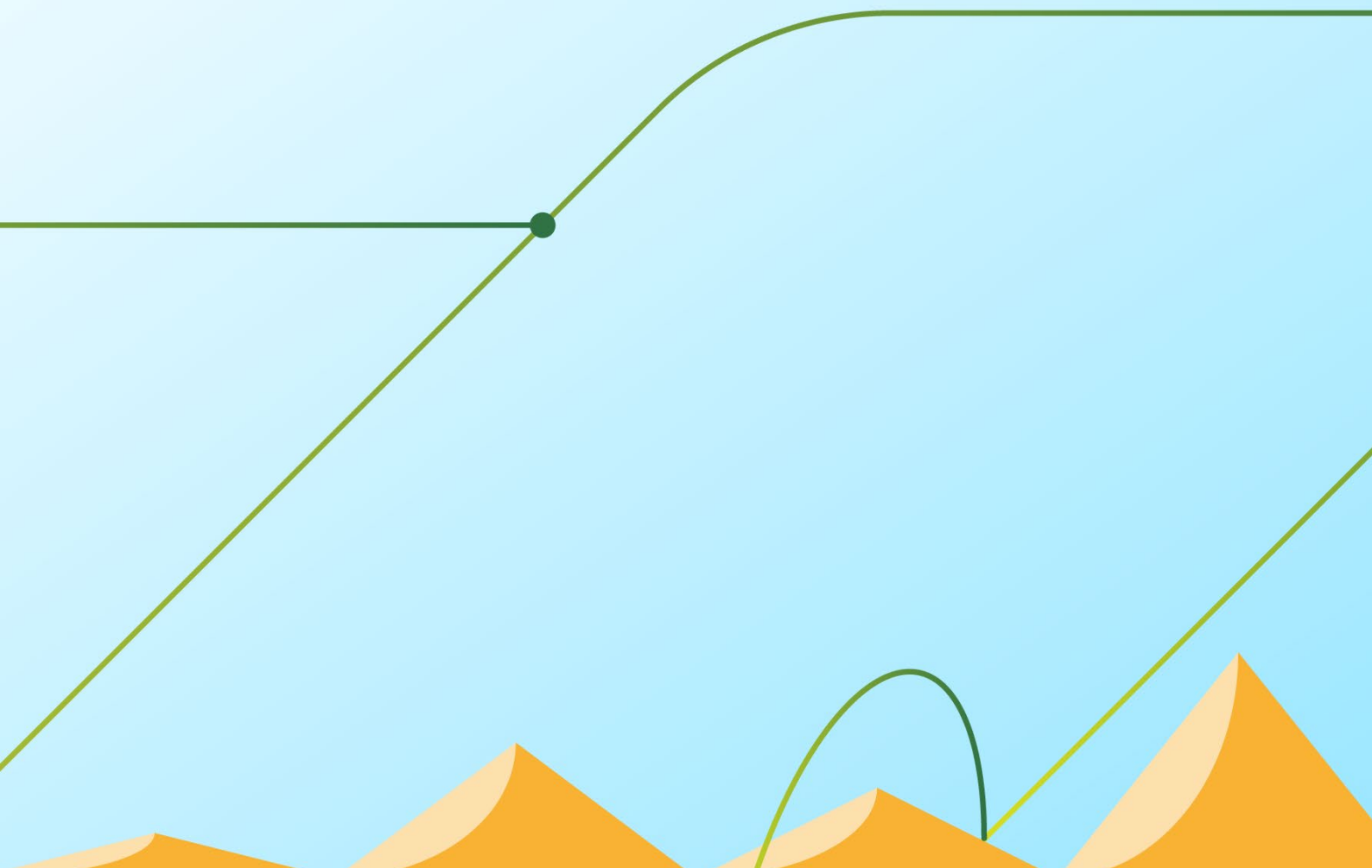




3º Relatório Parcial de Monitoramento

Fase de Teste – Ciclo 3



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

Angela Halen Claro Franco
Diego André Sant'Ana
Ewerton Luiz Costadalle
Fábio Silveira Vidal
Lucas Costa dos Anjos
Maria Carolina Ferreira da Silva
Marcus Vinícius Rossi da Rocha
Jeferson Dias Barbosa

Equipe de Supervisão

Albert França Josué Costa
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	10
4.1 Synapse	10
4.1.1 Documentação principal	11
4.1.2 Fase A	11
4.1.3 Fase B	14
4.1.4 Fase C	18
4.1.5 Resultados do Ciclo 3 sob a perspectiva da ANPD	21
4.1.6 Resultados do Ciclo 3 sob a perspectiva da Synapse	24
4.2 Prevvine	24
4.2.1 Documentação principal	25
4.2.2 Fase A	26
4.2.3 Fase B	29
4.2.4 Fase C	32
4.2.5 Resultados do Ciclo 3 sob a perspectiva da ANPD	32
4.2.6 Resultados do Ciclo 3 sob a perspectiva da Prevvine	33
4.2.7 Coleta e análise das evidências	34
4.3 Metatext	34
4.3.1 Documentação principal	35
4.3.2 Registros do progresso técnico e regulatório	35
5. Conclusões gerais do Ciclo 3	50

1. Introdução

Este relatório apresenta o planejamento, a orientação e a supervisão das atividades do Ciclo 3 do Projeto Piloto de Sandbox Regulatório de Inteligência Artificial e Proteção de Dados, consolidando os resultados técnicos e regulatórios das participantes Synapse¹, com o projeto Trajetto; Previne², com a plataforma STAiDOC; e Metatext³, com a solução GuardionAI. Seu objetivo é registrar os avanços, os riscos, as limitações e os aprendizados identificados, bem como fundamentar os encaminhamentos para o Ciclo 04.

O acompanhamento manteve a metodologia estruturada⁴ nas Fases A — Pré-teste, B — Teste e C — Pós-teste, com atividades adaptadas às características e ao estágio de desenvolvimento de cada solução. Foram realizadas reuniões remotas, orientações individualizadas, análises documentais, execução ou preparação de testes, avaliações de riscos e medidas de mitigação e consolidação de aprendizados. As evidências consideradas incluíram fichas A1 a A4, pareceres técnicos e regulatórios, documentos de arquitetura e de fluxos de dados, instrumentos de avaliação ética, matrizes de transparência, registros de testagem, informações da plataforma HarPIA, fichas de êxitos⁵ e documentos de descontinuidade, conforme disponíveis para cada participante.

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

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

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

² A Previne 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-aco-es-iniciativas/sandbox/metodologia-sandbox.pdf/@@display-file/file>, acessado em: 14 de outubro de 2026

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

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

Na Synapse, os testes verificaram a utilização de protocolo de segurança na comunicação auditada com o banco de dados e não identificaram indivíduos no processamento de imagens nas condições examinadas. A pontuação no *framework* geral passou de 45 para 68 pontos. O parecer classificou a solução como “suficiente com condicionantes”, com risco inerente alto e residual médio. A entrega de instrumentos de governança, como RIPD, AIDH e plano de resposta a incidentes, ampliou a base documental, mas não demonstrou integralmente sua implementação prática. Permaneceram pendências relativas às *Application Programming Interface* (API) e aos fluxos externos, à identificação indireta, à proteção da aplicação, à acessibilidade e às evidências de transparência. A classificação de êxito integral registrada na Fase C refere-se ao escopo de testagem realizado, preservadas essas condicionantes.

Na Previne, o relatório registra o restabelecimento da integração STAiDOC–HarplA, a aplicação das 30 vinhetas clínicas e ajustes na resposta da API e na apresentação de referências. Os testes também revelaram alucinações, referências inexistentes e links quebrados, motivando a redução do nível de maturidade tecnológica do sistema de verificação bibliográfica NTAV⁶ de TRL 7 a TRL 6. A pontuação no *framework* geral passou de 60 para 63 pontos; o parecer manteve a avaliação favorável com condicionantes, risco inerente crítico e residual médio, enquanto a Fase C classificou o êxito como parcial. Os avanços operacionais não encerram as limitações clínicas nem comprovam o cumprimento das medidas pendentes de explicabilidade, proteção de dados, segurança, continuidade e apresentação de políticas.

Na GuardionAI, o Ciclo 3 concentrou-se na compreensão da arquitetura, dos fluxos de dados e da documentação, no aprofundamento do mapeamento regulatório e na preparação de dois testes: proteção de dados pessoais e envio de alertas à equipe de segurança. O levantamento normativo foi ampliado de cinco para 58 diplomas. As declarações e os documentos examinados permitiram delimitar questões de supervisão humana, pseudonimização, transferências internacionais, governança e explicabilidade, cuja confirmação prática permanece prevista para o ciclo seguinte. O relatório registra como pendentes a disponibilização do RIPD para avaliação e os demais artefatos de conformidade indicados.

Para o Ciclo 04, os encaminhamentos prioritários são:

- Synapse: aprofundar a análise de integrações externas e riscos de identificação indireta e comprovar controles de segurança, acessibilidade, transparência e governança.
- Previne: examinar os resultados das vinhetas e cumprir as condicionantes COND2 a COND7, incluindo políticas, explicabilidade, proteção criptográfica, supressão de dados pessoais, segurança da aplicação e contingência para indisponibilidade do PubMed.
- Metatext: executar os testes previstos, receber e avaliar a documentação pendente e esclarecer responsabilidades, fluxos internacionais e mecanismos de supervisão e explicabilidade.

⁶ NTAV: Mecanismo de explicabilidade e verificação de referências bibliográficas externas.

- Todas as participantes: acompanhar recomendações com responsáveis, prazos e evidências, atualizar riscos residuais e maturidade tecnológica e preservar o registro das alterações nas soluções.

Dessa forma, a leitura dos resultados alcançados no sandbox regulatório deve distinguir controles efetivamente verificados, medidas documentadas ou declaradas e providências ainda sujeitas à validação. Essa distinção orienta a continuidade da experimentação e a avaliação da efetividade das salvaguardas nos ciclos subsequentes.

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

Preliminarmente, é importante registrar que, em razão dos resultados alcançados e da experiência prática adquirida no Ciclo 2, a estrutura das sessões e das interações foi mantida para o Ciclo 3. Nesse sentido, observa-se que, além das reuniões previamente definidas, para fins de aprofundamento do diagnóstico de cada participante, foram organizados: (i) encontros concomitantes, bem como (ii) diálogos internos entre os membros da consultoria sobre o melhor aproveitamento da metodologia proposta para este sandbox e, por fim, (iii) intercâmbio de informações por meio da plataforma Teams em razão da agilidade que proporciona à comunicação entre as equipes regulatórias e técnicas do CIAAM/USP e os representantes dos projetos selecionados.

