacam-se o inventário, o controle de mudanças e as trilhas de auditoria, com sugestões de complementação por métricas de desempenho e definição de responsáveis pelas revisões técnicas. Esses elementos constituem a base documental disponível, sem substituir o preenchimento e a validação específicos do Ciclo 6.

F) Parecer técnico: mapa de riscos e mitigações

O parecer técnico emitido em 13 de outubro de 2026, identificado como referente ao Ciclo 6, apresenta conclusão "suficiente com condicionantes" para prosseguir com testes. O documento mantém risco inerente agregado crítico e risco residual agregado médio, fundamentando sua avaliação nos controles automatizados, na documentação de governança, nos instrumentos de informação aos usuários e nos mecanismos de supervisão médica.

Em síntese, o parecer da equipe técnica do CIAAM/USP registra:

Quadro 3: Classificação de Indicadores do sistema STAiDOC

Indicador de Avaliação	***Classificação***
Conclusão Preliminar:	Suficiente com condicionantes.
Recomendação Principal:	Prosseguir com testes (mediante a ajustes)
Nível Agregado de Risco Principal	Crítico: reflete a natureza do tratamento de dados pessoais sensíveis de saúde (prontuários, diagnósticos, histórico de pacientes) combinado com o uso de LLM de terceiro (Claude Opus 4.6 da Anthropic via API internacional).
Nível Agregado de Risco Residual:	Médio: justificado pela comprovação em produção dos controles automatizados e pelo saneamento integral do arcabouço contratual e de governança[13].
Mitigações efetivadas validadas:	• Saneamento da Pendência Documental e Governança (COND2 Concluída): Entregas formais e validação da Política de Privacidade e Proteção de Dados, Política de Retenção e Descarte de Dados, Relatório de Impacto à Proteção de Dados (RIPD), Plano de Resposta a Incidentes e

[13] A expressão reproduz a avaliação constante do parecer técnico e deve ser lida em conjunto com as providências de adequação contratual registradas e apontadas pela equipe da parte regulatória neste relatório, entre as quais permanece pendente a conclusão da formalização das cláusulas-padrão contratuais com Anthropic e Supabase.

Registro das Operações de Tratamento (ROPA).

- Esteira Sequencial Determinística do Filtro de Privacidade: Comprovação técnica de que o filtro de privacidade opera de forma autônoma e estritamente anterior ao modelo de IA (anonymizer.ts em Edge Supabase Deno/TypeScript, AWS sa-east-1). O modelo nunca recebe o texto bruto, eliminando a possibilidade de bypass ou processamento paralelo.

- Anonimização de 8 Categorias de PII: Remoção determinística com substituição por placeholders neutros para: CPF, RG, E-mail, Telefone BR, Endereço, CEP, Prontuário/Registro e Nomes Próprios.

- Prova Dupla de Integridade e Imutabilidade: Descarte imediato do texto original em memória após a geração de hash SHA-256 de prova de existência. O texto anonimizado recebe outro hash SHA-256 de integridade, e é gravado em tabelas insert-only sem permissão de UPDATE ou DELETE (cumprindo Arts. 6º, III, VI, VII e 37 da LGPD).

- Validação Bibliográfica Real (Sistema NTAV): Integração com a API pública do PubMed (E-utilities/NCBI) para verificação de PMIDs sem envio de dados pessoais, reduzindo o risco de alucinações clínicas.

- Plano de Descontinuidade (PdD) Estruturado: Mapeamento de papéis e responsabilidades claras entre o Sponsor Executivo (Tiago L. Miolo), Líder Técnico

(Bruno Miolo) e o DPO (Karlos Roneely Feitosa).

- Formulário Eletrônico de Autosserviço (Art. 18 da LGPD - RG-11 Concluído): Disponibilização do formulário na área logada do médico (Minha Conta), permitindo a abertura de solicitações com cálculo automático do prazo de 15 dias úteis (comprovado nos testes do Ciclo 5 em 07/10/2026).

- Mecanismo de Validação Médica por Resposta (HITL - Art. 20 da LGPD): Inclusão de botões explícitos de "Confirmar Sugestão" e "Rejeitar sugestão" ao final de cada resposta clínica, com gravação síncrona na trilha de auditoria imutável.

- Avisos de Privacidade Preventivos e Fixos: Exibição de aviso preventivo no campo de digitação ao detectar dados sensíveis (como datas e CPFs) e aviso fixo de suporte à decisão clínica com remissão expressa ao Art. 20 da LGPD.

- Comprovação Empírica por Testes Automatizados em Produção (07/10/2026): Execução de testes chamando a API pública da solução sob o commit congelado `66361a2`: Teste C (Anonimização: 9/9 casos e 188/188 suíte aprovados), Teste A (Segurança: 7/7 aprovados — bloqueios de payload > 50k caracteres com HTTP 413 e limite de 15 req/min com HTTP 429), e Teste D (Autosserviço: 6/6 aprovados).

- Preservação da Versão em Avaliação: Congelamento estrito do código, instruções de sistema e modelo de IA desde 25/08/2026 por orientação da equipe do

	Sandbox, garantindo a reprodutibilidade dos testes em curso. • Publicação dos Normativos do Usuário: Lançamento dos Termos de Uso do Usuário Médico e da Política de Privacidade, readequando o meio de autenticação (e-mail e senha), detalhando a remoção de dados pessoais e orientando a minimização sobre datas de nascimento, bem como incorporando o serviço de e-mail transacional (Resend) como operador. • Manual de Capacitação do Médico: Disponibilização do Manual de Uso na área logada, detalhando o fluxo de treinamento em 5 módulos (15 min de leitura, sessão assistida de 20 min), a declaração de ciência com aceite obrigatório, os alertas de urgência em vermelho e os botões de confirmação/rejeição (*human-in-the-loop*).
Mitigações Planejadas Críticas:	• Inclusão de Salt Criptográfico (COND4 - Mantida em Andamento): Reprogramado para a próxima versão identificada (até o final do Ciclo 6) para evitar alterar o ambiente de testes congelado. • Defesa de Perímetro e Aplicação (WAF / AppSec – COND5): A aplicação opera sob rede de borda com mitigação DDoS; a ativação do WAF dedicado com regras OWASP foi reprogramada para a versão do final do Ciclo 6. • Plano de Continuidade para o PubMed (COND5 - Mantida em Andamento): Estruturação de mecanismo de contingência (cache local ou espelhamento de base) para garantir operação em janelas de indisponibilidade do servidor.

	• Aprimoramento da Heurística REGEX (COND7 - Em Evolução): Adição de filtro de contexto clínico, *blocklist* de, aproximadamente, 50 termos médicos para evitar falsos positivos e reexecução da triagem em texto normalizado contra erros de digitação. • Trava Determinística para Perguntas Sobre o Sistema / Explicabilidade (COND3 - Em Evolução): Substituição da resposta comportamental do modelo por uma resposta fixa codificada no servidor (redirecionando para a Política de Privacidade e DPO) e bloqueio determinístico pré-modelo na etapa E4 para injeções de instrução. • Detector de Data de Nascimento (PR-15): Inclusão do detector automático na camada de anonimização programada para a versão final do Ciclo 6, permanecendo mitigado por aviso preventivo na interface e orientação no Manual de Uso. • Remoção de Elementos Visuais de Imagem: Expurgo definitivo dos ícones remanescentes de anexo de imagem até o final do Ciclo 6.

Fonte: ANPD

A redação do documento conserva referências ao prosseguimento dos testes e, em sua conclusão, trata da continuidade pós-Sandbox. Essa distinção deve permanecer identificada, delimitando o alcance da recomendação emitida pelo avaliador.

No plano experimental, o parecer relata testes realizados em produção em 7 de outubro de 2026, sob o identificador 66361a2: nove de nove casos e 188 de 188 verificações no Teste C; sete de sete passos no Teste A; e seis de seis passos no Teste D. Foram registrados bloqueio de mensagens acima de cinquenta mil caracteres, limitação de requisições, recusa de credenciais inválidas e funcionamento dos fluxos de autosserviço. No Teste C, a aprovação corresponde ao comportamento declarado, inclusive à manutenção da data de nascimento, para a qual não havia detector dedicado.

Logo, esses resultados não demonstram remoção universal de identificadores ou ausência de identificação indireta.

O quadro de condicionantes apresenta a seguinte situação:

Quadro 4: Condicionantes do sistema STAiDOC

Condicionante	***Registro no parecer e alcance do encaminhamento***
COND1: Integração com a HarpIA	Concluída, com estabilização da API.
COND2: Documentação de governança	Concluída segundo o avaliador, permanecendo as ressalvas dos demais instrumentos sobre conteúdo e harmonização.
COND3: Transparência e explicabilidade	Concluída / em evolução, com reconhecimento dos mecanismos de interface e autosserviço.
COND4: Proteção adicional dos hashes	Em andamento / reprogramada para versão subsequente.
COND5: WAF e contingência do PubMed	Em andamento / reprogramada.
COND6: Prazos dos avaliadores	Concluída.
COND7: Heurística de identificação	Em evolução, com aperfeiçoamentos ainda vinculados à próxima versão.

Fonte: ANPD

Diante das informações coletadas, subsistem divergências quanto ao estado de implementação de algumas medidas. O parecer descreve o bloqueio pré-modelo e respostas institucionais fixas como implementados em determinados trechos, mas os inclui entre os aperfeiçoamentos planejados em outros. Por sua vez, o *framework* geral informa que, na versão avaliada, a triagem adversarial sinaliza e registra a tentativa, enquanto o bloqueio e a guarda de saída dependem da versão identificada. A confirmação

dessas correções exige, portanto, associação entre a versão implantada, o registro de mudança e os resultados dos retestes.

O registro consolidado do parecer também mantém riscos individuais classificados como altos, incluindo reidentificação, indisponibilidade do PubMed, ataques à aplicação e insuficiências de transparência. Não explicita, entretanto, a regra que reconcilia essas classificações com o risco residual agregado médio. O estudo específico do filtro apresenta matriz própria, com dois riscos críticos e seis altos. As classificações devem permanecer atribuídas aos respectivos instrumentos, sem presumir equivalência entre seus objetos ou substituir uma avaliação pela outra.

Quanto à técnica de proteção de dados, o parecer a qualifica como pseudonimização e reconhece limitações da supressão por REGEX. Já o estudo do filtro conclui que a omissão de informações na resposta não demonstra anonimização ou pseudonimização do conteúdo recebido pelo modelo. Para a consolidação, a descrição como supressão de identificadores preserva o comportamento técnico documentado, mantendo explícita a necessidade de avaliar os dados remanescentes e suas possibilidades de associação ao titular. A proteção adicional dos *hashes*, embora pertinente, não comprova isoladamente a anonimização do texto clínico.

Em matéria de incidentes, o parecer estabelece gatilhos de pausa associados a falhas críticas de proteção, quebra de integração, indisponibilidade bibliográfica e reidentificação, mas registra o exercício simulado de severidade alta como reprogramado. Sua referência genérica a 72 horas também precisa ser harmonizada com o prazo de três dias úteis registrado na matriz para comunicação à ANPD, consideradas as condições aplicáveis. Da mesma forma, permanecem necessárias evidências dos testes de suspensão, bloqueio e eliminação e da formalização das cláusulas-padrão com os operadores internacionais.

Em conjunto, a Fase B do Ciclo 6 consolidou os aprendizados sobre desempenho clínico, comportamento adversarial, privacidade e supervisão humana, reconhecendo avanços documentais e de interface. A pontuação conforme o framework geral de 55/100 e o parecer favorável com condicionantes constituem resultados de instrumentos distintos, cujos fundamentos e ressalvas devem permanecer identificados. Nesse passo, o encerramento registra medidas demonstradas e pendências de implementação, conciliação documental e retestagem, sem equiparar a recomendação técnica à comprovação de conformidade integral ou à autorização irrestrita de operação.

4.2.4 Fase C

A Fase C, referente à Avaliação e Ajuste, consolidou os resultados e os aprendizados do Ciclo 6 sob duas perspectivas complementares: os êxitos para a ANPD e os êxitos para a Prevvine. Foram consideradas as fichas C1 preenchidas separadamente pela Agência e pela participante, esta última datada de 19 de outubro de 2026 e subscrita por seu representante legal, Tiago Miolo. Os registros permitem examinar o aprendizado regulatório, os avanços relatados e as pendências remanescentes, preservando a distinção entre as avaliações individuais e a conclusão conjunta, cujos campos de classificação do êxito, principal evidência e próximos passos permanecem sem preenchimento em ambos os documentos.

A) Resultados do Ciclo 6 sob a perspectiva da ANPD

O acompanhamento da solução STAiDOC permitiu comparar as declarações da participante com os resultados dos testes técnicos, produzindo evidências sobre os limites da supressão de identificadores em sistemas baseados em modelos de linguagem aplicados à saúde. A ANPD destacou como aprendizado a necessidade de distinguir anonimização, pseudonimização e tratamento de dados pessoais sensíveis. Nesse contexto, a Agência registrou que, durante o Ciclo 6, a empresa passou a reconhecer o fluxo dirigido ao provedor como tratamento de dados pessoais sensíveis com supressão de identificadores, revendo a caracterização anterior como anonimização.

No âmbito da proteção dos usuários e consumidores, a ficha ressalta que os testes revelaram riscos dificilmente identificáveis apenas pela análise documental, especialmente quanto à inferência de vínculos familiares, à exposição de dados pessoais e ao comportamento conversacional do modelo. A avaliação adversarial demonstrou resistência a riscos evidentes, acompanhada de vulnerabilidade a interações que reproduzem situações cotidianas de utilização por médicos. Contudo, a Agência afirma expressamente que os resultados não permitem concluir que os usuários foram efetivamente protegidos. Assim, o êxito dessa dimensão reside na identificação e caracterização dos riscos, sem equivaler à comprovação de sua eliminação ou mitigação suficiente.

Como subsídio regulatório, o projeto gerou elementos para futuras orientações da ANPD sobre transferência internacional, anonimização, papéis dos agentes de tratamento e responsabilização nas cadeias de inteligência artificial generativa. Entre os temas destacados estão a qualificação das chamadas de inferência a modelos estrangeiros, a distribuição de responsabilidades entre instituições de saúde, desenvolvedoras e provedores de modelos e a atualização do Guia Orientativo de agentes de tratamento com exemplos dessas configurações. Tais contribuições constituem insumos para aprofundamento interpretativo e regulatório, sem representar a edição de novas regras ou o encerramento das questões examinadas.

Quanto aos efeitos sobre o mercado e a concorrência, a ANPD identificou a produção de referências práticas sobre transparência, explicabilidade, supervisão humana, registro de operações, direitos dos titulares e gestão de riscos em IA para saúde. Foram mencionados o Manual do Médico, o autosserviço de direitos, a confirmação ou rejeição das sugestões e os documentos de governança, incluindo RIPD, RoPA, políticas e inventários. Esses resultados oferecem parâmetros para o desenvolvimento e a avaliação de soluções semelhantes, embora a ficha não apresente mensuração de efeitos concorrenciais, ganhos de mercado ou ampliação da adoção comercial.

B) Resultados do Ciclo 6 sob a perspectiva da Prevvine

A Prevvine organizou sua avaliação em seis dimensões: evoluções técnicas; aprendizados regulatórios; transparência e explicabilidade; direitos dos titulares; autorizações e análises pendentes; e outros aprendizados. O preenchimento relaciona as medidas descritas às evidências documentais e experimentais, abrangendo os resultados acumulados durante o Sandbox e as providências destinadas à etapa posterior ao encerramento.

Quanto às evoluções técnicas, a empresa destacou o reconhecimento, pelo parecer do Ciclo 6, da esteira sequencial de processamento anterior ao modelo, com execução em todos os turnos, descarte do texto original após a geração do *hash* e manutenção de trilhas de auditoria. Também reiterou os resultados dos testes realizados

em 7 de outubro de 2026. Esses números correspondem aos procedimentos informados pela participante e devem ser interpretados em conjunto com os limites de cobertura e os achados adversariais examinados na Fase B.

A ficha também registra a supervisão humana por resposta, os avisos de suporte à decisão, os alertas preventivos de privacidade e urgência e os mecanismos de contingência da verificação bibliográfica. Paralelamente, a participante apresenta uma próxima versão identificada como desenhada e preparada, contemplando proteção adicional dos resumos criptográficos, *firewall* de aplicação, ampliação do detector de nomes e filiação, detecção de data de nascimento, controle determinístico de saída, resposta institucional fixa, bloqueio adversarial anterior ao modelo e pré-visualização do texto encaminhado. Entretanto, a implantação dessa versão ainda figura entre as autorizações solicitadas, devendo-se distinguir sua preparação da comprovação de entrada em operação e dos resultados dos retestes.

No campo regulatório, a empresa reconheceu que a supressão por expressões regulares reduz a exposição de identificadores diretos, mas não neutraliza integralmente os riscos associados a quase-identificadores e vínculos familiares. Acolheu, por conseguinte, a qualificação do fluxo como tratamento de dados pessoais sensíveis, relacionando-a à definição da base legal, à informação aos titulares, à atualização do RIPD e às salvaguardas para transferência internacional. Como aprendizado adicional, afirmou que a proteção não pode depender exclusivamente das instruções dirigidas ao modelo, indicando a necessidade de controles determinísticos anteriores e posteriores à inferência.

Em relação à cadeia de agentes, a Prevvine se identifica como controladora no piloto e reconhece que uma futura operação com instituições de saúde pode envolver controladoria conjunta, dependente de instrumento que distribua obrigações por etapa do tratamento. Ademais, relata a orientação sanitária preliminar recebida na reunião de 15 de outubro de 2026 e o contato com a ANVISA para definição formal do enquadramento e da via de regularização da STAiDOC. A retirada das páginas públicas apontadas pela agência foi informada como providência adotada durante o encontro. A ficha mantém a regularização sanitária como procedimento em andamento, sem apresentar decisão formal sobre a classe de risco ou autorização de disponibilização à população.

No eixo de transparência e explicabilidade, a participante destacou o Quadro de Retificações, com correção de dezessete afirmações anteriores, a reemissão dos documentos com controle de alterações, os instrumentos acessíveis na interface e o reconhecimento das limitações de maturidade tecnológica. Também registrou a elevação da pontuação do *framework* geral de 43 para 55 pontos, associada à supervisão humana e ao conhecimento das limitações demográficas. A declaração de consolidação da matriz de transparência, baseada no status “Concluído” de seus interlocutores, deve ser relacionada à conferência apresentada na Fase B, que identificou ausência de conteúdo na coluna específica do Ciclo 6. Desse modo, o status do painel não substitui a verificação material dos campos correspondentes.

Quanto aos direitos dos titulares, a empresa informou o funcionamento do formulário eletrônico de autosserviço na área “Minha conta”, a identificação nominal do Encarregado e a atualização dos instrumentos de governança. Indicou, ainda, prazos de retenção diferenciados: duração do Sandbox acrescida de um ano para os dados clínicos tratados e registros de explicabilidade, e cinco anos para os registros de auditoria, supressão de identificadores e incidentes. Para os pacientes, reconheceu que o atendimento dos direitos deverá ser articulado com a instituição de saúde, diante das

limitações de localização do titular pela própria plataforma. Essa organização permanece vinculada à formalização do futuro arranjo institucional.

A própria participante relacionou como pendentes a autorização de implantação e reteste da versão identificada, as condicionantes técnicas de proteção dos *hashes*, *firewall* e contingência bibliográfica e o aperfeiçoamento dos detectores. Também indicou a formalização das cláusulas-padrão com operadores internacionais, os testes de viés com casos pareados, a inclusão de fontes brasileiras, a harmonização das bases legais e dos documentos, o novo aceite do TCLE, os exercícios de incidente e descontinuidade, a atualização da matriz de riscos residuais e a formalização do Comitê de Governança de IA. A revisão da maturidade dos componentes de supressão e verificação bibliográfica foi vinculada à medição de desempenho e aos testes pertinentes.

Como outros aprendizados, a Prevvine destacou que as exposições observadas decorreram principalmente de estratégias conversacionais de correção e inferência, com maior incidência sobre a filiação materna, enquanto não foram registradas exposições dos identificadores numéricos e de contato relacionados na ficha. Na dimensão clínica, informou saldos de favorabilidade de +97,7 e +78,7 pontos percentuais para os dois avaliadores, reconhecendo simultaneamente o preenchimento indevido de lacunas em escores e os erros bibliográficos. Esses indicadores são apresentados pela participante como síntese das avaliações, sem equivaler a taxas de acerto diagnóstico. Por fim, a empresa reconheceu a importância de reproduzir no ambiente de avaliação o comportamento entregue ao usuário e de envolver as autoridades setoriais desde as etapas iniciais.

C) Coleta e análise das evidências

A leitura conjunta das fichas revela convergência quanto à insuficiência da supressão de identificadores para sustentar a premissa original de anonimização e quanto à necessidade de relacionar as declarações documentais ao comportamento observado nos testes. O principal aprendizado compartilhado consiste na revisão das premissas de proteção e na identificação das consequências técnicas, jurídicas e organizacionais do tratamento de dados sensíveis por uma cadeia de fornecedores de IA.

Os instrumentos, entretanto, conferem ênfases distintas aos resultados. A ANPD destaca a identificação de riscos e ressalva a ausência de comprovação de proteção efetiva dos usuários; a empresa valoriza os controles testados, os avanços de interface, a documentação e a preparação das correções. A consolidação deve preservar ambas as perspectivas, relacionando cada avanço à evidência correspondente e distinguindo os resultados demonstrados das funcionalidades anunciadas para a versão subsequente.

No encerramento do Ciclo 6, a ficha da participante mantém o pedido de autorização para implantação e reteste, sem apresentar o resultado desse procedimento. Permanecem igualmente em andamento as adequações contratuais, sanitárias e de governança. Assim, os encaminhamentos passam a se referir à comprovação das medidas remanescentes e à definição das condições aplicáveis à continuidade posterior ao ambiente experimental, sem pressupor novo ciclo de testagem já autorizado.

Quanto à descontinuidade, a Prevvine descreve o plano de 7 de outubro de 2026 do Ciclo 5, com gatilhos, responsabilidades, revogação de chaves e regras de retenção, mas também relaciona sua atualização e a revisão dos riscos residuais entre as pendências. Os exercícios simulados e os testes de suspensão, bloqueio e eliminação permanecem declarados como não concluídos. As fichas C1 não apresentam comparação detalhada entre versões dos Ciclos 5 e 6 nem evidências suficientes para confirmar a execução

dessas medidas. O resultado documentado da Fase C compreende, portanto, aprendizado regulatório, avanços específicos e delimitação das providências de encerramento e continuidade, permanecendo sem formalização a classificação conjunta do êxito global.