Dessa forma, as bases estabelecidas nos ciclos 1 e 2 foram fundamentais para possibilitar o início do Ciclo 3, especialmente no que se refere ao alinhamento técnico e regulatório entre as equipes das participantes e da consultoria do CIAAM/USP, bem como para a disponibilização das informações mais relevantes para a ANPD no que tange à conclusão da preparação dos ambientes tecnológicos para o início dos testes.

No Ciclo 3, foram realizadas reuniões remotas individualizadas com as empresas participantes e discussões internas entre as equipes técnicas e regulatórias, com registro de orientações e encaminhamentos. As entregas compreenderam o calendário de testes previstos e executados, os pareceres técnicos e regulatórios, bem como a consolidação dos resultados neste relatório. Foram fichas de caracterização e maturidade tecnológica, documentos de arquitetura e fluxos de dados, instrumentos de governança, avaliações éticas e de transparência e registros de testagem. Como principais resultados, destacam-se a verificação de controles criptográficos e do processamento de imagens da Synapse; a retomada da integração com a plataforma HarpIA e a aplicação das vinhetas clínicas da Prevvine, com identificação de limitações e revisão da maturidade do sistema NTAV; e o aprofundamento da análise documental e normativa da Metatext, com definição dos testes previstos para o Ciclo 04. Essas atividades fundamentaram recomendações e condicionantes específicos para a continuidade da avaliação de cada participante.

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

No âmbito do Ciclo 3, se manteve a metodologia de supervisão e acompanhamento validada conjuntamente pela equipe do CIAAM/USP e pela Comissão de Sandbox da ANPD. Assim, os trabalhos foram estruturados em três fases: (i) Fase A – Pré-teste; (ii) Fase B – Teste; e (iii) Fase C – Pós-teste

Dessa forma, permaneceram aplicáveis tanto a estrutura metodológica comum quanto a individualização dos conteúdos e das atividades de acordo com o Plano de Sandbox de cada participante, as características técnicas e regulatórias da solução, o setor de aplicação, os dados pessoais tratados, o nível de maturidade tecnológica, os riscos identificados e as evidências produzidas durante os ciclos de testagem.

Nesse passo, deve-se registrar que as sessões continuaram a combinar exposição orientativa, análise documental, discussão técnica e regulatória e registro de encaminhamentos. Com isso, o acompanhamento foi realizado prioritariamente por meio de reuniões remotas e individualizadas, complementadas pelos canais específicos mantidos no Microsoft Teams entre as empresas, o CIAAM/USP e a ANPD.

Ademais, a estruturação do calendário conferiu a dinamicidade para acomodar os ajustes justificados pelo desenvolvimento dos testes e pelas demandas verificadas em cada caso. Além disso, essa continuidade metodológica assegurou, na medida do possível, a coerência entre os Planos de Sandbox aprovados, as atividades executadas e os resultados produzidos, bem como preservou a comparabilidade entre os ciclos, a rastreabilidade das decisões e a adequada fundamentação dos pareceres e relatórios do projeto piloto.

Finalmente, embora não tenha havido necessidade de se realizar reuniões presenciais nesta oportunidade, a organização dos encontros no formato remoto e a comunicação assíncrona possibilitada pela plataforma Microsoft Teams viabilizaram a realização, com ajustes, do calendário previsto bem como o início dos testes das inovações apresentadas pelas participantes. Também, verificou-se que a metodologia foi observada e adaptada conforme necessário para o pleno atendimento dos tópicos propostos para o Ciclo 3 do Sandbox Regulatório de IA e Proteção de Dados.

4. Apoio, orientação e supervisão das atividades de testagem

Neste item do relatório, tal como no Ciclo 2, serão apresentadas as atividades destinadas à realização das testagens, ao apoio às empresas participantes, às orientações da equipe do CIAAM/USP e da ANPD e à supervisão contínua e direta dos testes conduzidos no Ciclo 3.

É importante recordar que o objeto dos testes depende da inovação apresentada pela participante bem como do nível de maturidade tecnológica alcançado. Dessa forma, espera-se que cada solução apresente desenvolvimento e resultado compatíveis com o seu progresso até o início de cada teste.

4.1 Synapse

O relatório do CIAAM USP referente ao Ciclo 3 submetido à ANPD apresenta os principais dados, informações e conhecimento extraídos deste ciclo considerando os seguintes pilares da metodologia de avaliação das soluções de IA das empresas:

- Dimensão Técnico-funcional: fichas técnicas referentes à solução e à arquitetura (A1 e A2);
- Dados, privacidade e segurança: fichas técnicas desenvolvidas pela equipe do CIAAM/USP nas subfases A3, A4, B1 e B2;
- Maturidade Tecnológica (TRL⁷ e ETC⁸ identificados): nas fichas A1 e A2, as empresas apresentam a sua autoavaliação. Nas fichas B1 e B2, há a análise da equipe do CIAAM/USP.
- Realismo experimental e benefícios esperados: testes e evidências coletadas na Fase B;
- Riscos, mitigação e impacto residual: foram desenvolvidas fichas de avaliação de riscos na Subfase B2;
- Governança, transparência e aprendizagem regulatória: trata-se da análise da documentação, do organograma, dos papéis e responsáveis, da explicabilidade da solução e do uso da matriz de transparência. Todos esses elementos são analisados na Subfase B1 de testagem.

Como o Ciclo 3 foi completado em suas três fases (A, B e C), cumpre-se relatar a documentação principal e os principais aspectos identificados em cada fase.

⁷ TRL - *Technology Readiness Level* ou Nível de Maturidade Tecnológica

⁸ ETC – Elementos Tecnológicos Críticos

4.1.1 Documentação principal

Preliminarmente, observa-se que os principais documentos referidos neste item do relatório encontram-se organizados conforme a descrição a seguir, por fases, para atendimento dos requisitos pertinentes ao Ciclo 3 deste Sandbox Regulatório:

- Fase A:
 - Fichas A1 e A2 - Aplicação e Matriz TRL;
 - Fichas A3 e A4;
 - Fichas A3 e A4 - Tabela resumo da evolução dos ciclos;
 - Parecer regulatório dos documentos enviados pela Synapse;
 - Parecer técnico dos documentos enviados pela Synapse.
- Fase B:
 - Ficha B1 - *Framework* geral;
 - Parecer Técnico Consolidado;
 - Matriz de Transparência – Synapse.
- Fase C:
 - Ficha C1 - Êxitos para a ANPD e para a empresa participante;
 - Plano de Descontinuidade.

4.1.2 Fase A

Inicia-se a análise da Ficha Técnica de Aplicação (A1 e A2) sob responsabilidade das empresas. Note-se que, quanto à arquitetura do modelo e ao fluxo de dados, não houve mudança do Ciclo 2 para o Ciclo 3.

Da mesma maneira, os protocolos de segurança (art. 46 da LGPD) permaneceram semelhantes entre os Ciclos 02 e 03, e, até o momento, não há autenticação multifator (*Multi-Factor Authentication* - MFA). Em relação à Ficha A2, a qual trata da temática concernente à avaliação de maturidade tecnológica e ao mapeamento dos elementos tecnológicos críticos (ETC), é possível verificar que a validação técnica dos algoritmos centrais permaneceu no mesmo nível do Ciclo 2, ou seja, TRL 4. Já no que concerne à pseudonimização e à segurança dos bancos de dados apartados, é possível notar que a empresa manteve o TRL 6 do Ciclo 2.

No que tange às Fichas A3 e A4, em termos de análise regulatória e documental, houve mudança relevante entre os Ciclos 02 e 03, pois foram realizadas análises pela equipe regulatória destacando-se os seguintes aspectos elencados a seguir e aprofundados nos respectivos documentos:

- A análise regulatória da equipe do CIAAM/USP examinou os principais instrumentos de governança, proteção de dados, direitos humanos, segurança da informação e inteligência artificial adotados no projeto.
- Foram analisados o Relatório de Impacto à Proteção de Dados Pessoais (RIPD), a Avaliação de Impacto de Direitos Humanos (AIDH), a nomeação do Encarregado pelo Tratamento de Dados Pessoais, as políticas de governança de IA e de dados, o plano de resposta a incidentes e os registros das operações de tratamento apresentados à ANPD.

- Em razão do seu caráter predominantemente técnico, os documentos relativos à arquitetura de privacidade e criptografia, inventário de API e terceiros, fluxo de dados ponta a ponta, explicabilidade em camadas e testes de robustez e estresse foram analisados pela equipe técnica do CIAAM/USP.
- Em relação aos Acordos de Cooperação Técnica (ACT) com entidades privadas, estes devem delimitar melhor o ambiente de operação, as entregas, os usos proibidos e a finalidade exclusivamente técnica, científica e regulatória do projeto experimental.
- Os ACT também devem definir os dados tratados, as bases legais, as finalidades, os papéis dos agentes de tratamento, as responsabilidades das partes e as medidas técnicas e organizacionais de segurança.
- Recomenda-se incluir nos ACT regras específicas de governança de IA, supervisão humana, avaliação de impacto, revisão periódica de riscos, versionamento do modelo, rastreabilidade e validação técnica.
- Nos ACT com entidades públicas, devem ser explicitados a finalidade pública da cooperação, sua relação com as competências do órgão, os deveres relativos à LGPD, a incidência da Lei de Acesso à Informação e o regime de recursos e custos.
- A experiência do usuário apresenta fluxo funcionalmente organizado, mas possui fragilidades de acessibilidade, transparência e atendimento a pessoas com deficiência, idosos, usuários com baixa literacia digital ou mobilidade reduzida.
- O aplicativo depende excessivamente de elementos visuais, mapas, textos pequenos e digitação, sem demonstrar suficientemente a existência de alternativas como descrição textual, orientação por áudio, alto contraste e compatibilidade com tecnologias assistivas.
- Recomenda-se disponibilizar rotas acessíveis como opção para todos os usuários, com filtros funcionais para evitar escadas, trechos movimentados ou mal iluminados, preferir elevadores e indicar obstáculos físicos, sem exigir autodeclaração desnecessária de deficiência.
- O tratamento de dados de localização e deslocamento apresenta riscos de vigilância, exposição de rotinas e inferência de hábitos, exigindo informações claras sobre finalidade, duração, destinatários e formas de interrupção do compartilhamento.
- A interface deve separar o aceite dos termos, a autorização de localização e o recebimento de comunicações, além de permitir compartilhamento temporário, encerramento imediato e visualização de quem possui acesso à localização.
- A AIDH classificou como elevados os riscos iniciais relacionados à opacidade algorítmica e aos impactos sobre grupos vulneráveis. Esses riscos podem ser reduzidos mediante transparência ativa, explicações

acessíveis, indicação de obstáculos e monitoramento contínuo da acurácia.

- O RIPD foi considerado, em linhas gerais, compatível com o modelo adotado pela ANPD, mas deve incorporar expressamente os riscos de recomendações inadequadas de acessibilidade e esclarecer a hipótese legal aplicável ao eventual tratamento de informações que revelem deficiência de pessoa identificada ou identificável. Essas informações constituem dados pessoais sensíveis referentes à saúde, nos termos do art. 5º, II, da LGPD, e seu tratamento deve observar as hipóteses do art. 11, não sendo admissível fundamentá-lo no legítimo interesse. Recomenda-se, portanto, disponibilizar rotas acessíveis e filtros funcionais a todos os usuários, sem exigir autodeclaração de deficiência, evitando a coleta de dados sensíveis desnecessários à prestação da funcionalidade.
- O Encarregado pelo Tratamento de Dados Pessoais deve possuir autonomia técnica, independência e ausência de conflito de interesses conforme Resolução CD/ANPD nº 18/2024. A indicação de pesquisador pode ser admitida transitoriamente no Sandbox, mas a operação regular requer estrutura institucional independente.
- A política de governança de IA permanece excessivamente abstrata e deve detalhar os destinatários das informações, os momentos e meios de comunicação, a supervisão humana, os mecanismos de intervenção e os procedimentos de monitoramento e responsabilização.
- Os *model cards* devem apresentar uso pretendido, métricas de desempenho, limitações, cenários de falha e riscos éticos. Também devem ser fornecidos exemplos das notificações e explicações efetivamente apresentadas aos usuários.
- A política de governança de dados contém declarações genéricas e precisa demonstrar como as salvaguardas são implementadas, especialmente quanto a dados de deficiência, minimização, consentimento, transparência e alternativas em caso de recusa.
- O plano de resposta a incidentes foi considerado suficiente, mas deve definir com maior precisão o conceito de “rota crítica” e os critérios utilizados para sua identificação e priorização.
- O relatório de viés e *fairness* representa avanço técnico relevante em *fairness* espacial, monitoramento de *drift* e supervisão humana, mas não resolve as pendências referentes à acessibilidade, bases legais, consentimento granular, transparência na interface e operacionalização da governança. As métricas e evidências de intervenção humana ainda necessitam de validação em ambiente produtivo.

É importante observar que a equipe regulatória preparou o documento denominado “Análise dos documentos enviados pela Synapse” em que os pontos acima

descritos são aprofundados, bem como são listados os documentos técnicos não abrangidos pela análise regulatória nesta oportunidade:

- Arquitetura de privacidade e criptografia;
- Inventário de API e terceiros;
- Mapa de fluxo ponta a ponta;
- Protocolo de explicabilidade em camadas;
- Relatório de testes de robustez e estresse.

4.1.3 Fase B

A) Framework geral baseado em recomendações da UNESCO:

Da mesma forma que no Ciclo 2, empregou-se a seguinte metodologia com a aplicação do *framework* geral:

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.

Dessa maneira, buscou-se transformar as Recomendações da UNESCO sobre Ética da Inteligência Artificial⁹ em um *framework* mensurável, estabelecendo-se a supracitada escala de pontuação. No Ciclo 3, a pontuação da Synapse no *framework* geral passou de 45 para 68 pontos, constituindo indicativo de amadurecimento nos critérios examinados, sem comprovar, por si só, conformidade regulatória ou efetividade integral das salvaguardas. A ausência de pontuação no Eixo 4 (Justiça, não discriminação e vieses algorítmicos) permanece como ponto sensível, especialmente diante dos riscos relacionados à acessibilidade, à mobilidade reduzida e à reidentificação indireta. Esses aspectos deverão ser aprofundados mediante evidências específicas nos ciclos subsequentes.

B) Testes da Equipe Técnica do CIAAM USP

Foram realizados 4 testes pela equipe técnica do CIAAM/USP, sendo que foi possível avançar de maneira significativa em relação ao Ciclo 2 como resultado do diálogo

⁹ UNESCO. Recomendação sobre a Ética da Inteligência Artificial. Paris: UNESCO, 2022. Disponível em: <https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence>. Acessado em: 14 de outubro de 2026.

e da constante interação com a equipe da participante no que se refere à adoção das melhores práticas para a garantia da privacidade e segurança da informação. Assim, são apresentadas a seguir as conclusões sobre cada um dos testes executados no Ciclo 3:

1) Teste 1: ajuste do conjunto de dados

O conjunto de dados que acompanha o algoritmo fornecido foi gerado pela empresa. Apesar de os testes terem sido bem-sucedidos nos ciclos anteriores, os pesquisadores do CIAAM/USP propuseram que os dados utilizados no teste fossem gerados de forma independente. Neste sentido, foi montado neste ciclo um conjunto de dados de autoria da equipe do CIAAM/USP. Como o conjunto de dados contém rostos de pessoas reais em espaço público obtidos exclusivamente para a pesquisa realizada pela equipe técnica, o conjunto não será divulgado e será descartado após o encerramento dos testes técnicos do Sandbox.

2) Teste 2: geração de dados passando pelo túnel:

O teste do túnel para o Ciclo 3 foi concluído com sucesso. Com uso da infraestrutura de testes concebida e implementada nos ciclos anteriores, o aplicativo foi alterado pela empresa para não mais apontar diretamente para o banco de dados, mas sim para a infraestrutura de testes da USP. A Infraestrutura de testes foi conectada no banco de dados através de uma rede privada virtual (*Virtual Private Network – VPN*), fornecida pela empresa. O teste realizado foi síncrono, com os pesquisadores do CIAAM/USP e os testadores da empresa realizando o teste ao mesmo tempo. Solicitou-se da empresa utilizar o aplicativo normalmente, simulando os fluxos esperados de uso do aplicativo, desde o cadastro até a utilização normal. Os dados gerados são válidos, porém sintéticos, dado o risco de identificação (como CPF e CEP válidos mas sem ligação com indivíduos reais, nomes fictícios ou dos testadores da empresa).

Neste sentido, ao utilizar o aplicativo modificado para os testes, todos os dados coletados que podem ser enviados para servidores externos foram trafegados corretamente para o banco de dados, porém passando por um túnel lógico no ambiente de testes, o que possibilita a captura de todos os pacotes de rede. Tal ambiente fornece acesso privilegiado para o pesquisador do CIAAM/USP, portanto, a infraestrutura foi desmontada após os testes.

Tecnicamente, foram estabelecidos dois túneis bidirecionais usando ferramentas de rede (socat) e uma *bridge* virtual para os servidores próprios da empresa (banco de dados) usando a conexão de VPN. Os pacotes de redes foram logados em formato convencional *Packet Capture* (PCAP) e foram analisados manualmente. Para a empresa, o teste é transparente; porém o pesquisador CIAAM/USP logrou logar e abrir todos os pacotes de rede trafegados durante o teste.

Nos testes realizados neste ciclo, identificou-se que a empresa seguiu as recomendações do ciclo anterior e utilizou criptografia em trânsito para a comunicação com o banco de dados. Tal utilização é aceitável para uso civil e garante a privacidade dos dados.

Para os dados coletados pela empresa, o teste comprova que não há dados às claras trafegando que possam identificar um usuário. Os dados às claras que foram identificados são unicamente do protocolo de segurança utilizado e da negociação algorítmica com o servidor, não restando dúvidas sobre a utilização de criptografia adequada para os dados pessoais.

Foram identificados dois riscos, comunicados em reunião para a ANPD:

- A empresa pode trafegar dados fora do túnel, para uso próprio. Parte-se da boa-fé da empresa de participar corretamente do teste, trafegando os dados diretamente pelo túnel. No entanto, tecnicamente, o túnel e o ambiente montado não garantem que a empresa trafegue todos os dados de forma a serem auditados.
- O túnel fornecido pela empresa trafega unicamente dados coletados a serem armazenados na infraestrutura da empresa, sem cobrir acessos a terceiros. Sabe-se que, pela arquitetura da solução proposta pela empresa, há acessos frequentes a servidores de geolocalização com função de mapas (como a API do Google Maps). Apesar de a empresa afirmar nos documentos que não fornece aos terceiros nenhum dado que possa identificar um indivíduo, tecnicamente tais terceiros podem identificar o usuário. Ao solicitar um serviço de terceiro por meio de uma API diretamente do aplicativo, mesmo sem identificação ou com identificação pseudonimizada pela Synapse, é possível que o terceiro utilize meios próprios (como *cookies*, *fingerprint*, identificação do dispositivo, acordo com operadora de celular etc.) para efetivamente correlacionar o acesso à API com o aparelho e com o indivíduo. A correlação permite que o terceiro, sem conhecimento da Synapse, saiba onde o indivíduo está localizado, mesmo que tal função esteja desligada no celular, como consequência do uso do aplicativo. Tal identificação extrapola os testes efetuados e deve ser discutida pela equipe regulatória, pois as soluções técnicas existentes para completa anonimização contra terceiros são inviáveis do ponto de vista de estratégia de negócios, impactando a empresa financeiramente e degradando o desempenho do aplicativo. Sugere-se que tal comportamento esteja claro para o usuário, que deve consentir com tal prática. Adicionalmente, cumpre-se mencionar que as soluções técnicas avaliadas pela equipe técnica do CIAAM/USP foram: (i) próprio serviço de mapas em servidor próprio, considerada inviável devido ao custo elevado e qualidade inferior à do Google Maps; e (ii) solução intermediária (*proxy*), o servidor da Synapse faria a requisição ao Google Maps e o tráfego ficaria concentrado na empresa Synapse. Nesse contexto, há limitações, pois se o celular do usuário for Android, o Google pode identificar quem fez a requisição. Além disso, há o custo elevado do *proxy* e da internet.

3) *Teste 3: IA Adversarial em ambiente controlado*

Este teste não foi realizado integralmente, pois, dado que os dados estão corretamente protegidos, não há o que a IA adversarial identifique no conjunto de dados. O planejamento inicial era subir uma IA local para análise de dados e identificação de possíveis excertos que poderiam ser utilizados para identificação de indivíduos; porém, como os dados foram corretamente protegidos, tecnicamente, não haveria como uma IA

inferir qualquer identificador a partir dos dados trafegados. Como o risco foi mitigado pela criptografia e para completude do teste, a equipe do CIAAM/USP utilizou uma IA comercial em nuvem para analisar o conjunto de dados (Gemini 3.5 Thinking) para certificar-se de que todos os dados foram analisados e, conforme esperado, não foi possível identificar qualquer dado passível de ataque de violação de privacidade.

4) *Teste 4: frente algorítmica – execução de código em ambiente controlado*

Anteriormente, no teste do Ciclo 2, o algoritmo para detecção de densidade de pessoas em um ambiente, parte da solução da Synapse, foi testado com dados fornecidos pela empresa. Já no Ciclo 3, o algoritmo foi testado com imagens sintéticas, com as variações esperadas (com e sem anonimização prévia usando filtro gaussiano para borrar as faces) e o algoritmo se comportou conforme o esperado, efetivamente não identificando indivíduos. O teste foi realizado em máquina da USP isolada da rede.