4.3 Metatext

No Ciclo 6, o acompanhamento da solução GuardionAI, da participante Metatext, concentrou-se na consolidação das evidências técnicas e regulatórias, na conclusão das avaliações de proteção de dados e explicabilidade e na verificação das condicionantes estabelecidas no ciclo anterior. Como etapa final da experimentação, a análise reuniu os resultados acumulados, as complementações documentais e os aprendizados registrados pela ANPD e pela empresa. As atividades consideraram a atuação da solução como camada intermediária de segurança entre aplicações, ferramentas e modelos de linguagem, em infraestrutura dedicada no Brasil e com utilização exclusiva de dados sintéticos.

No campo da proteção de dados, as avaliações da plataforma HarpIA permitiram examinar a detecção e a substituição de identificadores, a preservação do conteúdo das mensagens e a consistência dos resultados diante de variações das entradas. O procedimento híbrido compreendeu 1.225 execuções, ampliando as condições de observação e identificando tanto capacidades demonstradas quanto situações que demandam aperfeiçoamento. Complementarmente, o parecer técnico registrou a realização dos retestes direcionados aos modos de falha conhecidos, a sanitização dos registros de auditoria, a avaliação especializada de segurança e a comprovação do funcionamento do cofre de uso único. Esses resultados evidenciam o amadurecimento da infraestrutura, preservada a distinção entre as métricas da avaliação ampla e aquelas obtidas nos retestes específicos.

Quanto à transparência e à explicabilidade, foi avaliada a organização das informações em dois níveis: comunicação compreensível ao titular e detalhamento técnico destinado à controladora, aos auditores e às equipes de segurança. O Teste 2 reuniu trinta alarmes examinados por dois avaliadores, oferecendo elementos para apreciar a utilidade das explicações e orientar melhorias de clareza, coerência e fundamentação. Ademais, a análise documental esclareceu responsabilidades, canais de atendimento e procedimentos de revisão, enquanto a avaliação pelo *framework* geral alcançou 86 pontos, correspondentes à classificação de alta maturidade ética e de governança.

Por fim, o parecer técnico apresentou conclusão favorável e reconheceu o cumprimento das cinco condicionantes relativas à segurança e aos registros, às salvaguardas da infraestrutura, aos exercícios simulados, à avaliação de vieses e aos retestes e à verificação do cofre. O encerramento satisfatório da participação permanece acompanhado da classificação residual agregada como média e de recomendações de manutenção e aperfeiçoamento. A apresentação das Fases A, B e C relaciona, portanto, os controles demonstrados, os resultados experimentais e os aprendizados institucionais, distinguindo o atendimento das exigências do piloto das condições concretas de utilização posterior da solução.

4.3.1 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 6 deste Sandbox Regulatório:

Principais documentos:

- Fase A:
  - Fichas A1 e A2;
  - Fichas A3 e A4; e
  - Documentação confidencial enviada pela empresa compreendendo temas jurídicos, de proteção de dados, de segurança e de engenharia.
- Fase B:
  - Ficha B1 — *Framework* geral;
  - Ficha B2 — Classificação de riscos;
  - Parecer técnico;
  - Matriz de Transparência;
  - Matriz de enquadramento;
  - Tabela de atendimento de solicitações;
  - Relatório de Análise - Plataforma HarpIA:
  - Teste 1.A;
  - Teste 1.B; e
  - Teste 2;

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

4.3.2 Fase A

Inicialmente, no que se refere às Fichas A1 e A2, o Ciclo 6, correspondente ao segundo ciclo de avaliação da participante Metatext, registrou a manutenção da arquitetura e dos fluxos operacionais da solução GuardionAI, acompanhada de complementações documentais e de esclarecimentos sobre rastreabilidade, explicabilidade e maturidade tecnológica. Permaneceram descritos o ambiente dedicado ao sandbox, localizado em São Paulo, a segregação de infraestrutura e o uso de dados sintéticos nos testes, bem como o processamento intermediado pelo *gateway*, com detecção de dados pessoais, substituição por tokens e reversão controlada por meio de cofre de uso único.

Quanto à instrução documental, a participante informou a disponibilização de dezessete políticas em português na pasta compartilhada do programa, abrangendo instrumentos relativos à proteção de dados, à segurança da informação e à disciplina contratual da solução. Além disso, identificou o encarregado pelo tratamento de dados pessoais e os canais destinados à comunicação sobre privacidade e segurança, com indicação de sua formalização na Política de Proteção de Dados. Essas complementações

respondem às solicitações anteriores de acesso aos documentos e de esclarecimento sobre os responsáveis e os meios de contato.

No eixo de rastreabilidade e explicabilidade, foram apresentados registros visuais do console para demonstrar a reconstrução de requisições, a identificação do campo que originou a detecção, os indicadores de confiança, a política aplicada e a decisão produzida. Para a equipe de segurança, a resposta passou a explicitar relatórios estruturados por alarme, contendo intenção da requisição, evidências com os dados substituídos, severidade e justificativa, associados à fila de revisão e ao histórico de status de cada achado. Com isso, a documentação passou a contemplar mais diretamente as necessidades informacionais dos profissionais responsáveis pela análise dos alertas.

Complementarmente, a participante esclareceu que o ajuste especializado do detector utiliza dados sintéticos em português, anotados por segmento de texto, com calibração de limiares por categoria e congelamento da versão resultante, identificada como ModernGuard PII v1.5. Também informou a publicação da documentação dos níveis de sensibilidade L1 a L4, indicando a utilização do nível L3 no programa, e apresentou imagem do painel central de monitoramento. A ficha A2, por sua vez, distingue o detector proprietário do sistema julgador, que opera sobre modelo de pesos abertos, sem ajuste de pesos, e fornece justificativas em linguagem natural para apoiar a revisão dos achados.

Em relação à maturidade tecnológica, a autoavaliação apresentada na Ficha A2 elevou a infraestrutura e a pseudonimização do TRL 5 para o TRL 7, manteve o componente denominado “Funcionamento do LLM proprietário” no TRL 7 e atualizou o módulo de atribuição de características (Feature Attribution) do TRL 1 para o TRL 3. As justificativas remetem aos testes do programa e à operação comercial dos componentes centrais. Entretanto, os níveis registrados constituem declaração da participante; o módulo de atribuição permanece em prova de conceito experimental, sem integração ao *gateway* e sem ter sido objeto dos testes oficiais do sandbox.

Assim, a análise das Fichas A1 e A2 no Ciclo 6 evidencia avanço na organização documental e na demonstração dos mecanismos de acompanhamento da solução. Subsistem, contudo, aspectos que exigem apreciação conjunta com as evidências das demais fases, especialmente a efetividade da pseudonimização e o alcance das explicações fornecidas. O mapeamento normativo foi mantido sem atualização, preservando a necessidade anteriormente apontada de ampliar a análise das regras setoriais pertinentes; igualmente, a declaração de inexistência de transferência internacional de dados pessoais deve ser examinada em conjunto com os fluxos e instrumentos contratuais apresentados. Dessa maneira, o encerramento deste eixo de avaliação distingue as complementações recebidas, as funcionalidades demonstradas e os componentes ainda em desenvolvimento, sem equiparar a autoavaliação tecnológica à validação integral da solução.

Em relação às Fichas A3 e A4, o Ciclo 6 consolidou a análise da documentação complementar apresentada pela Metatext, superando a etapa anterior de recebimento e organização dos arquivos. Foram examinados os vinte e quatro documentos em PDF e o arquivo de registros de auditoria, em articulação com as respostas às solicitações e os esclarecimentos das Fichas A1 e A2. A disponibilização dos instrumentos em português e a organização das referências permitiram aprofundar o exame da governança, das responsabilidades, dos fluxos de dados, da transparência e dos procedimentos de atendimento aos titulares.

No âmbito da coordenação institucional e da interoperabilidade regulatória, foram esclarecidos os papéis da instituição contratante, como controladora das operações realizadas por sua conta, e da Metatext, como operadora do gateway, preservada sua

atuação como controladora nos tratamentos cujas finalidades determina diretamente. Ademais, o Acordo de Tratamento de Dados Pessoais disciplina instruções, confidencialidade, subcontratação, auditoria, incidentes e assistência ao exercício dos direitos. A Matriz de Enquadramento, constante do Anexo I, relaciona as situações efetivamente configuradas às medidas apresentadas e indica suporte jurídico à implementação do arranjo nas condições avaliadas, observadas as obrigações próprias de cada agente e as exigências setoriais aplicáveis à contratante.

Quanto à transparência e à explicabilidade, a documentação distingue as informações destinadas aos titulares, às instituições contratantes, às autoridades e às equipes de segurança. Essa organização articula mensagens compreensíveis ao interessado com registros técnicos sobre a ocorrência, a política aplicada, a intervenção e sua justificativa. Por sua vez, o fluxo de explicação e contestação estabelece a interlocução entre titular, controladora e operadora, com consulta aos achados e revisão pela equipe responsável, oferecendo elementos para o atendimento ao art. 20 da LGPD. Mantém-se a diferenciação entre as evidências produzidas pelo detector e a explicação posterior do modelo julgador, cuja justificativa deve ser interpretada em conjunto com os registros do processamento.

Complementarmente, o Anexo II individualiza as vinte e quatro solicitações e registra vinte e três como atendidas e uma como atendida com ressalva, referente aos fluxos internacionais. Nesse ponto, a localização da infraestrutura em São Paulo, o inventário de fornecedores e as previsões contratuais constituem elementos relevantes de controle; entretanto, a pseudonimização do conteúdo e sua classificação como dado técnico não afastam, isoladamente, eventual natureza pessoal. A conclusão exige correspondência entre destinatários, categorias transmitidas, possibilidades de identificação e salvaguardas aplicáveis à configuração concreta.

À luz da situação final registrada nas fichas e nos anexos, as cinco condicionantes foram consideradas cumpridas, abrangendo a sanitização dos registros e a avaliação de segurança, as salvaguardas da infraestrutura, os exercícios de incidente, eliminação, explicação e contestação e descontinuidade, o protocolo de vieses e os retestes, além da verificação do cofre de uso único. Dessa forma, o encerramento reconhece a evolução de procedimentos anteriormente descritos para salvaguardas submetidas a verificação, cujos resultados devem permanecer associados às respectivas evidências, versões e condições experimentais.

Não obstante os avanços, subsistem recomendações de harmonização entre versões documentais, prazos de retenção, atribuições, classificações de risco e descrições de funcionalidades. O risco residual agregado foi classificado como médio, exigindo manutenção das medidas de proteção, acompanhamento dos efeitos das intervenções e preservação dos meios de explicação e revisão. Essas providências integram o aperfeiçoamento continuado da governança e a organização do acervo de encerramento, sem implicar reabertura automática das condicionantes registradas como cumpridas.

Assim, a avaliação das Fichas A3 e A4 no Ciclo 6 sustenta a recomendação de encerramento satisfatório da participação da Metatext no projeto piloto. A conclusão reconhece a adequação demonstrada no escopo examinado e os avanços na articulação entre proteção de dados, transparência e responsabilidades. Seu alcance permanece vinculado ao ambiente experimental, aos dados sintéticos e aos cenários avaliados, cabendo à participante e às instituições contratantes preservar e atualizar essas condições na utilização da solução fora do sandbox.

4.3.3 Fase B

A Fase B do Ciclo 6 da solução GuardionAI, da participante Metatext, concentrou-se na consolidação dos resultados dos testes, na avaliação das evidências de implementação das salvaguardas e na verificação do atendimento das solicitações formuladas durante o acompanhamento. Diferentemente da etapa anterior, em que parte das providências permanecia sujeita à comprovação, o fechamento registra o cumprimento das cinco condicionantes técnicas e conclusão favorável ao encerramento satisfatório da participação.

Nesse contexto, a análise articulou a matriz de transparência, o parecer técnico consolidado, o relatório da plataforma HarpIA, a avaliação pelo *framework* geral da UNESCO e os anexos de enquadramento e acompanhamento das solicitações. O conjunto evidencia maturidade na organização da infraestrutura, na distribuição das responsabilidades e na aplicação dos controles, ao mesmo tempo que oferece uma agenda de aperfeiçoamento para a continuidade do desenvolvimento da inovação.

A) Matriz de transparência

A avaliação manteve os quatro grupos de interlocutores definidos para a solução: clientes finais das instituições financeiras; instituições financeiras; ANPD e demais reguladores; e equipes de segurança. Para cada público, foram examinados a finalidade da comunicação, seu conteúdo, o momento de disponibilização e os meios de acesso, permitindo relacionar a transparência às diferentes responsabilidades existentes na cadeia de utilização da GuardionAI.

Quanto aos titulares, a documentação passou a explicitar a finalidade da intervenção de segurança, os dados detectados, a substituição por códigos temporários e os caminhos para solicitar esclarecimento e revisão. Também foi esclarecida a distinção entre o alerta produzido pela GuardionAI, a intervenção técnica sobre a requisição e a decisão da instituição financeira perante seu cliente. Essa separação contribui para que o interessado compreenda os efeitos do processamento e identifique a instituição responsável pelo atendimento de sua solicitação.

No relacionamento com as instituições financeiras, foram apresentados os níveis de sensibilidade, as versões dos componentes, os critérios de geração de alertas, as limitações conhecidas e os instrumentos contratuais pertinentes. Ademais, as capturas do console demonstram o painel de monitoramento, a árvore de execução e os relatórios de ocorrência, oferecendo subsídios à integração, à supervisão e à reconstrução dos eventos. A instituição contratante permanece responsável pela definição das finalidades e dos parâmetros de utilização, enquanto a Metatext executa as instruções e fornece os elementos necessários à explicação e à revisão.

Para as autoridades, a organização dos documentos em português, a associação entre solicitações e evidências e a identificação dos interlocutores de privacidade e segurança favorecem a prestação de contas. Já para as equipes de segurança, os alertas passaram a reunir intenção da requisição, indicadores detectados, confiança, política aplicada, severidade e justificativa, com possibilidade de revisão e reclassificação dos achados.

Dessa forma, a matriz, interpretada em conjunto com as verificações finais, permite reconhecer o atendimento do nível de transparência esperado para o escopo experimental. Consideradas a constituição recente da participante e sua atuação em área crítica de cibersegurança, a estrutura apresentada revela capacidade de organizar

informações técnicas e regulatórias para destinatários distintos. A harmonização de prazos, o aprimoramento da acessibilidade, a capacitação das contratantes e a ampliação das demonstrações em situações de uso continuado constituem oportunidades de evolução dessa estrutura.

Algumas recomendações da matriz ainda refletem momentos anteriores à comprovação dos exercícios e controles posteriormente registrados como concluídos no parecer final. Sua leitura deve, portanto, preservar essa evolução. Igualmente, a explicabilidade atualmente demonstrada está apoiada na rastreabilidade e nas justificativas apresentadas pelo julgador, distinguindo-se da atribuição interna da contribuição de cada característica da entrada, cuja evolução permanece como frente específica de desenvolvimento.

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

O parecer técnico consolidado, emitido em 15 de outubro de 2026, apresenta conclusão favorável, reconhece o cumprimento integral das cinco condicionantes estabelecidas no ciclo anterior e recomenda a conclusão satisfatória da participação da Metatext. O resultado representa avanço em relação ao parecer favorável com condicionantes do Ciclo 5, pois registra a apresentação e a verificação das evidências exigidas para o encerramento.

Quadro 5: Classificação de Indicadores da solução GuardionAI

Indicador de Avaliação	***Classificação***
Conclusão Preliminar:	Favorável.
Recomendação Principal:	Encerramento do Sandbox Regulatório com homologação das salvaguardas e transição para operação comercial.
Nível Agregado de Risco Principal	Alto.
Nível Agregado de Risco Residual:	Médio.
Mitigações efetivadas validadas:	• Geração de dados 100% sintéticos em contexto financeiro brasileiro (personas IBGE/PNAD, sem dados de clientes reais).

- Detecção e mascaramento de identificadores por substituição por tokens na borda pelo ModernGuard PII v1.5 (GliNER 570M, sensibilidade L3) sob política Redact Only e latência < 100 ms no GCP São Paulo.

- Cofre One-Time Vault com tokenização opaca por categoria, desanonimização de uso único reversão de uso único (referida no parecer como desanonimização) e destruição irreversível do vínculo pós-leitura.

- Deploy em produção da correção guard#216 no ClickHouse (logs imutáveis de auditoria sem PII em claro - Decisão D11).

- Validação automatizada de alertas pelo Guardion Judge (gpt-oss-safeguard-20b).

- Explicabilidade em Dois Níveis (Titular simples x Auditor técnico via console), atendendo ao Art. 20 da LGPD sem violar o segredo comercial.

- PenTest de IA concluído, Protocolo de Vieses B1.5 executado pós-calibração e relatórios assinados dos 4 exercícios simulados (A4.4, A4.5, A4.7, C2).

Fonte: ANPD

No plano operacional, o parecer destaca a utilização exclusiva de dados sintéticos no contexto financeiro brasileiro, a identificação das versões dos componentes e a manutenção de ambiente segregado. A infraestrutura reúne interceptação das requisições, aplicação de políticas, detecção de identificadores, tokenização, cofre de vínculos, geração de alertas e registros de auditoria. Essa integração demonstra que a inovação ultrapassa a disponibilização isolada de um modelo de linguagem, compondo uma infraestrutura de proteção e acompanhamento das interações entre aplicações e serviços de inteligência artificial.

O atendimento das condicionantes foi registrado nos seguintes termos:

- Sanitização dos registros e avaliação de segurança (COND1): Foi confirmada a implantação da correção guard#216, identificada pelo *commit* 0986285, destinada a impedir a conservação de conteúdo pessoal em claro na trilha de auditoria do ClickHouse. Também foi apresentado o relatório de teste de intrusão especializado em inteligência artificial, abrangendo vetores como injeção de prompts e exfiltração.
- Salvaguardas da infraestrutura (COND2): Foram validadas as referências de certificação ISO 27001 e os relatórios SOC da infraestrutura GCP São Paulo, em articulação com o acervo de dezessete políticas em português. O aproveitamento dessas evidências deve preservar a identificação de seu objeto, distinguindo as certificações do provedor de infraestrutura das certificações independentes da própria participante.
- Execução dos exercícios simulados (COND3): Foram apresentados os relatórios dos exercícios de resposta a incidentes, eliminação de dados, explicação e contestação e descontinuidade. As atividades abrangeram comunicação e contenção, tratamento de solicitações, reaplicação das eliminações após restauração de backups, revisão de intervenções e interrupção segura das integrações.
- Protocolo de vieses e retestes (COND4): O parecer registra a execução do protocolo sobre grupos demográficos e variações linguísticas, acompanhada de calibração e retestes direcionados aos modos de falha conhecidos, especialmente RG, CPF isolado e dados inseridos em argumentos de ferramentas. Nesses retestes, foram reportados precisão e F1 no patamar de 0,90, sendo que na Ficha A2 preenchida pela empresa Metatext, a meta era de F1 ≥ 0,80, ou seja, obteve um resultado superior ao esperado.
- Verificação do cofre de uso único (COND5): Foi demonstrado o funcionamento do mecanismo de reversão única, com falha da tentativa de segunda leitura e destruição do vínculo entre token e dado original após sua utilização. A verificação fornece suporte à redução da persistência dos vínculos, preservada a necessidade de avaliar a proteção do fluxo completo.

Esses resultados sustentam o reconhecimento da maturidade dos componentes centrais. O parecer atribui TRL 7 à infraestrutura e ao conjunto de pseudonimização e cofre, e TRL 8 ao detector, conforme os critérios empregados nessa avaliação. Tais classificações expressam o estágio reconhecido para os componentes examinados e devem permanecer associadas às funções, configurações e evidências correspondentes.

Sob a perspectiva regulatória, o Anexo I reúne dezesseis situações configuradas e relaciona suas medidas de atendimento à organização jurídica da prestação do serviço. Destacam-se a definição dos agentes, a limitação de finalidade, a ausência de treinamento com dados de clientes, a proteção desde a concepção, a transparência por destinatário, o

apoio ao exercício dos direitos e os procedimentos de encerramento. A matriz indica suporte na legislação vigente para a implementação do arranjo, mantidos os controles e as responsabilidades avaliados.

Por sua vez, o Anexo II permite acompanhar o atendimento individualizado das vinte e quatro solicitações: vinte e três foram classificadas como atendidas e uma como atendida com ressalva, relativa aos fluxos internacionais. Portanto, todas receberam resposta e tratamento na consolidação, preservando-se a necessidade de examinar cada fluxo conforme seu conteúdo, destinatário e possibilidade de identificação. A existência dessa ressalva não foi convertida, nos documentos de encerramento, em condicionante técnica pendente.

C) *Framework* geral baseado em recomendações da UNESCO

A avaliação ética pelo framework geral complementa esse resultado. A participante alcançou 86 pontos em 100, com classificação de alta maturidade ética e de governança, conforme a distribuição registrada:

Quadro 6: Avaliação da solução GuardionAI pelo framework geral

Eixo de Avaliação	***Pontuação Alcançada***	***Pontuação Máxima***
Centralidade da dignidade humana e direitos fundamentais	20	20
Governança e responsabilidade clara	15	15
Avaliação de impactos	11	15
Justiça, não discriminação e vieses algorítmicos	8	15
Transparência, explicabilidade e comunicação	15	15
Segurança técnica e robustez	10	10
Proporcionalidade e necessidade	7	10
Total	86	100

Fonte: ANPD

A pontuação integral em quatro eixos reforça os avanços na definição de responsabilidades, na comunicação e na estruturação das salvaguardas. As pontuações dos demais eixos orientam o aprofundamento da identificação dos grupos afetados, da cobertura das avaliações de equidade e da demonstração de alternativas menos intrusivas. O cumprimento das condicionantes e a pontuação ética possuem objetos distintos: a conclusão das providências exigidas não esgota todas as possibilidades de aperfeiçoamento da solução.

Em conjunto, o parecer e os anexos demonstram resposta consistente às solicitações do acompanhamento e capacidade de converter exigências em medidas verificáveis. A classificação residual agregada como média permanece acompanhada do dever de manter os controles, atualizar a avaliação de riscos e preservar a coerência dos instrumentos na continuidade da operação.