C) **Governança, segurança da informação e privacidade em IA**

No que tange à governança, à segurança da informação e à privacidade, a equipe técnica do CIAAM/USP concluiu que, com ressalvas, houve avanço substancial até a conclusão do Ciclo 3. Isso demonstrou, novamente, o impacto positivo da implementação das boas práticas recomendadas.

Com o objetivo de utilizar como parâmetro a ISO/IEC 27001, foi elaborado o seguinte quadro pela equipe técnica do CIAAM/USP para sistematizar os indicadores de avaliação e a respectiva classificação:

Tabela: Classificação de Indicadores do sistema Trajetto

Indicador de Avaliação	Classificação
Conclusão Preliminar:	Suficiente com condicionantes.
Recomendação Principal:	Prosseguir com testes e solicitar complementação.
Nível Agregado de Risco Inerente:	Alto (devido ao tratamento contínuo de geolocalização e padrões de mobilidade).
Nível Agregado de Risco Residual:	Médio (devido aos controles criptográficos fortes empregados, mas impactado pela ausência de planos testados).
Mitigações Efetivadas Validadas:	Arquitetura em duas camadas com isolamento de cofres. Criptografia forte

	vinculada ao contexto e implementada via gerenciamento de chaves terceirizado. • Comunicação em trânsito com protocolo de segurança da camada de transporte implementado e ausência de dados em claro no tráfego capturado durante os testes de tunelamento.
Mitigações Planejadas Críticas:	Pendentes: • Inventário seguro de API (Google Maps); • Implementação e comprovação de filtros contra falsificação de informações e proteção via <i>Web Application Firewall</i> (WAF); • Comprovação material e visual de transparência ao titular (Art. 9 e 20 da LGPD).

Fonte: ANPD

Quanto aos resultados alcançados pela equipe técnica e aprofundados no parecer técnico do CIAAM/USP, indicou-se que o sistema “Trajetto” demonstra maturidade avançada em sua arquitetura de segurança por *design*, utilizando separação de bases lógicas e pseudonimização na primeira camada e criptografia na segunda camada. O descarte de imagens de câmeras mitigou os riscos associados ao tratamento de imagens faciais, tratando dados de coordenadas. Contudo, a lacuna agora não reside mais na ausência de documentação formal, visto que o RIPD, a AIDH e o Plano de Incidentes foram entregues e considerados estruturalmente suficientes, mas sim na falta de evidências materiais de sua implementação prática. Ademais, faltam comprovações visuais de transparência (banners, telas), teste contra falsificação de informações e detalhamento do inventário de API. Em suma, a continuidade dos testes é viável sob monitoramento contínuo.

4.1.4 Fase C

A elaboração do relatório de êxitos integra a Fase C da metodologia desenvolvida pelo CIAAM/USP e validada pela Comissão de Sandbox da ANPD. Essa etapa possui função complementar aos pareceres técnicos, porque, enquanto os pareceres avaliam a adequação da solução, os riscos, as medidas de mitigação e as condicionantes, o relatório

de êxitos sistematiza os aprendizados produzidos durante a testagem e verifica os benefícios gerados tanto para a agência reguladora quanto para a empresa participante.

Para orientar essa avaliação, o relatório foi dividido em campos específicos a serem preenchidos pela ANPD e pela empresa. Em cada campo, deveriam ser indicados o resultado observado e a evidência correspondente, permitindo relacionar as conclusões às informações efetivamente coletadas durante o ciclo.

Tabela: Campos de avaliação sob a perspectiva da ANPD

CAMPO	OBJETO DE AVALIAÇÃO
Aprendizado regulatório	Conhecimentos produzidos sobre a tecnologia, os riscos, as obrigações aplicáveis e os critérios necessários à supervisão.
Proteção do usuário ou do consumidor	Riscos identificados para os titulares e usuários, bem como salvaguardas necessárias para a proteção de seus direitos.
Subsídios para nova norma	Possíveis orientações, critérios, parâmetros ou lacunas regulatórias que possam fundamentar futuras ações normativas ou de supervisão.
Efeito no mercado ou na concorrência	Potencial da solução para estimular inovação responsável, ampliar a oferta de serviços ou produzir efeitos concorrenciais, positivos ou negativos.

Fonte: ANPD

Tabela: Campos de avaliação sob a perspectiva da empresa

CAMPO	OBJETO DE AVALIAÇÃO
Evoluções técnicas	Aperfeiçoamentos realizados na arquitetura, nos modelos, nos controles de segurança, nos fluxos de dados e nos mecanismos de rastreabilidade.
Aprendizados regulatórios	Compreensão das obrigações jurídicas e incorporação dos requisitos regulatórios ao desenvolvimento da solução.
Transparência e explicabilidade	Medidas adotadas para informar os titulares, explicar o funcionamento da solução e tornar as recomendações ou resultados compreensíveis e auditáveis.
Direitos dos titulares	Mecanismos destinados a viabilizar acesso, correção, eliminação, anonimização, oposição e revisão, quando aplicáveis.
Autorizações e análises pendentes	Documentos, validações, testes ou providências ainda necessários para a continuidade ou ampliação da solução.
Outros aprendizados	Efeitos organizacionais, operacionais ou estratégicos produzidos pela participação no Sandbox.

Fonte: ANPD

O relatório também contemplou uma conclusão conjunta, na qual deveriam ser registrados: (i) a classificação do êxito do projeto como integral, parcial ou inexistente; (ii) a principal evidência de êxito; e (iii) os próximos passos recomendados. Essa estrutura permitiu uma avaliação multidimensional do Ciclo 3, evitando que o êxito fosse medido apenas pelo funcionamento técnico da solução. No Ciclo 3, a conclusão conjunta foi pelo êxito integral para o escopo de testagem realizado nesta etapa.

Por fim, quanto ao Plano de Descontinuidade (PdD) da empresa Synapse, é possível afirmar que houve avanço na análise realizada no Ciclo 3, sendo observado que

uma dificuldade identificada no plano foi a tentativa de formalizar, com algumas empresas, como o Google, cláusulas contratuais que assegurassem o não uso dos dados dos usuários. Contudo, esses fornecedores já informam expressamente em sua documentação que não utilizam os dados para finalidades próprias. Também, foi realizado um trabalho relevante na estrutura técnica do banco de dados, com a segregação das informações para ampliar a segurança quando for iniciado o processo de anonimização permanente e o mapeamento das API de terceiros.

4.1.5 Resultados do Ciclo 3 sob a perspectiva da ANPD

No que tange à dimensão proporcionada pelo aprendizado regulatório, o Ciclo 3 permitiu avançar de uma avaliação centrada na existência formal de documentos para uma análise probatória da implementação efetiva dos controles. A entrega de RIPD, AIDH, plano de resposta a incidentes, políticas de governança de IA e de dados, nomeação/atuação do Encarregado pelo Tratamento de Dados Pessoais, registros de operações de tratamento e documentos técnicos ampliou a base de avaliação sobre os riscos, a maturidade ética e a governança da aplicação Trajetto.

Dessarte, o principal aprendizado regulatório é que, em sistemas de recomendação baseados em geolocalização, a formalização de instrumentos de governança é necessária, mas insuficiente para demonstrar conformidade e justificar a progressão segura no Sandbox. É indispensável comprovar, por evidências materiais e auditáveis, a implementação prática dos controles previstos, especialmente em relação à transparência e acessibilidade ao titular, inventário de API e terceiros, testes contra falsificação de informações e WAF, supervisão humana, matriz de responsabilidade e prestação de contas operacional e atualização dos instrumentos de cooperação com parceiros.

Em relação às evidências coletadas durante o aprendizado regulatório, verificou-se que o Parecer Técnico Consolidado do Ciclo 3, que classifica o projeto como “suficiente com condicionantes”, registra risco inerente alto e risco residual médio, valida a mitigação por arquitetura em duas camadas e protocolo de segurança da camada de transporte, mas aponta pendências relativas a inventário de API, transparência ao titular, testes contra falsificação de informações e WAF. Já nas Fichas A3 e A4 do Ciclo 3, baseadas na análise regulatória dos documentos enviados pela Synapse e na tabela resumo de evolução dos ciclos, que registram a entrega e a análise dos documentos supervenientes, notou-se a necessidade de ajustes pontuais e a evolução da pontuação no *framework* geral de 45 para 68 pontos, com classificação de risco ético médio.

Em seguida, no que se refere à proteção do usuário ou consumidor, o Ciclo 3 reforçou a identificação de riscos relevantes aos titulares e usuários, especialmente decorrentes da geolocalização contínua, gerenciamento de sessões, do histórico de deslocamentos, da integração com serviços externos e da possibilidade de reidentificação indireta. Também evidenciou riscos específicos para usuários em situação de vulnerabilidade, como pessoas com deficiência, mobilidade reduzida, baixa literacia digital ou dependência de informações rápidas em contexto de mobilidade.

Houve avanço na mitigação de riscos técnicos, com criptografia em trânsito e em repouso, segregação de bases e descarte de imagens faciais. Contudo, persistem lacunas quanto à acessibilidade da interface, à clareza das informações prestadas ao titular, à demonstração prática dos controles de transparência e ao eventual tratamento de dados relacionados à deficiência. O ciclo também reforçou que os direitos dos titulares devem

ser tratados como requisito funcional do produto, com mecanismos claros para informação sobre:

- coleta e uso de localização;
- gestão de permissões;
- revogação de compartilhamento;
- exclusão ou *download* de dados;
- contato com o encarregado de dados; e
- solicitação de informações adicionais.

Cabe, ainda, esclarecer se as informações de acessibilidade serão apresentadas indistintamente a todos os usuários ou se haverá autodeclaração de deficiência ou mobilidade reduzida. Caso a empresa opte pela autodeclaração do titular, recomenda-se que a base legal para tal tratamento seja explicitada. Na segunda hipótese, devem ser explicitadas a base legal, a finalidade, a necessidade, a proporcionalidade, as medidas de minimização e as alternativas disponíveis em caso de negativa de consentimento.

Nesse sentido, as evidências concernentes à proteção do usuário ou consumidor podem ser destacadas da seguinte forma:

- Parecer Técnico Consolidado do Ciclo 3;
- Análise regulatória dos documentos enviados pela Synapse, especialmente os comentários sobre experiência do usuário, RIPD, política de governança de dados e registro de operações;
- Matriz de Transparência;
- Análise da experiência do usuário da aplicação Trajetto;
- Fichas A3 e A4 do Ciclo 3, que apontam o atendimento parcial das medidas de transparência e a necessidade de comprovação concreta por telas, *banners*, notificações, conteúdos informativos, *model cards* e canais de privacidade.