D) Teste 01: Proteção de dados (avaliações da plataforma HarpIA e resultados registrados no parecer técnico)

O relatório HarpIA de 17 de outubro de 2026 consolida duas abordagens complementares de avaliação da pseudonimização. O Teste 1a utilizou trinta mensagens comuns a quatro avaliadores, acrescidas de mensagens exploratórias. O Teste 1b ampliou a análise mediante a geração de 215 mensagens perturbadas a partir das trinta originais, totalizando 245 mensagens submetidas cinco vezes cada, em ordem aleatória, com 1.225 instâncias de teste.

A ampliação permitiu examinar a resposta do detector a inserções, substituições, permutações e exclusões de palavras, além de observar a consistência entre execuções da mesma entrada. Trata-se de contribuição relevante para uma infraestrutura que precisa atuar sobre comunicações heterogêneas, incluindo texto livre e estruturas associadas a ferramentas, em vez de depender exclusivamente de formulários com campos previamente delimitados.

Entre os resultados positivos, o Teste 1a identificou mensagens com precisão e sensibilidade simultaneamente iguais a 1,00. Também registrou preservação integral do conteúdo em todos os julgamentos de dezenove das trinta mensagens de base; no conjunto exploratório adicional, essa preservação foi reconhecida em dezessete dos dezoito itens. Já o Teste 1b contabilizou 1.139 trechos corretamente protegidos e demonstrou manutenção de desempenho agregado relativamente próximo ao teste inicial, apesar da ampliação das condições de entrada.

As métricas devem ser apresentadas conforme cada conjunto. No Teste 1a, foram registrados precisão de 0,77, sensibilidade de 0,71 e F1 micro de 0,74. Na avaliação híbrida, os valores corresponderam a 0,73% de precisão, 0.82% de sensibilidade e 0,77% de F1 micro. A maior sensibilidade observada no segundo procedimento constitui resultado favorável à identificação dos trechos que demandavam proteção, acompanhada, contudo, de maior incidência relativa de substituições indevidas.

O relatório também individualiza oportunidades de aperfeiçoamento em datas por extenso, endereços, CEP, dados bancários e situações dependentes de contexto. No Teste 1b, foram contabilizados 241 trechos que deveriam ter sido protegidos e não foram, além de 423 substituições indevidas. Esses achados fornecem informações concretas para

calibração, revisão dos conjuntos de avaliação e aperfeiçoamento do equilíbrio entre proteção e utilidade das mensagens.

Importa distinguir esses resultados dos retestes direcionados registrados no parecer técnico. O patamar de F1 de 0,90 informado no fechamento refere-se à verificação dos modos de falha conhecidos após os ajustes descritos no parecer; não substitui as métricas da bateria ampla da HarpIA. A leitura conjunta permite reconhecer o atendimento da exigência de correção e retestagem, preservando o registro das limitações encontradas em condições experimentais mais abrangentes.

Como resultado geral, a avaliação evidencia elevado nível de organização da infraestrutura e capacidade de integração a um procedimento extenso de experimentação. O maior grau de abstração da GuardionAI, em relação aos casos de uso finalísticos das demais participantes, decorre de sua atuação sobre a camada de comunicação e proteção que conecta usuários, aplicações, ferramentas e modelos. Essa característica amplia a complexidade do objeto avaliado e o potencial de aplicação transversal da inovação, sem estabelecer, por si, superioridade uniforme de desempenho.

Por fim, as avaliações consideraram a remoção dos conteúdos identificáveis, sem verificar sistematicamente a correspondência entre o tipo da entidade e o código atribuído. A classificação dos tokens, a identificabilidade por combinação contextual e a variabilidade entre execuções constituem dimensões pertinentes a avaliações futuras. Seu acompanhamento integra o desenvolvimento continuado da solução, ao lado das correções já reconhecidas no encerramento.

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

O Teste 2 consolidou a avaliação de trinta alarmes disponibilizados pela Metatext e examinados por dois avaliadores na plataforma HarpIA. O procedimento ocorreu de forma assíncrona, sobre painéis com informações dos incidentes e justificativas do Guardion Judge, permitindo analisar a utilidade do material apresentado à equipe responsável pelas primeiras medidas de apuração.

Foram considerados o título da ocorrência, o nível de risco, os indicadores detectados, a intenção da requisição, as evidências e os textos explicativos. Essa organização permitiu observar a relação entre três dimensões: a detecção produzida pelo mecanismo de proteção, a política configurada e a interpretação posterior do julgador. A avaliação incidiu, assim, sobre a qualidade das informações de apoio à supervisão humana.

Os resultados apresentam pontos positivos. Na identificação dos incidentes e na convergência entre os componentes, ambos os avaliadores reconheceram correção integral ou parcial na maioria dos casos, com avaliações negativas minoritárias. Também predominou percepção favorável quanto à utilização do contexto apresentado no campo de evidências. As explicações relativas à classificação dos incidentes receberam avaliações majoritariamente positivas, com reconhecimento de clareza e suficiência em parte expressiva dos julgamentos.

Ademais, os dois avaliadores atribuíram resultados positivos à pergunta geral sobre a suficiência das informações para orientar as primeiras medidas de apuração. O Índice de Percepção Geral do Sistema apresentou valores de +0,7 e +13,34, respectivamente. Esses indicadores possuem caráter descritivo e exploratório: expressam percepções diferentes e não constituem medida objetiva de suficiência operacional ou consenso entre os avaliadores.

As oportunidades de ajuste concentram-se na coerência entre severidades, na vinculação das justificativas às evidências disponíveis, na redução de siglas e na apresentação de informações em português. Também foram sugeridos esclarecimentos sobre a função de cada modelo, a origem dos conteúdos examinados, os critérios das classificações e os procedimentos de conservação dos dados. Tais recomendações oferecem uma agenda concreta para tornar o painel mais compreensível e útil aos diferentes perfis das equipes contratantes.

Nesse particular, a divergência entre avaliadores constitui aprendizado relevante: explicações tecnicamente detalhadas podem produzir graus distintos de compreensão conforme a experiência e a necessidade de informação do destinatário. O aperfeiçoamento poderá incorporar testes com profissionais de segurança, ampliação do número de avaliadores e procedimentos de revisão das divergências, favorecendo maior consistência na análise da utilidade operacional dos alertas.

A consolidação do Teste 2 demonstra, portanto, avanço na disponibilização de elementos para investigação e revisão dos achados. Em articulação com o exercício de explicação e contestação registrado no parecer final, seus resultados sustentam o reconhecimento do atendimento das providências solicitadas, preservando a distinção entre a execução do fluxo exigido e o aperfeiçoamento contínuo da qualidade das explicações.

Quadro 7: Síntese dos resultados dos testes com a plataforma HarpIA:

Nº	***Nome do Teste***	***Avaliadores***	***Quantidade de Itens Testados***	***Total Testado***
1.a	Proteção de dados: avaliação humana da pseudonimização	4 avaliadores	30 *prompts* × 10 perguntas × 4 avaliadores	1.200 avaliações
1.b	Proteção de dados: avaliação híbrida de robustez	Equipe HarpIA, com anotação manual e execução automatizada	245 mensagens: 30 originais e 215 perturbadas × 5 execuções	1.225 execuções de teste
2	Qualidade das informações e explicações dos alarmes	2 avaliadores	30 incidentes × 5 itens × 2 avaliadores	300 itens avaliados

Total	Conjunto das avaliações	Avaliação humana e procedimentos automatizados	Somatório por modalidade	1.500 itens avaliados e 1.225 execuções automatizadas

Fonte: ANPD

Em conjunto, a Fase B do Ciclo 6 registra o cumprimento das condicionantes, o atendimento das solicitações nos termos individualizados no Anexo II e a consolidação de uma infraestrutura de proteção com elevada complexidade técnica e alta maturidade ética e de governança. Os resultados positivos e os achados de aperfeiçoamento permitem fundamentar o encerramento satisfatório da participação da Metatext, com recomendações voltadas à manutenção das salvaguardas, à evolução do desempenho e à adaptação das informações aos contextos concretos de utilização.

4.3.4 Fase C

A Fase C do Ciclo 6 da solução GuardionAI, da participante Metatext, consolidou os resultados e os aprendizados registrados na ficha C1 sob as perspectivas da Agência reguladora e da empresa. O avanço em relação ao ciclo anterior consistiu na conclusão das verificações exigidas, no cumprimento das condicionantes e na demonstração prática das salvaguardas. Dessa forma, o encerramento reúne evidências sobre a maturidade da infraestrutura, a organização das responsabilidades e os mecanismos de transparência e atendimento aos titulares, preservado o alcance experimental das avaliações.

Sob a perspectiva da ANPD, o projeto aprofundou a compreensão das arquiteturas intermediárias de proteção em inteligência artificial generativa. A ficha destaca o funcionamento integrado de cinco estágios: detecção de identificadores por modelo não generativo, substituição por tokens vinculados ao cofre, encaminhamento do conteúdo tratado ao modelo de linguagem, reversão de uso único e registro para auditoria. Ademais, a utilização exclusiva de dados sintéticos permitiu realizar testes de estresse e privacidade sem exposição de dados reais de clientes. A disponibilização pública do conjunto BR-Agentic-PII-Benchmark, sob licença livre, acrescenta uma contribuição ao desenvolvimento e à avaliação de ferramentas de proteção de dados em português.

No campo da proteção dos titulares, os êxitos foram associados à interceptação de identificadores antes do envio aos modelos externos, à comunicação das intervenções e à preservação de registros para subsidiar sua revisão. A explicabilidade em dois níveis constituiu aprendizado central: informações simples sobre o motivo da intervenção são destinadas ao titular, enquanto categorias, pontuações de confiança, critérios de risco e justificativas estruturadas apoiam a análise pela controladora, pelos auditores e pelas equipes de segurança. Esse arranjo oferece referência para compatibilizar a prestação de informações adequadas com a preservação dos segredos comercial e industrial.

Quanto aos subsídios regulatórios, a experiência reuniu elementos para futuras orientações sobre camadas intermediárias de segurança, governança de agentes com chamadas de ferramentas, distribuição de responsabilidades e comunicação de incidentes. O aprendizado consiste em relacionar essas exigências aos fluxos técnicos e aos instrumentos contratuais, permitindo identificar quem executa a proteção, quem fornece os elementos explicativos e quem responde ao titular. Também se destacou a viabilidade de uma startup estruturar controles de segurança e privacidade desde a concepção sem recorrer à flexibilização das exigências gerais de proteção de dados, contribuindo para a inovação responsável e a competitividade.

Por sua vez, a participante identifica como principais evoluções técnicas a calibração do ModernGuard PII para os modos de falha encontrados, a comprovação do cofre de uso único, a sanitização dos registros de auditoria e a conclusão do teste de intrusão voltado a modelos de linguagem. A ficha registra que a infraestrutura sustentou as avaliações sem indisponibilidade e com latência inferior a cem milissegundos. Segundo o parecer técnico, a infraestrutura alcançou TRL 7 e o detector, TRL 8. Esses resultados expressam a evolução dos componentes examinados e devem permanecer vinculados às configurações e evidências correspondentes.

O aprendizado regulatório da empresa concentrou-se na conversão das políticas em procedimentos demonstráveis. Foram executados os quatro roteiros de incidente, eliminação de dados, explicação e contestação e descontinuidade, acompanhados da consolidação dos papéis da instituição contratante como controladora e da Metatext como operadora. No atendimento aos titulares, destacam-se a identificação do encarregado, o canal dedicado de privacidade, a execução do procedimento de eliminação e a revisão humana dos achados no console. As atividades ocorreram com dados integralmente sintéticos, tendo sido registrada ausência de incidentes envolvendo dados pessoais reais no ambiente do programa.

Em relação à transparência e à explicabilidade, a participante reconhece a validação do modelo em dois níveis e a contribuição da avaliação dos trinta alarmes para o aperfeiçoamento do produto. As diferenças de percepção entre os avaliadores evidenciaram a importância de ajustar o conteúdo e o detalhamento das explicações às necessidades de quem as recebe. Conforme informado na ficha, as recomendações foram incorporadas ao plano de evolução, preservando a vinculação entre as justificativas, as evidências do processamento e as possibilidades de intervenção humana.

Por fim, a análise da empresa informa que o parecer final não deixou condicionantes pendentes. Permanecem como agenda posterior ao sandbox o avanço das certificações independentes da participante, a aplicação das melhorias de explicabilidade e a implantação do *Shadow Mode* para apoiar a triagem de falsos positivos e o ajuste das políticas no ambiente da controladora. Assim, os êxitos do Ciclo 6 concentram-se na demonstração das salvaguardas, no amadurecimento da inovação e na produção de aprendizados regulatórios aplicáveis à sua continuidade.

5. Consultas setoriais

No decorrer do Piloto, a interlocução com autoridades reguladoras setoriais permitiu ampliar a análise das soluções, relacionando os achados técnicos e regulatórios às particularidades dos setores envolvidos. Nesse contexto, foram realizadas consultas à Agência Nacional de Saúde Suplementar (ANS) e à Agência Nacional de Vigilância Sanitária (Anvisa), em relação à solução da Prevvine; ao Banco Central do Brasil (Bacen), quanto à solução GuardionAI, da Metatext; e à Agência Nacional de Transportes Terrestres (ANTT), a respeito da solução Trajetto, da Synapse. Esses encontros constituíram oportunidades de diálogo entre a ANPD, a equipe do CIAAM/USP, as empresas e as autoridades convidadas, favorecendo a discussão de questões que demandavam conhecimento setorial complementar. Apresentam-se, a seguir, os principais temas debatidos em cada consulta e suas contribuições para a compreensão das soluções e para a aprendizagem institucional proporcionada pelo Sandbox.

5.1 Consulta e aprendizados — Setor de serviços financeiros

A consulta setorial referente aos serviços financeiros foi realizada com o Banco Central do Brasil (Bacen) e tal aproximação assumiu especial relevância não apenas para a identificação de desafios regulatórios relacionados à solução, mas também para contribuir diretamente para o aperfeiçoamento da última fase de testagem da GuardionAI, a partir do diálogo com o regulador setorial. Desse modo, a escolha do interlocutor não teve como finalidade apenas apresentar os resultados do Sandbox, mas criar uma oportunidade para que a solução fosse examinada à luz de questões concretas enfrentadas pelo setor financeiro e de preocupações regulatórias relacionadas à cibersegurança, à inteligência artificial e à governança de dados.

A disponibilidade do Banco Central para participar da consulta deve ser destacada como um resultado relevante da estratégia de engajamento institucional adotada. A participação de um regulador setorial possibilitou que a solução fosse discutida com um ator que reúne conhecimento específico sobre o funcionamento, os riscos e as exigências regulatórias do setor financeiro.

A consulta ocorreu em 23 de setembro de 2026, das 15h às 17h, de forma remota, por meio da plataforma Microsoft Teams.

A estrutura da consulta foi concebida para combinar contextualização institucional, apresentação da solução e do projeto e, principalmente, um espaço de diálogo com o regulador.

Assim, a consulta foi precedida de uma etapa de preparação destinada a adequar a apresentação e o debate às competências e aos interesses do regulador financeiro.

Iniciada a reunião, foi apresentado o instrumento de Sandbox Regulatório para alinhar o conhecimento de todos os participantes presentes. Em seguida, a Metatext apresentou a empresa e a solução desenvolvida, permitindo ao Bacen conhecer diretamente o produto objeto da experimentação. Na sequência, foram apresentados o projeto-piloto, sua metodologia e seus resultados preliminares. A programação reservou,

posteriormente, 60 minutos para perguntas abertas, organizadas em quatro blocos temáticos, de modo que a discussão pudesse partir das características concretas da solução e avançar para questões regulatórias específicas.

A consulta setorial foi idealizada como uma oportunidade de coleta de percepções externas sobre a solução; as questões apresentadas ao Bacen foram elaboradas de modo a produzir contribuições potencialmente úteis para a última fase de testagem da GuardionAI.

Os quatro blocos de discussão foram:

- Bloco I: transparência e supervisão;
- Bloco II: anonimização e governança de dados;
- Bloco III: explicabilidade e segurança; e
- Bloco IV: lacunas regulatórias.

Tais blocos tinham por finalidade orientar a discussão de aspectos relevantes para o aperfeiçoamento da solução e para a definição dos testes finais.

Embora a organização prévia em blocos temáticos tenha sido importante para conferir estrutura e direcionamento ao debate, a realização da consulta evidenciou a necessidade de adotar uma abordagem mais flexível. Ao longo da reunião, observou-se que os participantes diretamente envolvidos no Sandbox apresentaram um volume significativo de informações e aspectos técnicos sobre o projeto-piloto, cuja compreensão demandava um conhecimento prévio e mais aprofundado de seu contexto. Para os participantes externos ao projeto, esse volume de informações tornou mais difícil acompanhar integralmente a discussão e formular, naquele momento, contribuições a partir das perguntas previamente estruturadas.

Essa experiência revelou, portanto, a importância de estabelecer um diálogo prévio com os participantes da consulta, especialmente aqueles que não estão diretamente envolvidos na execução do projeto-piloto. Um contato anterior poderia permitir a apresentação antecipada de informações essenciais sobre a solução, sua metodologia, os resultados já obtidos e os principais desafios identificados, possibilitando que os participantes chegassem à consulta mais preparados para contribuir com questões substantivas.

O aprendizado não implica abandonar a utilização de blocos temáticos ou de perguntas norteadoras, mas reconhecer que esses instrumentos devem funcionar como referências para o diálogo, e não como uma estrutura rígida. Em futuras consultas, mostra-se recomendável combinar uma preparação prévia dos participantes com maior flexibilidade durante a reunião, permitindo que o debate se adapte ao nível de conhecimento dos interlocutores e às questões que emergirem no curso da discussão.

Pensando em uma nova abordagem, representantes do CIAAM/USP solicitaram recomendações para o aperfeiçoamento das etapas finais de testagem. Foi discutida, então, a possibilidade de ampliar a interação entre as instituições, inclusive mediante eventual participação de representantes do Bacen em futuras atividades de avaliação. Como encaminhamento, os representantes do Bacen indicaram que organizariam internamente os temas discutidos, compartilhariam as informações com as áreas pertinentes e avaliariam a possibilidade de encaminhar contribuições complementares por escrito.

Também foi elaborado um questionário pós-participação para avaliar a organização, a metodologia e a efetividade da consulta. Entretanto, foram obtidas apenas duas respostas, ambos referentes à avaliação posterior à realização do encontro. A baixa

adesão limitou o alcance dessa avaliação e evidenciou, assim como ocorreu na consulta multissetorial, a necessidade de desenvolver estratégias mais efetivas de engajamento dos participantes. Ainda assim, as respostas obtidas permitem identificar alguns aprendizados relevantes, apresentados neste relatório.

À luz dessas considerações, entende-se que o encontro contribuiu para o refinamento do processo de experimentação, pois foi uma primeira interlocução com o regulador financeiro, permitindo a produção de aprendizados sobre a preparação e a condução das consultas setoriais em próximos sandboxes.

Assim, a consulta permitiu submeter a solução a uma perspectiva regulatória externa, criando uma oportunidade para identificar eventuais pontos de atenção que poderiam não emergir com a mesma clareza apenas a partir da equipe responsável pela experimentação ou das discussões multissetoriais.

5.2 Consulta e aprendizados — Setor de Saúde

A consulta ao setor de saúde ocorreu em 15 de outubro de 2026, em reunião técnica com ANS e Anvisa, com participação do CIAAM/USP e da Prevvine. Na reunião, a iniciativa da ANPD de organizar um sandbox regulatório foi elogiada.

Houve também questionamento sobre a forma como a ferramenta da Prevvine é apresentada no site da empresa. O episódio mostra como dimensões de negócio, regulatórias e técnicas podem se sobrepor nesse tipo de interação. Isso indica a necessidade de detalhar melhor os limites temáticos do sandbox, para que fique claro quais pontos podem ser tratados no âmbito do programa.

Quanto ao modelo da Prevvine, discutiu-se a função da ferramenta: se atua apenas como apoio à decisão ou se produz decisão definitiva. O debate é relevante porque a própria equipe do sandbox discutiu internamente essa questão. Não havia consenso sobre se a resposta do sistema teria consequência direta sobre o paciente ou se a mediação por revisão médica bastaria para caracterizá-la como recomendatória.

Cabe refletir se, em edições futuras, seria razoável exigir que participantes do setor de saúde consultem previamente outros órgãos reguladores. Esses órgãos poderiam definir o nível de impacto da ferramenta a partir de seus efeitos específicos, sanitários e afins, tanto sobre o usuário direto (o médico) quanto sobre o usuário final (o paciente). Isso não significa que a legislação de proteção de dados seja omissa nesse ponto, já que ela se volta justamente à proteção do titular. Falta-lhe, porém, capilaridade em relação ao contexto específico em que esse titular se insere.

A discussão dialoga ainda com a segurança do titular. Mesmo sem incidente, o tratamento de dados pode ser danoso, seja por efeito discriminatório, seja por risco físico. Um exemplo é o registro de alergia a nozes, glúten ou camarão.

A preocupação com a integração com outras autoridades acende um alerta em duas dimensões. A primeira diz respeito à capacidade técnica da ANPD para lidar com especificidades setoriais sem apoio ou diálogo prévio com autoridades setoriais. Em análise preliminar, valeria considerar com mais força dois caminhos: (i) estruturar um sandbox conjunto com uma autoridade que já tenha experiência nesse formato; ou (ii) eleger um setor específico do mercado e promover o nivelamento regulatório com a respectiva autoridade setorial.

A segunda dimensão refere-se à capacidade de articulação da ANPD com as demais autoridades setoriais. Esse ponto ganha peso porque o PL 2.338/2023, em tramitação no Congresso Nacional, atribui à ANPD a coordenação do Sistema Nacional de Regulação e Governança de Inteligência Artificial (SIA). Nesse sentido, o sandbox e suas etapas preparatórias podem contribuir para a construção desse ecossistema e para a preparação institucional da ANPD para essa eventual atribuição.