Ademais, em relação à dimensão identificada como subsídio para nova norma, observou-se que o Ciclo 3 ainda não permite formular subsídios conclusivos para nova norma, mas produziu aprendizados regulatórios mais densos e replicáveis, capazes de orientar futuras ações de supervisão, recomendações ou instrumentos regulatórios da ANPD sobre sistemas de IA baseados em geolocalização, mobilidade e recomendação automatizada.

Entre os principais aprendizados, pode-se destacar:

- a necessidade de diferenciar documentação formal de evidência material;
- a definição de critérios de progressão condicionada em ambientes experimentais;
- a exigência de mapeamento preliminar das API de terceiros.;
- a comprovação de transparência e explicabilidade em camadas;
- a realização de testes de robustez, estresse e contra falsificação de informações;
- a governança ativa e independente do encarregado de dados;
- o versionamento de modelos, parâmetros e *pipelines*; e

- a aderência entre o modelo testado no Sandbox e eventual implementação operacional.

O ciclo 3 também confirmou que transparência, explicabilidade e acessibilidade permanecem pontos centrais de evolução. Embora a análise da experiência do usuário indique fluxo relativamente organizado no aplicativo, com etapas de cadastro, pesquisa, visualização e realização de trajeto, perfil e compartilhamento de localização, ainda persistem fragilidades quanto à clareza sobre o tratamento de dados, à informação contextual no momento da recomendação, à acessibilidade comunicacional e à garantia de direitos de usuários vulneráveis. A Matriz de Transparência avança na identificação de interlocutores e temas informativos, mas ainda demanda exemplos concretos de mensagens, telas, notificações, *model cards* e explicações em camadas dirigidas a usuários, operadores, equipe técnica e reguladores.

Nesse passo, as evidências identificadas na dimensão correspondente aos subsídios para nova norma são:

- Parecer Técnico Consolidado do Ciclo 3, especialmente as condicionantes recomendadas;
- Fichas A3 e A4 do Ciclo 3;
- Análise dos documentos regulatórios;
- Matriz de Transparência da Synapse;
- Análise de experiência do usuário da aplicação Trajetto;
- Documento de *links*, que organiza os pareceres como instrumentos dedicados ao aprendizado regulatório, à proteção do usuário, à validação técnica e à redução de incertezas.

Finalmente, a respeito da dimensão que avalia o efeito no mercado ou na concorrência, registrou-se que o Ciclo 3 demonstra potencial de estímulo à inovação responsável em soluções de mobilidade urbana inteligente, ao permitir que uma solução com potencial de escalabilidade seja testada sob supervisão regulatória e com condicionantes de proteção de dados. Ainda não há evidência suficiente de efeito concreto sobre mercado ou concorrência, pois o projeto segue em fase de validação e depende da comprovação de controles técnicos, regulatórios e operacionais antes de eventual escala. A despeito disso, o efeito positivo, até o momento, é a redução de incertezas regulatórias e a definição de requisitos mínimos para testes em ambiente controlado, sem caracterizar autorização ampla para exploração comercial plena.

Destarte, quanto às evidências da mesma dimensão que apura o efeito no mercado ou na concorrência da inovação, foram identificados os seguintes documentos:

- Parecer Técnico Consolidado do Ciclo 3, com recomendação favorável com condicionantes;
- Análise dos ACT e instrumentos com parceiros, que recomenda explicitar:
 - a finalidade regulatória e científica da cooperação;
 - os limites do ambiente de teste; e
 - a inexistência de exploração comercial plena;
- Plano de Descontinuidade (PdD), que trata a evolução do projeto como processo controlado, reversível e documentado.

4.1.6 Resultados do Ciclo 3 sob a perspectiva da Synapse

Sob a perspectiva da empresa Synapse, o Ciclo 3 resultou em melhorias na usabilidade e na experiência do usuário da aplicação Trajetto. Nesse sentido, a participante informou que trabalhou no aprimoramento da interface e que continuou o aperfeiçoamento da comunicação no *site*, no âmbito das medidas de transparência e explicabilidade.

Já no campo regulatório, a empresa informou que os aprendizados permaneceram conforme a atualização anterior. Da mesma forma, não foram relatadas mudanças em relação aos direitos dos titulares no decorrer do Ciclo 3.

Quanto às análises pendentes, a Synapse registrou a necessidade de entregar a evidência de não identificação solicitada pela equipe do CIAAM/USP, referente ao modelo denominado “point-to-point” na ficha C1.

Finalmente, em relação ao plano de descontinuidade, as equipes indicaram que uma dificuldade identificada foi a tentativa de formalizar, com algumas empresas, como o Google, cláusulas contratuais que assegurassem o não uso dos dados dos usuários. Contudo, esses fornecedores já informam expressamente em sua documentação que não utilizam os dados para finalidades próprias. Também, foi realizado um trabalho relevante na estrutura técnica do banco de dados, com a segregação das informações para ampliar a segurança quando for iniciado o processo de anonimização permanente. Além disso, foi feito o mapeamento das API de terceiros, cuja documentação está disponível no diretório do projeto.

Em suma, no que tange à conclusão conjunta alcançada pela ANPD e pela empresa participante, concluiu-se que houve êxito integral do projeto no Ciclo 3, restrito ao escopo de testagem e preservadas as condicionantes, diante da documentação apresentada e das contribuições demonstradas durante as interações entre as equipes da participante, do CIAAM/USP e da agência reguladora.

4.2 Prevvine

Esta seção consolida as informações, os registros e as evidências produzidos nas fases A, B e C do Ciclo 3 de testagem da solução apresentada pela empresa Prevvine. Descreve-se a caracterização técnico-funcional da solução, seu nível de maturidade, os testes realizados e a avaliação preliminar dos riscos, das medidas de mitigação e dos aspectos de governança. A seguir, serão apresentados os principais dados, informações e conhecimentos extraídos do Ciclo 3, considerando os seis pilares da metodologia de avaliação das soluções de inteligência artificial:

- Dimensão técnico-funcional: análise das fichas técnicas A1 e A2, que descrevem a solução da Prevvine, sua arquitetura, seus componentes tecnológicos, os fluxos operacionais e as características do modelo de linguagem utilizado;
- Dados, privacidade e segurança: análise das informações coletadas por meio dos instrumentos aplicados nas subfases A3, A4, B1 e B2, incluindo o tratamento de dados pessoais, as medidas de segurança,

os controles de acesso, os mecanismos de transparência e as salvaguardas adotadas pela empresa;

- Maturidade tecnológica (TRL e ETC identificados): nas fichas A1 e A2, a Prevvine apresentou sua autoavaliação do nível de maturidade da solução e de seus componentes críticos. Nas fichas B1 e B2, a equipe do CIAAM/USP analisou essa avaliação à luz das evidências apresentadas e dos resultados observados durante a testagem;
- Realismo experimental e benefícios esperados: avaliação dos testes realizados e das evidências coletadas na Fase B, considerando a correspondência entre o ambiente experimental, os usos esperados da solução e os benefícios declarados pela Prevvine, especialmente em relação à utilização do sistema de inteligência artificial no contexto da saúde;
- Riscos, mitigação e impacto residual: análise das fichas de avaliação de riscos elaboradas na subfase B2, contemplando os riscos identificados, sua classificação quanto à probabilidade e ao impacto, as medidas de mitigação efetivadas ou planejadas e o nível de risco residual da solução; e
- Governança, transparência e aprendizagem regulatória: análise da documentação institucional, da estrutura organizacional, dos papéis e responsabilidades, dos mecanismos de supervisão humana, da explicabilidade da solução e da aplicação da matriz de transparência. Esses elementos foram examinados especialmente na Fase B1, permitindo identificar avanços, lacunas e aprendizados relevantes para a empresa e para a ANPD.

Considerando que o Ciclo 3 foi concluído em suas três fases: (i) Fase A (pré-teste); (ii) Fase B (teste); e (iii) Fase C (pós-teste). A seguir são apresentados os principais aspectos identificados em cada uma dessas etapas.

4.2.1 Documentação principal

Preliminarmente, observa-se que os principais documentos referidos neste item do relatório encontram-se organizados conforme a descrição a seguir, por fases, para atendimento dos requisitos pertinentes ao Ciclo 3 deste Sandbox Regulatório:

- Fase A
 - Fichas A1 e A2;
 - Fichas A3 e A4.
- Fase B
 - Ficha B1 - *Framework* geral;
 - Planilha de Testes com 30 Vinhetas;
 - Teste 4 - Explicabilidade para Transparência;
 - Matriz de Transparência;
 - Parecer Técnico Consolidado;
 - Relatório da plataforma HarpIA sobre testes realizados.

- Fase C
 - Ficha C1 – Êxitos (ANPD e empresa);
 - Plano de Descontinuidade.

4.2.2 Fase A

Inicialmente, no que se refere à Ficha A1, ao se analisarem os itens relativos à arquitetura do modelo, ao fluxo de dados, às APIs, à inteligência artificial e aos protocolos de segurança, a participante Previne informou que não houve alterações entre os Ciclos 02 e 03 de testagem.

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

- motor de IA;
- camada de anonimização;
- sistema NTAV, responsável pela verificação bibliográfica; e
- infraestrutura de conformidade LGPD.

No Ciclo 3, o motor de IA, a camada de anonimização e a infraestrutura de conformidade mantiveram a classificação TRL 7. Entretanto, o sistema NTAV teve sua classificação reduzida de TRL 7 para TRL 6 em razão dos resultados dos testes realizados na plataforma Harpia.

Ademais, foram identificadas alucinações, mistura de dados reais com informações inventadas, *links* quebrados e pesquisas realizadas pelo modelo sem resultados úteis. Também, foi apontada a ausência de informação explícita ao usuário sobre possíveis falhas, incluindo a geração de referências inexistentes. Esses achados evidenciaram limitações na acurácia da informação entregue ao médico, apesar da manutenção da decisão clínica final sob responsabilidade humana.

Diante do exposto, embora a participante não tenha declarado mudanças nos componentes descritos na Ficha A1 entre os Ciclos 02 e 03, a avaliação de maturidade tecnológica foi revista à luz das evidências da testagem. A redução do TRL do sistema NTAV demonstra a importância de confrontar a classificação inicialmente atribuída aos componentes com seu desempenho observado no ambiente experimental.

Já em relação às Fichas A3 e A4 da Previne, procedeu-se à comparação das informações apuradas no Ciclo 2 com a situação verificada no início do Ciclo 3. A análise abrangeu o marco regulatório aplicável, os controles internos, os processos operacionais e a arquitetura tecnológica da solução de inteligência artificial em saúde.