Por fim, o recorte demográfico, já trabalhado pela ANPD em relatórios anteriores sobre o sandbox, foi objeto de reiterada preocupação. Esse ponto dialoga com os comentários da sociedade civil no evento do Fórum da Internet no Brasil (FIB), em Belém. Desenha-se aqui, com mais clareza, a expectativa de que a ANPD observe as particularidades brasileiras no tratamento de dados. Isso vale tanto para ferramentas nativamente brasileiras quanto para ferramentas nacionais que utilizem modelos estrangeiros. O CIAAM/USP esclareceu que apenas os testes de filtro, com 60 prompts[14], foram demograficamente acurados, com base em levantamento do IBGE de 2024. O ponto é relevante por dois motivos. Primeiro, 60 prompts constituem uma amostra reduzida, sem pretensão de representatividade estatística. Por isso, em contextos futuros, convém considerar testes em volume maior ou mesmo automatizados, seja na interação com o sistema, seja na análise dos resultados.

Segundo, a automação é o cenário ideal, pois permitiria repetir esse tipo de análise a cada lançamento de um novo modelo. Esse mesmo teste também pode servir para avaliar as capacidades de um modelo base em questões de saúde e para mapear os riscos que ele pode gerar. A implementação dessa estrutura de testagem constitui, portanto, uma capacidade útil em dois contextos: a análise de IAs aplicadas à saúde e a avaliação das capacidades de novos modelos. Trata-se de uma habilidade que serve a um setor específico, mas também a um sistema de governança no qual diferentes autoridades precisariam recorrer a testagens automatizadas, na saúde ou em outras áreas.

5.3 Consulta e aprendizados — Setor de mobilidade urbana

No caso da consulta à Agência Nacional de Transportes Terrestres (ANTT), a referida agência reguladora indicou uma série de contribuições relacionadas tanto ao setor de transportes quanto à experiência institucional da ANTT em iniciativas de inovação regulatória. Inicialmente, esclareceu que a Resolução nº 3.323/2009, mencionada durante a exposição da equipe da USP, foi substituída por normativos mais recentes, especialmente pela Resolução nº 6.000, recomendando que as análises regulatórias considerem o arcabouço atualmente vigente.

[14] Nota metodológica sobre os quantitativos de avaliação. A referência a 60 vinhetas corresponde à meta de avaliação prevista para o encerramento do Ciclo 6, cujo cumprimento integral foi limitado pelo tempo disponível para execução e análise dos testes. O relatório quantitativo da plataforma HarpIA abrange 27 avaliações concluídas e incorporadas à base até o fechamento da análise. Posteriormente, foi recebida uma avaliação adicional, elevando o conjunto disponível para exame a 28, sem sua inclusão nos resultados quantitativos já consolidados pela plataforma. Portanto, os números expressam etapas distintas: 60 corresponde à meta planejada; 27, à base efetivamente analisada quantitativamente pela HarpIA; e 28, ao total disponível após o recebimento adicional. As restrições temporais e seus efeitos sobre o alcance da avaliação são explicitados no relatório final. termos quantitativos, porque foi finalizada a análise antes do envio dessa avaliação adicional.

Além disso, destacou a importância da Lei Complementar nº 182/2021 (conhecida como ‘Marco Legal das Startups’) e de outras normas relacionadas à experimentação regulatória, indicando sua consideração como referência para o aperfeiçoamento do próprio Sandbox regulatório.

A ANTT considerou ainda que a identificação do órgão regulador competente dependerá do local e da abrangência territorial em que a solução da empresa Synapse venha a ser implantada. Foi explicado que os serviços de transporte circunscritos a um município ou a um único estado se submetem, em regra, às respectivas competências municipal ou estadual, enquanto a atuação da ANTT alcança o transporte interestadual e internacional. Nesse sentido, a ANTT recomendou que a análise regulatória considere o ambiente concreto de implantação da solução Trajetto e os órgãos responsáveis pela respectiva infraestrutura ou serviço. A ANTT observou ainda que não é necessária a imediata dispensa ou exceção à norma regulatória específica do setor de transportes. A ANTT ressaltou ainda que os normativos mencionados poderiam servir como referência metodológica, mas que sua aplicabilidade direta dependeria do ambiente concreto de implantação e da eventual vinculação da solução Trajetto a serviços ou a infraestruturas sujeitas à competência da Agência.

A ANTT também compartilhou sua experiência na implementação de ambientes regulatórios experimentais, ressaltando que a construção de um sandbox deve partir da definição clara dos objetivos regulatórios que se pretende alcançar. Os representantes da ANTT distinguiram os indicadores específicos de cada experimento das métricas utilizadas para avaliar o programa de sandbox como um todo. No primeiro caso, os indicadores são definidos conforme os objetivos e as perguntas de cada teste e acompanhados periodicamente junto à empresa participante. No segundo, podem ser consideradas métricas de prazo mais longo, como a efetiva implantação da tecnologia testada, a incorporação das evidências ao processo regulatório e os aprendizados institucionais produzidos. A ANTT ressaltou, ainda, que os resultados de um sandbox regulatório não precisam necessariamente conduzir à edição ou à alteração de normas. Em determinados casos, as evidências produzidas pelo experimento podem indicar que o tema não demanda regulamentação específica, conclusão que também constitui resultado relevante para o processo decisório do regulador. Também foi destacado que resultados diferentes daqueles inicialmente esperados não caracterizam, por si só, o insucesso do experimento.

5.4 Conclusões gerais das reuniões setoriais

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

Assim, é necessário considerar, desde a etapa de planejamento, estratégias específicas para mobilizar os diferentes públicos e comunicar claramente a importância de sua contribuição e a finalidade dos instrumentos de participação. A experiência

evidencia, portanto, a necessidade de investir de forma mais sistemática no engajamento prévio e posterior das partes interessadas. A participação em uma consulta setorial depende não apenas da existência de um espaço formal de diálogo, mas também da criação de condições para que os participantes compreendam a relevância de sua contribuição e se sintam efetivamente envolvidos no processo.

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

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

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

Além desses aspectos, a experiência permite identificar outros aprendizados relevantes para o aprimoramento de futuras consultas.

- A diversidade de participantes precisa ser acompanhada de estratégias de engajamento: É necessário desenvolver estratégias específicas para preparar, mobilizar e manter o engajamento dos diferentes grupos, considerando seus distintos interesses, níveis de disponibilidade e formas de contribuição.
- O tempo de participação deve ser protegido: a decisão de atribuir cinco minutos a cada empresa participante foi orientada pela intenção de ampliar o espaço destinado ao diálogo. A experiência reforça a importância de preservar esse equilíbrio: quanto mais eficiente for a etapa expositiva, maior tende a ser o espaço disponível para a participação efetiva dos diferentes atores.
- Perguntas estruturadas em blocos podem favorecer a participação qualificada: a divisão do debate em quatro blocos temáticos mostrou-se adequada como mecanismo de organização da discussão. A estrutura permitiu abordar diferentes dimensões do Sandbox sem restringir as manifestações a um questionário rígido. A combinação de perguntas norteadoras e espaço para manifestações abertas deve ser preservada em futuras consultas, especialmente em temas tecnicamente complexos.

- As reuniões devem combinar estrutura prévia e flexibilidade para contemplar as demandas que emergirem durante a discussão: a experiência demonstrou que a estrutura previamente definida não deve limitar o desenvolvimento do diálogo. Ao longo das reuniões, podem surgir dúvidas, demandas de esclarecimento ou temas que, embora não previstos nas perguntas iniciais, sejam relevantes para os participantes e para os objetivos da consulta. É importante, portanto, que a metodologia reserve margem para adaptar a ordem, o tempo e o conteúdo das discussões, permitindo aprofundar questões que se mostrem particularmente relevantes durante o encontro. A combinação entre perguntas norteadoras, espaço para manifestações abertas e flexibilidade na condução da reunião tende a favorecer uma participação mais qualificada, especialmente em temas tecnicamente complexos.
- É importante transformar a participação em uma relação continuada: a consulta setorial pode ser mais efetiva quando compreendida como parte de um processo continuado de relacionamento com as partes interessadas, e não como um evento isolado. Nesse sentido, futuras iniciativas podem prever, além da consulta, mecanismos de acompanhamento, como comunicação posterior sobre os resultados, retorno sobre a incorporação das contribuições e eventual convite para novas etapas de discussão. Essa estratégia pode contribuir para aumentar a percepção dos participantes de que suas contribuições produzem efeitos concretos no processo regulatório, fortalecendo o interesse em participar de futuras iniciativas.
- A aproximação com reguladores deve ocorrer, sempre que possível, antes da conclusão da experimentação, permitindo que suas contribuições possam efetivamente influenciar os testes, e não apenas ser registradas após o encerramento do processo. Quanto mais cedo e mais diretamente os potenciais reguladores forem envolvidos, maior tende a ser a capacidade de transformar a consulta em instrumento efetivo de aprendizagem e de aprimoramento da testagem.

6. Conclusões gerais do Ciclo 6

O Ciclo 6 concluiu o percurso experimental das soluções Trajetto, da Synapse; STAiDOC, da Prevvine; e GuardionAI, da Metatext, consolidando as avaliações técnicas, os instrumentos de governança, as evidências de funcionamento e os aprendizados registrados pela ANPD e pelas participantes. O encerramento permitiu identificar os avanços acumulados, a situação das condicionantes e as providências necessárias à descontinuidade dos ambientes experimentais ou à eventual utilização posterior das soluções.

A apreciação conjunta evidencia resultados distintos, relacionados à natureza das aplicações, à maturidade dos componentes e à suficiência das demonstrações apresentadas. As pontuações éticas, os níveis de maturidade tecnológica, os testes funcionais e os pareceres de segurança examinam dimensões complementares, com critérios e bases de evidências próprios. Consequentemente, seus resultados devem permanecer associados aos respectivos objetos, evitando que uma classificação agregada substitua a análise dos riscos individuais ou que um avanço documental seja apresentado como comprovação de eficácia operacional.

Quanto à Synapse, o ciclo final consolidou avanços na arquitetura de proteção de dados e na governança do Trajetto. Foram reconhecidos a segregação entre identidade e telemetria, os mecanismos de pseudonimização e criptografia, o processamento de imagens na infraestrutura dos parceiros e a formalização das responsabilidades institucionais. A documentação também passou a distinguir a disponibilização universal de rotas acessíveis da coleta facultativa de informações sobre deficiência, contribuindo para reduzir a vinculação desnecessária entre a finalidade funcional da navegação e o tratamento de dados sensíveis.

A pontuação no *framework* geral passou de 78 para 86 pontos, alcançando a classificação de alta maturidade ética e de governança, enquanto os componentes tecnológicos permaneceram em TRL 4 para os algoritmos centrais e TRL 6 para a pseudonimização e a proteção das bases apartadas. Os quatro ensaios técnicos anteriormente realizados permaneceram como referência, sem registro de nova execução integral no Ciclo 6. A evolução ética e documental deve, portanto, ser interpretada juntamente com os limites dos testes e com as verificações ainda necessárias em condições operacionais.

O parecer manteve a classificação “suficiente com condicionantes”, com risco inerente agregado alto e residual agregado médio, preservando riscos individuais elevados. Permanecem relevantes a demonstração das telas e explicações efetivamente apresentadas ao usuário, a resistência à manipulação de coordenadas, a proteção contra correlação de trajetos, a delimitação das integrações e responsabilidades contratuais e o planejamento de inclusão de pessoas com deficiência visual. A Matriz de Transparência foi preenchida e avaliada, mas seu conteúdo não comprova, isoladamente, a implementação dos avisos e controles na interface.