Para avaliar a incidência normativa, foram considerados três aspectos:

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

No âmbito da proteção de dados, foram considerados o Decreto nº 12.881/2026 e a Resolução CD/ANPD nº 4/2023, sobre dosimetria das sanções, inclusive diante do

tratamento de dados sensíveis de saúde. Embora não regulem diretamente o ato assistencial, essas normas são relevantes para a avaliação do risco regulatório associado à solução e ao seu uso por instituições de saúde.

Quanto à saúde digital, foi examinada a seguinte legislação:

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

Nesse sentido, a análise destacou que, se o *chatbot* realizar triagem ou apresentar recomendações que influenciem a conduta assistencial, a sua função deve permanecer claramente delimitada como apoio, sem substituir o julgamento clínico do profissional. Com isso, as normas relativas à telenfermagem e ao exercício da enfermagem têm incidência condicionada à utilização da ferramenta nesse contexto.

Também, foi examinado o possível enquadramento da solução como Software as a Medical Device (SaMD). A RDC ANVISA nº 657/2022 trata da regularização de *software* como dispositivo médico, conforme a finalidade pretendida e os efeitos do sistema sobre decisões clínicas, condutas médicas ou informações diagnósticas. Nessa linha, a sua aplicação deve ser analisada em conjunto com a RDC nº 751/2022, que disciplina classes de risco e regimes de regularização, e com a RDC nº 848/2024, relativa a requisitos essenciais de segurança e desempenho. Assim, a necessidade de regularização sanitária dependerá da finalidade e das funcionalidades efetivamente oferecidas pela solução.

No exame dos documentos e das políticas internas, identificou-se a necessidade de delimitar com maior clareza os contextos em que a ferramenta não deve ser utilizada, especialmente diante de emergências. Como contribuição para o ciclo, recomendou-se incluir nas políticas e nas informações apresentadas previamente aos usuários um aviso expresso de que o sistema não deve ser utilizado nesses casos urgentes.

Além disso, a análise dos dados e dos fluxos de tratamento registrou pendências de documentação. Recorde-se que o mapeamento de dados ainda não havia sido apresentado de forma suficiente para identificar integralmente a coleta de dados dos médicos, os meios de acesso à plataforma e os agentes envolvidos em cada etapa. Também, não havia conexão com prontuários eletrônicos prevista para o escopo então avaliado. Com base no fluxo descrito, a operação foi caracterizada como pseudonimização, pois subsiste possibilidade de reidentificação, inclusive para fins de pesquisa interna; por isso, não deve ser tratada como anonimização.

Paralelamente, verificou-se que a definição da base legal e dos papéis dos agentes de tratamento depende da consolidação desse mapeamento. Como hipótese de trabalho, o hospital contratante pode atuar como controlador quando define a finalidade da triagem e os elementos essenciais do tratamento, enquanto a empresa fornecedora pode atuar como operadora se tratar os dados segundo as instruções do hospital. Nesse passo, a possibilidade de controladoria conjunta deve ser examinada caso as partes determinem conjuntamente finalidades ou meios essenciais. Já a caracterização definitiva demanda a verificação das decisões efetivamente tomadas por cada agente e a cadeia completa do tratamento.

Ademais, foi apontada a necessidade de avaliar a transferência internacional de dados, pois o processamento por fornecedores ou serviços localizados no exterior pode envolver dados pessoais de autores de artigos científicos e, eventualmente, informações identificáveis de terceiros constantes desses artigos. Nesses casos, especialmente,

devem ser verificados os requisitos do Capítulo V da LGPD e da Resolução CD/ANPD nº 19/2024, inclusive a adoção do mecanismo de transferência pertinente, como cláusulas-padrão contratuais quando aplicáveis.

Em seguida, no campo da transparência e da conformidade, o documento registrou pendências relativas à explicação dos critérios de seleção das referências utilizadas pelo sistema e à possibilidade de o usuário consultar a relação completa das fontes acessadas, para além das referências exibidas em cada resposta. Com isso, a avaliação também apontou a necessidade de demonstrar, por evidências, a implementação das políticas apresentadas e de esclarecer os fluxos de uso da inteligência artificial e as responsabilidades de governança.

Quanto à coordenação regulatória, não foram identificadas alterações substanciais na lista de normas setoriais considerada no ciclo anterior, mas a análise aprofundou a incidência das regras de saúde digital e de regularização de SaMD. Dessa forma, o documento registra a expectativa de interlocução entre a ANPD e autoridades do setor de saúde, porquanto, enquanto não houver um arranjo específico mais amplo, foram apontados como instrumentos de coordenação o dever legal de articulação previsto na LGPD (art. 55-J), os acordos de cooperação técnica, a oitiva das autoridades setoriais e a atuação conjunta. Destarte, o Acordo de Cooperação Técnica entre a ANPD e a SENACON foi considerado exemplo de mecanismo para uniformizar entendimentos respeitando as competências de cada órgão.

Após, na avaliação dos eixos de governança e conformidade, a centralidade da dignidade humana e dos direitos fundamentais foi considerada consolidada. Permaneceram, contudo, condicionantes relativos à governança e à prestação de contas, à comprovação documental das avaliações de impacto e dos planos de mitigação, à existência de protocolo de avaliação de vieses, à clareza dos fluxos de uso da inteligência artificial e à conclusão das verificações de segurança. Ainda, foi apontada a necessidade de esclarecer os critérios de classificação das referências e permitir a consulta às fontes efetivamente utilizadas. Como os documentos complementares não foram apresentados na relação de materiais compartilhados até a conclusão do ciclo analisado, não foi possível verificar integralmente esses aspectos.

Em suma, as Fichas A3 e A4 do Ciclo 3 mostram que a avaliação regulatória da solução STAiDOC avançou no detalhamento dos seguintes pontos:

- normas aplicáveis;
- riscos associados à triagem em saúde; e
- mecanismos possíveis de coordenação institucional.

Contudo, permaneceram pendentes:

- a consolidação do mapeamento de dados;
- a definição fundamentada dos papéis dos agentes de tratamento;
- a comprovação da implementação das políticas e avaliações de impacto;
- a explicitação dos limites de uso em situações críticas; e
- a demonstração das medidas de transparência e governança.

Diante da análise, recomendou-se à ANPD:

- Promover interlocução com ANVISA, ANS e conselhos profissionais para avaliar possibilidades de coordenação regulatória no uso de sistemas de IA em saúde.
- Manter um mapeamento integrado e atualizado das normas aplicáveis, distinguindo regras setoriais, transversais e condicionadas ao uso da solução.
- Solicitar a complementação do mapeamento de dados e dos fluxos de tratamento, incluindo a identificação dos agentes e das respectivas responsabilidades.
- Verificar, à luz das funcionalidades e da finalidade pretendida, o eventual enquadramento da solução como SaMD e a classe de risco correspondente.
- Recomendar a inclusão de aviso claro de não utilização em situações de emergência e a delimitação dos usos que exigem intervenção de profissional de saúde.
- Solicitar evidências da implementação das políticas de governança, das avaliações de impacto, das medidas de mitigação e dos mecanismos de transparência.
- Avaliar a necessidade de salvaguardas para eventual transferência internacional de dados e de mecanismos que permitam a consulta das referências utilizadas pelo sistema.

4.2.3 Fase B

Inicialmente, a Fase B do Ciclo 3 reuniu a aplicação do *framework* geral, a preparação das vinhetas clínicas para o teste conversacional, a avaliação de explicabilidade e transparência, a atualização da matriz de transparência e a consolidação dos riscos e das medidas de mitigação em parecer técnico. Os instrumentos permitiram distinguir os critérios já atendidos, as medidas ainda dependentes de comprovação e as condições necessárias à continuidade dos testes.

A) Framework geral baseado em recomendações da UNESCO:

A Prevvine obteve 63 pontos de 100 no Ciclo 3, ante 60 pontos no Ciclo 2. O acréscimo de três pontos ocorreu exclusivamente no eixo de avaliação de impactos, que passou de zero para três pontos em razão do registro de plano de mitigação de riscos. As pontuações dos demais eixos permaneceram iguais: Eixo 1 (Centralidade da dignidade humana e dos direitos fundamentais), 20/20; Eixo 2 (Governança e responsabilidade clara), 7/15; Eixo 4 (Justiça, não discriminação e vieses algorítmicos), 7/15; Eixo 5 (Transparência, explicabilidade e comunicação), 11/15; Eixo 6 (Segurança técnica e robustez), 5/10; Eixo 7 (Proporcionalidade e necessidade), 10/10.

A pontuação do Ciclo 3 ainda corresponde a risco ético moderado a alto, sendo recomendada a sua mitigação antes da ampliação da solução. Nesse passo, apesar do avanço pontual no planejamento das medidas de mitigação, o resultado evidencia:

- a manutenção de lacunas de governança;

- avaliação de vieses;
- transparência; e
- robustez técnica.

B) Avaliação conversacional do modelo:

O conjunto de trinta vinhetas organizou cenários clínicos e suas variações para a aplicação dos prompts no Ciclo 3. Assim, o instrumento contém um mapa de cobertura de características demográficas e contextuais, incluindo:

- idade;
- gestação;
- deficiência;
- raça e etnia;
- renda;
- orientação sexual;
- região do Brasil; e
- pertencimento a populações prioritárias do SUS.

Com isso, a introdução dessas variações permite examinar, nos testes, a consistência das respostas e as possíveis diferenças de desempenho entre perfis de pacientes.

No entanto, apesar de a planilha disponibilizada registrar os casos e sua cobertura, ela não apresenta, por si só, resultados consolidados das respostas do sistema ou classificações de aprovação e falha. Além disso, o parecer técnico do Ciclo 3 registra que alterações na interface da solução afetaram a integração com a plataforma HarplA, impedindo temporariamente a execução regular das avaliações algorítmicas. Por essa razão, a existência das vinhetas não deve ser interpretada como comprovação de desempenho clínico, segurança ou ausência de vieses. Em resumo, a retomada e a conclusão dos testes dependem da estabilização da integração e do registro dos resultados pelos avaliadores.

Sendo assim, observa-se que os dados e informações gerados pela plataforma HarplA, possibilitam a compreensão global da dinâmica dos testes conduzidos para atendimento dos requisitos técnicos e regulatórios deste projeto piloto de sandbox regulatório de inteligência artificial e proteção de dados. Todo o conhecimento angariado por essa tecnologia é um dos principais pilares que permitiu sustentar as conclusões deste relatório.

C) Explicabilidade para transparência:

As informações apresentadas para o Ciclo 3 contêm registros relativos às camadas de governança institucional e de comunicação com usuários. Na camada destinada à alta administração e às estruturas de governança, o mapa de riscos éticos, legais e regulatórios foi aprovado, com indicação de que também devem ser considerados os pareceres das Fichas A3 e A4 e a análise da ANPD. A aderência ao marco regulatório e a definição das linhas de responsabilidade foram aprovadas com ressalvas. O critério relativo ao registro de incidentes e às medidas de mitigação foi reprovado.

Na camada destinada a pacientes, profissionais de saúde e demais usuários, a indicação de que se trata de um sistema de IA foi aprovada com ressalvas, diante da necessidade de aprimorar sua apresentação na interface. Foram reprovados os critérios relativos à comunicação da finalidade e dos limites de uso, à informação sobre as fontes utilizadas, à indicação de confiança ou incerteza das respostas, ao aviso sobre situações em que o sistema não deve ser empregado (exemplo: emergências) e à disponibilização de canal visível para revisão humana e reclamações. A avaliação do uso de linguagem clara recebeu registro de que a testagem e a adequação desse requisito ainda não haviam sido realizadas.

Finalmente, sobre o teste de explicabilidade e transparência, os resultados apontam a necessidade de transformar a documentação de governança em informações e funcionalidades verificáveis na interface. Em especial, devem ser demonstrados os avisos de limitação de uso, a apresentação das fontes, os canais de intervenção humana e a comunicação acessível aos diferentes destinatários.

D) Matriz de transparência:

A matriz de transparência manteve, no Ciclo 3, os registros do Ciclo 2 para seus cinco grupos de interlocutores: usuários, médicos, hospitais e clínicas, ANPD e demais reguladores, e equipes de tecnologia da informação e segurança. Assim, essa manutenção preserva as necessidades de informação já identificadas nos relatórios pretéritos sem registrar nova alteração de conteúdo para o ciclo em análise.

Para os usuários, permaneceram as recomendações de adequar as políticas à linguagem clara, definir o fluxo de revisão de decisões automatizadas, verificar a base legal aplicável ao tratamento e uniformizar a apresentação das informações nos diferentes canais. Para os médicos, continuaram em revisão as avaliações sobre o uso da ferramenta e permaneceu indicada a disponibilização de manual de uso, treinamento e informações sobre funcionalidades, fontes e limitações das respostas.

Quanto aos hospitais e clínicas, a matriz manteve a necessidade de informações principalmente sobre arquitetura, segurança, políticas e métricas. Para a ANPD e os demais reguladores, permaneceram previstos relatórios estruturados, registros de operação, métricas de viés, informações sobre incidentes e demonstrações das medidas de mitigação, ainda com necessidade de detalhar a forma de apresentação dessas evidências. Para a equipe de tecnologia e segurança, continuou indicada a definição de um ciclo interno de revisão técnica, apoiado por métricas, registros, alertas e documentação atualizada das integrações.

E) Parecer técnico: mapa de riscos e mitigações

Conforme o *framework* estabelecido pela ISO 27001, o parecer técnico consolidado do Ciclo 3 classificou o risco inerente agregado como crítico e o risco residual agregado como médio, concluindo pelo prosseguimento dos testes com condicionantes.

Neste Ciclo 3, a matriz individualizou seis riscos classificados como altos:

- reidentificação de titulares;
- indisponibilidade do serviço de bases bibliográficas PubMed e seus efeitos sobre a operação clínica;
- ataques à aplicação e *prompt injection*;
- bloqueio da integração com a plataforma HarpIA;

- geração de respostas clínicas inadequadas; e
- deficiências de transparência e explicabilidade.

No Eixo 3, o parecer técnico reconheceu os controles incorporados à solução, notadamente, a supressão contextual de determinadas categorias de dados pessoais, os mecanismos de bloqueio preventivo, a verificação de referências externas e a estruturação do Plano de Descontinuidade (PdD). No Eixo 4, assinalou limitações relevantes como: particularidades no método criptográfico que podem manter o risco de reidentificação; a possibilidade de o filtro de supressão deixar passar identificadores fora dos padrões previstos; a ausência de comprovação de proteção por *Web Application Firewall* (WAF); e a dependência da disponibilidade do PubMed sem mecanismo de contingência demonstrado.

Também foram registradas pendências relativas à apresentação das políticas de privacidade e de retenção e descarte de dados, à correção das deficiências de explicabilidade e à integração operacional com a HarPIA. Nesse sentido, o parecer condicionou o avanço dos testes à restauração dessa integração, à entrega e validação das políticas pendentes, ao aperfeiçoamento da supressão e da proteção criptográfica dos dados, à comprovação de medidas de segurança da aplicação e à continuidade do serviço, e à definição de prazos para que os especialistas concluam as avaliações na plataforma.

Em conjunto, os resultados da Fase B demonstram avanço limitado na pontuação ética e maior detalhamento dos cenários de teste e dos riscos da solução. Além disso, a continuidade experimental foi considerada possível sob condicionantes, com prioridade para a conclusão das avaliações práticas, a correção das vulnerabilidades técnicas e a comprovação das medidas de transparência destinadas a médicos, pacientes, instituições de saúde e autoridades reguladoras.

4.2.4 Fase C

A Fase C reuniu os resultados e os aprendizados do Ciclo 3 sob duas perspectivas complementares seguindo a mesma estrutura do Ciclo 2: os êxitos para a ANPD e os êxitos para a Previne.

Inicialmente, a Ficha C1 relacionou, para cada dimensão examinada, o resultado observado e a evidência correspondente. Ao final, registrou uma conclusão conjunta sobre o êxito do projeto e as providências necessárias para o ciclo seguinte.

4.2.5 Resultados do Ciclo 3 sob a perspectiva da ANPD

O acompanhamento da plataforma STAiDOC permitiu aprofundar a avaliação regulatória de sistemas de IA utilizados no suporte à decisão médica. Em especial, a identificação de vulnerabilidades técnicas, lacunas documentais e limitações de transparência forneceu evidências sobre os controles que precisam ser demonstrados durante a supervisão de uma solução dessa natureza. Além disso, o parecer técnico consolidou critérios de avaliação relativos à proteção de dados, à segurança da informação, à explicabilidade e à gestão de incidentes.

No que tange à proteção dos usuários e titulares de dados, foram examinadas medidas incorporadas à arquitetura, notadamente:

- a supressão contextual de categorias de dados pessoais;

- os mecanismos de bloqueio preventivo; e
- os registros de pesquisador CIAAM/USP.

A análise também evidenciou riscos ainda não suficientemente mitigados, especialmente:

- a possibilidade de reidentificação;
- as lacunas de proteção da aplicação; e
- as deficiências de informação apresentadas aos usuários.

Assim, a identificação desses riscos permitiu estabelecer condicionantes e hipóteses de pausa dos testes, associadas à exigência de evidências para a sua retomada.

Os resultados do Ciclo 3 também produziram subsídios para futuras orientações regulatórias sobre explicabilidade, transparência, distinção entre anonimização e pseudonimização e comprovação de controles de governança em sistemas de IA em saúde. Em suma, todos esses subsídios constituem aprendizado experimental.

Finalmente, sob a perspectiva mercadológica, é possível afirmar que a exigência de controles verificáveis de segurança, continuidade operacional e transparência pode orientar o desenvolvimento responsável de soluções semelhantes.

4.2.6 Resultados do Ciclo 3 sob a perspectiva da Prevvine

Em primeiro lugar, no plano técnico, a Ficha C1 registra o restabelecimento da integração entre a STAiDOC e a plataforma HarpIA, que havia constituído um obstáculo operacional durante a testagem. Com isso, a integração viabilizou a aplicação das trinta vinhetas clínicas previstas para o ciclo. Esse registro representa uma evolução em relação ao parecer técnico emitido durante o Ciclo 3, que ainda apontava o bloqueio da comunicação entre as plataformas como condicionante imediata.

Em segundo lugar, houve o registro da correção da resposta fornecida pela API. Nesse passo, artefatos técnicos internos, anteriormente removidos apenas na interface da STAiDOC, passaram a ser excluídos no lado do servidor, de modo que o campo de resposta disponibilize o texto clínico sem esses elementos. Assim, as referências bibliográficas passaram a incluir *links* diretos para os registros correspondentes no PubMed. Destarte, essas alterações melhoraram a consistência da informação entregue aos diferentes clientes da API, embora permaneça necessária a avaliação dos critérios de seleção e apresentação das fontes utilizadas pelo sistema.

Em terceiro lugar, no campo regulatório, a empresa conseguiu identificar com maior precisão as evidências documentais necessárias à avaliação de conformidade. Destacou-se a exigência de apresentação formal da Política de Privacidade e Proteção de Dados e da Política de Retenção e Descarte de Dados, cuja elaboração interna, sem disponibilização para exame, não permite verificar integralmente sua implementação. A análise também aprofundou a discussão sobre a posição da Prevvine na cadeia de tratamento de dados pessoais, especialmente quanto à distinção entre atuação como operadora e eventual controladoria conjunta com a instituição de saúde contratante.

A partir do exposto até o momento, a avaliação de transparência e explicabilidade delimitou as lacunas a serem enfrentadas nos próximos ciclos. Entre elas estão a documentação técnica do modelo e dos dados, a explicação dos critérios de classificação das fontes de referência, a comunicação dos limites de uso da ferramenta e a apresentação de avisos e canais adequados aos seus destinatários. Quanto aos direitos e

à participação dos usuários, permaneceram pontos de atenção relacionados ao mecanismo de *feedback* dos médicos, à diferenciação dos perfis de acesso de médicos e instituições de saúde e à demonstração de alertas sobre contextos de risco e situações em que a resposta do sistema não deve ser utilizada.

O Ciclo 3 também evidenciou a importância de controlar alterações na interface e no formato das mensagens durante a testagem. As modificações que interromperam temporariamente a integração com a HarpIA mostraram a necessidade de registrar, comunicar e validar mudanças técnicas com a equipe avaliadora antes de sua implantação no ambiente experimental.

4.2.7 Coleta e análise das evidências

A avaliação dos êxitos considerou, especialmente, a Ficha C1, o parecer técnico consolidado do Ciclo 3, as fichas A3 e A4 e os registros relativos à integração com o sistema HarpIA e à aplicação das vinhetas. A análise desses materiais permitiu distinguir a correção do impedimento operacional de integração das demais condicionantes, que continuaram sujeitas a implementação ou comprovação.

A conclusão conjunta classificou o êxito do projeto no Ciclo 3 como parcial. Como principal evidência positiva, registrou-se a manutenção da classificação “favorável com condicionantes”, associada à continuidade dos controles de proteção de dados reconhecidos pelo avaliador e ao restabelecimento da integração entre o sistema STAiDOC e a plataforma HarpIA. Além disso, o restabelecimento da integração com a HarpIA e a aplicação das trinta vinhetas clínicas representam avanços operacionais relevantes e permitiram retomar a avaliação prática da solução. Esses resultados, contudo, não demonstram, isoladamente, o atendimento integral das condicionantes técnicas e regulatórias. A análise das vinhetas deverá ser articulada às evidências de explicabilidade, segurança da aplicação, supressão de dados pessoais, continuidade operacional e apresentação das políticas documentais. Por essa razão, mantém-se a classificação de êxito parcial do Ciclo 3.