Nesse contexto, o encaminhamento relativo à Synapse reconhece a conclusão da etapa experimental e mantém requisitos para eventual operação regular. A ficha da ANPD classificou o êxito como parcial, enquanto a participante informou não ter registrado novas evoluções no período, concentrando esforços nas entregas documentais. O Plano de Descontinuidade foi apresentado e validado como instrumento de encerramento, permanecendo necessária a comprovação de sua operacionalização. A redução dos riscos associada à revogação de acessos, à desativação dos recursos e ao expurgo dos dados depende da execução verificável dessas medidas.

Em relação à Prevvine, o Ciclo 6 registrou revisão das premissas técnicas e jurídicas da STAiDOC, aperfeiçoamentos de interface e supervisão humana e ampliação da identificação dos riscos. Destacam-se a correção de dezessete afirmações anteriores, a atualização da Avaliação de Impacto em Direitos Humanos, a identificação nominal do Encarregado e o reconhecimento de que o fluxo encaminhado ao provedor envolve dados pessoais sensíveis submetidos à supressão de identificadores. Essa requalificação representa aprendizado central do projeto, pois explicita que a aplicação de expressões regulares e a omissão de dados na resposta não demonstram anonimização integral.

A pontuação no *framework* geral aumentou de 43 para 55 pontos, passando à faixa de risco ético moderado a alto, com necessidade de mitigação antes da ampliação de uso. A melhora decorreu, principalmente, do reconhecimento dos mecanismos de confirmação ou rejeição das sugestões pelo médico, da capacidade de intervenção humana e do conhecimento das limitações demográficas. Os componentes tecnológicos permaneceram nos níveis declarados de TRL 7 para o motor de IA e a infraestrutura de conformidade e TRL 6 para a supressão de identificadores e a verificação bibliográfica, observada a necessidade de harmonização dos registros da matriz.

A consolidação da HarpIA reuniu evidências clínicas e adversariais produzidas durante a experimentação, sem documentar uma nova rodada clínica específica do Ciclo 6. Foram reconhecidos aspectos favoráveis de compreensibilidade e naturalidade, acompanhados de limitações relativas à fundamentação bibliográfica, ao preenchimento de informações ausentes, às inferências e à proteção de dados. Os estudos de privacidade identificaram exposições e utilização de identificadores em estratégias conversacionais de correção e esclarecimento. Seus resultados devem permanecer vinculados aos conjuntos e critérios examinados, sem conversão das frequências observadas em taxas gerais de vulnerabilidade.

O parecer técnico apresentou conclusão favorável com condicionantes para a continuidade posterior ao sandbox, mantendo risco inerente agregado crítico e residual agregado médio. Entretanto, as correções destinadas à próxima versão ainda dependem de implantação identificada e reteste. Subsistem providências relativas à ampliação dos detectores, aos controles anteriores e posteriores à inferência, à proteção dos resumos criptográficos, ao firewall de aplicação, à contingência bibliográfica, aos testes de vieses e à formalização das garantias dos fluxos internacionais. Também permanecem necessários a atualização dos instrumentos de tratamento, o preenchimento material da coluna final da Matriz de Transparência e a execução documentada dos exercícios de incidente e descontinuidade.

A interlocução entre ANPD, ANS e ANVISA constituiu resultado institucional relevante, permitindo discutir a proteção de dados em conjunto com os requisitos do setor de saúde. As manifestações sanitárias sobre o possível enquadramento da STAiDOC foram preliminares e não substituem análise formal ou regularização. Assim, o encerramento da participação da Prevvine deve registrar os avanços e as pendências, preservando a distinção entre a identificação dos riscos, a preparação das correções e a comprovação de proteção efetiva dos usuários. As fichas C1 não formalizaram uma classificação conjunta do êxito global.

Por sua vez, a interlocução com a Agência Nacional de Transportes Terrestres (ANTT), relacionada ao Trajetto, contribuiu para delimitar a análise das competências regulatórias conforme o ambiente concreto de implantação da solução e a abrangência territorial do serviço envolvido. A Agência também compartilhou sua experiência com ambientes experimentais, destacando a necessidade de objetivos regulatórios claros e a distinção entre indicadores de cada teste e métricas de avaliação do programa de sandbox. Ademais, ressaltou que a experimentação pode produzir resultados relevantes sem conduzir à edição ou à alteração de normas, inclusive ao demonstrar que determinada questão não demanda regulamentação específica. Esses aprendizados

ampliam a contribuição institucional do piloto para o desenho de futuras iniciativas e para a utilização das evidências no processo decisório regulatório.

Quanto à Metatext, o Ciclo 6 consolidou o atendimento das exigências técnicas e a demonstração das salvaguardas da GuardionAI. O parecer final foi favorável e registrou como cumpridas as cinco condicionantes: sanitização dos registros e avaliação especializada de segurança; comprovação das salvaguardas da infraestrutura; execução dos exercícios de incidente, eliminação, explicação e contestação e descontinuidade; realização do protocolo de vieses e dos retestes; e verificação do cofre de uso único. O resultado sustenta a recomendação de encerramento satisfatório da participação.

A avaliação ética alcançou 86 pontos, correspondentes à alta maturidade ética e de governança. No parecer técnico, foram reconhecidos TRL 7 para a infraestrutura e o conjunto de pseudonimização e cofre e TRL 8 para o detector, conforme os critérios daquela avaliação. A organização dos contratos, das políticas em português, dos responsáveis e dos canais de atendimento também permitiu relacionar a configuração técnica às obrigações dos agentes. O acompanhamento individualizado registrou vinte e três solicitações atendidas e uma atendida com ressalva, referente aos fluxos internacionais, sem condicionante técnica remanescente no parecer final.

Os resultados da HarpIA contribuíram para delimitar a capacidade e os pontos de aperfeiçoamento da solução. A avaliação humana registrou F1 micro de 0,74, enquanto o procedimento híbrido, com 1.225 execuções, alcançou 0,774. Esses valores devem permanecer distintos do patamar de 0,90 reportado nos retestes direcionados dos modos de falha conhecidos. A avaliação dos trinta alarmes identificou elementos úteis à apuração inicial dos incidentes, acompanhados de recomendações sobre coerência das severidades, fundamentação das justificativas, linguagem e adaptação ao destinatário. O atendimento das condicionantes convive, portanto, com uma agenda de evolução do desempenho e da comunicação.

A experiência da GuardionAI demonstrou a possibilidade de uma startup estruturar uma camada intermediária de proteção com integração entre detecção, tokenização, reversão controlada, registros e supervisão. A experimentação com dados sintéticos e a disponibilização pública do conjunto de avaliação acrescentaram contribuições metodológicas ao ecossistema. Permanecem como compromissos posteriores a manutenção das salvaguardas, a evolução das certificações independentes da participante, a harmonização documental e o aperfeiçoamento das explicações e da triagem de falsos positivos.

De modo transversal, o projeto demonstrou a importância de avaliar a proteção de dados em toda a cadeia de processamento. A retirada de identificadores diretos, a separação de bases, a criptografia e o descarte de vínculos constituem medidas relevantes, cuja suficiência depende dos dados remanescentes, do contexto, das integrações e das possibilidades de associação ao titular. Esse aprendizado foi particularmente significativo na revisão do fluxo da Prevvine, na análise da combinação entre mobilidade e acessibilidade da Synapse e na avaliação dos limites da detecção e tokenização da Metatext.

No campo da transparência, a experiência reforçou a utilidade de organizar as informações conforme o destinatário. Titulares, profissionais, instituições contratantes, equipes de segurança e autoridades necessitam de conteúdos e níveis de detalhamento diferentes. A existência de políticas e relatórios deve ser acompanhada da demonstração dos avisos, dos canais e dos procedimentos utilizados. Da mesma forma, a supervisão humana exige condições para examinar o caso, identificar os elementos relevantes, corrigir intervenções e fundamentar a resposta, mantendo-se a distinção entre justificativas geradas pelo modelo e evidências do processamento.

Sob a perspectiva institucional, a articulação entre fichas, matrizes, testes, pareceres e avaliações de êxitos produziu uma base de conhecimento sobre os requisitos

de desenvolvimento e acompanhamento de soluções de IA. O piloto também evidenciou a importância da coordenação entre autoridades quando o tratamento integra serviços regulados, preservando as competências próprias de cada instituição. Seus resultados oferecem subsídios para orientações, práticas de supervisão e futuras edições do sandbox, sem pressupor que todas as questões examinadas demandem nova norma.

Nesse âmbito, a consulta ao Banco Central do Brasil (Bacen), realizada em 23 de setembro de 2026, permitiu apresentar a GuardionAI e os resultados preliminares da experimentação sob a perspectiva do regulador financeiro, abordando transparência e supervisão, anonimização e governança de dados, explicabilidade e segurança e questões regulatórias. Além da aproximação institucional, o encontro produziu aprendizados sobre a própria organização das consultas: a complexidade das soluções exige contextualização antecipada dos interlocutores externos e flexibilidade na condução das perguntas, criando condições para contribuições substantivas. A experiência reforça, portanto, a importância de envolver as autoridades setoriais desde o planejamento e de registrar os encaminhamentos e sua eventual incorporação às avaliações.

Para a consolidação do encerramento, recomendam-se as seguintes providências:

- Rastreabilidade das conclusões: organizar o acervo final por participante, solicitação, versão, teste e evidência, distinguindo os resultados acumulados das verificações realizadas no ciclo final e das funcionalidades previstas para versões posteriores.
- Situação das condicionantes: preservar o registro individualizado das exigências cumpridas, das ressalvas e dos requisitos remanescentes, diferenciando condições de encerramento, condições para eventual retomada e compromissos de aperfeiçoamento continuado.
- Coerência documental: harmonizar classificações de risco, níveis de maturidade, prazos de retenção, papéis dos agentes, descrições de funcionalidades e referências aos fluxos internacionais, mantendo identificadas as divergências que não puderam ser resolvidas.
- Descontinuidade e destino dos dados: reunir os comprovantes disponíveis de revogação de acessos, desativação de integrações, eliminação ou conservação justificada dos dados e comunicação às partes, registrando separadamente os planos aprovados, os exercícios realizados e a execução do encerramento.
- Continuidade posterior ao piloto: vincular eventuais alterações ou ampliações às avaliações de risco, aos testes e às exigências setoriais pertinentes, sem pressupor autorização de nova rodada experimental ou regularização decorrente da participação no programa.
- Aprendizagem institucional: incorporar os resultados aos instrumentos de orientação e à metodologia de futuras iniciativas, especialmente quanto à avaliação de privacidade, à explicabilidade por destinatário, à supervisão humana e à coordenação regulatória.

O Ciclo 6 encerra, assim, uma experiência capaz de produzir avanços técnicos e institucionais, identificar limitações e fundamentar encaminhamentos diferenciados. Para a Synapse, permanecem condicionantes de comprovação e requisitos para eventual operação regular; para a Prevvine, subsistem correções, retestes e adequações regulatórias; para a Metatext, o cumprimento das condicionantes sustenta o encerramento satisfatório, acompanhado de compromissos de manutenção e aperfeiçoamento. Em suma, a contribuição do sandbox reside nessa capacidade de

relacionar inovação, evidências e responsabilidades, preservando os direitos dos titulares e delimitando o alcance de cada conclusão.


sandbox regulatório


ANPD