Para o avanço ao Ciclo 4, a ficha indicou o cumprimento das seguintes condicionantes: (i) apresentação das políticas de privacidade e de retenção e descarte; (ii) correção das lacunas do teste de explicabilidade para transparência; (iii) aperfeiçoamento da proteção criptográfica; (iv) comprovação de proteção da aplicação e de continuidade diante de indisponibilidade do PubMed; (v) definição de prazos para os avaliadores; e (vi) aperfeiçoamento da heurística de supressão de dados pessoais. A condicionante relativa à restauração da integração com a HarpIA foi registrada na Ficha C1 como concluída.

Quanto ao Plano de Descontinuidade (PdD), a empresa participante informou que não houve alterações em seu conteúdo durante o Ciclo 3 e disponibilizou o arquivo no formato sugerido pela equipe do CIAAM/USP.

4.3 Metatext

Esta seção apresenta o progresso técnico e regulatório do projeto GuardionAI, bem como a documentação considerada para a testagem neste ciclo. As atividades incluíram a aplicação do *framework* geral e a análise do fluxo de dados, permitindo levantar pontos de dúvidas e elencar normativas aplicáveis ao contexto da solução. Além disso, foi

elencado um conjunto de documentos a serem apresentados pela empresa para continuidade das atividades de testagem no próximo ciclo.

4.3.1 Documentação principal

Preliminarmente, observa-se que os principais documentos referidos neste item do relatório encontram-se organizados conforme a descrição a seguir, por fases, para atendimento dos requisitos pertinentes ao Ciclo 3 deste Sandbox Regulatório:

- Fichas A1 e A2
- Fichas A3 e A4
- Parecer técnico
- Fluxo de dados (*data flow overview*)
- Perguntas e respostas com equipe do CIAAM/USP (guardion-faq¹⁰)
 - Respostas da participante sobre pseudonimização, infraestrutura, métricas, supervisão e documentação de proteção de dados.
- Simulação de incidentes
 - Autoavaliação dos sete critérios de conformidade, incluindo as lacunas de impactos e vieses.

4.3.2 Registros do progresso técnico e regulatório

Preliminarmente, recorda-se que, sobre o projeto GuardionAI da participante Metatext, no Ciclo 1, houve a necessidade de ajuste do plano de sandbox e do respectivo plano de descontinuidade, bem como do cronograma de atividades em razão da estrutura e do momento da empresa. Já no Ciclo 2, foi necessário focar no atendimento dos requisitos das fichas A1 e A2, bem como na continuação da adequação do ambiente na nuvem para adequação ao ambiente de testes.

Após o estabelecimento das bases para o Ciclo 3, inicialmente, em relação aos eixos do *framework* geral baseado em recomendações da UNESCO, deve-se lembrar de que a consultoria do CIAAM/USP no âmbito do Projeto Piloto de Sandbox Regulatório de IA e Proteção de Dados adotou tal *framework* como instrumento metodológico para análise dos impactos de sistemas de inteligência artificial. Embora sua adoção possua caráter voluntário, o *framework* foi utilizado como referência de boas práticas internacionais alinhadas aos princípios constitucionais brasileiros.

Nesse sentido, verificou-se que a autoavaliação submetida pela participante indicou conformidade integral em cinco dos sete eixos analisados, evidenciando maturidade inicial da governança da solução. As principais lacunas identificadas

¹⁰ O documento guardion-faq reúne respostas da participante sobre a arquitetura, a infraestrutura e os controles declarados. Essas informações integram a análise documental do Ciclo 3, mas permanecem necessários esclarecimentos complementares e evidências de sua implementação. No Ciclo 4, serão aprofundadas as questões relativas ao momento da supervisão humana, ao treinamento e aprendizado durante o uso, à governança e revisão, à responsabilidade pela pseudonimização, ao tratamento dos dados, à geração de alertas e à exclusão em produção e backups. Os testes deverão verificar as declarações existentes e subsidiar as respostas às questões ainda não suficientemente esclarecidas.

concentram-se nos eixos 3 (Avaliação de impactos) e 4 (Justiça, não discriminação e vieses algorítmicos), que ainda dependem da implementação de mecanismos formais de monitoramento, documentação e mitigação de riscos¹¹.

Por um lado, no âmbito do Eixo 3, a participante informou que estão em fase de elaboração o Relatório de Impacto à Proteção de Dados (RIPD), a Avaliação de Impactos em Direitos Humanos (AIDH) e o mapeamento de grupos vulneráveis potencialmente afetados pela aplicação. Por outro lado, em relação ao Eixo 4 (Justiça, não discriminação e vieses algorítmicos), foi reportada a ausência de testes sistemáticos para identificação de vieses em dados e modelos, bem como a existência de mecanismos apenas parcialmente eficazes para prevenir *proxies* discriminatórias, o que representa um ponto de atenção sob a perspectiva de *accountability* algorítmica.

Com isso, a análise conduzida pelas equipes técnica e regulatória do CIAAM/USP registrou que o RIPD será disponibilizado para avaliação no próximo ciclo, permanecendo igualmente pendentes os demais artefatos de conformidade. Adicionalmente, apesar da declaração de que a solução não viola direitos fundamentais, foram programados testes técnicos destinados a verificar a efetividade dos processos de pseudonimização de dados e a robustez das salvaguardas de proteção de dados pessoais implementadas na infraestrutura da aplicação.

Em seguida, no que tange ao fluxo de dados, gestão de dados pessoais e transferência internacional de dados, é necessário observar que um dos principais pilares do sandbox regulatório é conhecer de maneira aprofundada todo o tráfego de dados que entram e saem a partir da infraestrutura da participante com o objetivo de determinar por exemplo:

- se esse fluxo está ou não de acordo com a legislação de proteção de dados especialmente, mas, também, com a legislação setorial quando necessário;
- se os titulares de dados pessoais têm conhecimento sobre como ocorre o tratamento dos seus dados;
- se a eventual transferência internacional de dados ocorre de maneira regular, com fundamento em um dos mecanismos previstos pelo art. 33 da LGPD, e todos os titulares estão cientes e adequadamente informados acerca dessa operação;
- se a infraestrutura possui redundância suficiente para garantir que tanto o tráfego quanto a integridade dos dados não serão comprometidos em caso de incidente;
- se os titulares de dados pessoais podem estabelecer uma comunicação adequada com os agentes de tratamento e com o encarregado, seja para solicitar esclarecimento ou alteração ou exclusão dos seus dados;
- se o fluxo dos dados pode ser considerado auditável e rastreável no que tange à atuação das respectivas autoridades competentes.

¹¹ No Ciclo 3, a avaliação da solução GuardionAI apresentou natureza predominantemente documental e preparatória. Permanecem lacunas nos critérios de avaliação de impactos e de justiça, não discriminação e vieses, além da necessidade de confirmação experimental da efetividade da pseudonimização, da supervisão humana, da explicabilidade e das salvaguardas de proteção de dados.

Nesse sentido, foram agendadas reuniões conjuntas entre as equipes da GuardionAI e do CIAAM/USP além dos representantes da Comissão do Sandbox da ANPD para solucionar dúvidas técnicas e regulatórias relacionadas ao fluxo de dados dentro da plataforma.

Note-se que, a partir da análise do fluxo de dados pessoais, as equipes regulatória e técnica do CIAAM/USP realizaram questionamentos cujas respostas e evidências serão geradas a partir do próximo ciclo por meio dos testes e de novas reuniões para esclarecimento dos resultados alcançados. Assim, destaca-se uma seleção dos principais aspectos que serão aprofundados neste sandbox após o recebimento dos resultados dos testes:

- em qual momento exato do fluxo pode ocorrer a supervisão humana em relação aos resultados da inteligência artificial?
- a inteligência artificial permanece sendo treinada enquanto é utilizada?
- como os elementos de verificação e confiabilidade são incorporados na rotina do projeto?
- como é realizada a rotina de controle da infraestrutura dedicada à inteligência artificial?
- como é determinado o aprendizado do sistema de inteligência artificial?
- como se garante que é feita a revisão e como a governança atua de forma detalhada?
- quem é o responsável pela pseudonimização dos dados antes de enviá-los para o *Large Language Model* (LLM)?
- quais são os tipos de dados estruturados e não estruturados recebidos pela participante e como são tratados?
- como são elaborados pela inteligência artificial os alertas que são enviados à equipe de segurança atendida pela participante?
- em caso de solicitação de exclusão de dados pessoais pelos respectivos titulares, como são feitos o controle e a sua execução em relação ao ambiente de produção e aos dados pessoais salvos em *backups* da aplicação?

Após, em relação aos documentos disponibilizados pela participante para análise pela equipe regulatória do CIAAM/USP, permanecem dúvidas sobre o fluxo de dados internacionais, pois, por exemplo, na Política de Segurança da Informação (*Security Policy*) consta “*Our production environment runs in GCP data centers located in the United States*” e “*Backups are encrypted and stored in geographically appropriate locations within GCP*”. A versão preliminar do RIPD também traz um tópico intitulado “*Cross-border, geographic scope and international transfer*”. Assim, tendo em vista que o sandbox é realizado no Brasil, é necessário, por exemplo, separar as implicações de eventuais vazamentos de dados no exterior e no território nacional.

Ainda, os documentos disponibilizados pela participante denominados Gestão de Acessos (*Access Management*) e Política de Criptografia (*Cryptography Policy*) definem o contato para envio de dúvidas na utilização do sistema, mas não trazem informações de contato para direcionamentos de eventuais incidentes.

Em suma, o atual ciclo possibilitou o primeiro contato com a infraestrutura e a documentação técnica e regulatória da inovação apresentada pela participante. A partir da sua análise, já foram estabelecidas as principais dúvidas regulatórias e a estrutura dos testes que serão conduzidos em conjunto pelas equipes regulatória e técnica do CIAAM/USP. Assim, para o próximo ciclo em que os testes serão realizados, espera-se a geração das evidências e a confirmação das informações angariadas e das dúvidas regulatórias colecionadas conforme este relatório.

Ademais, no Ciclo 3 as fichas A3 e A4 permitiram o progresso da identificação da coordenação institucional e interoperabilidade conforme os seguintes itens:

- legislação aplicável;
- autoridades setoriais;
- instrumentos que podem ser usados pelas autoridades;
- análise de competitividade e Riscos Jurídicos; e
- procedimentos de interoperabilidade e harmonização.

Nessa linha, tendo em vista a informação da participante Metatext a respeito da legislação observada no ciclo pretérito, a equipe regulatória da consultoria do CIAAM/USP consolidou em definitivo a estrutura necessária para avançar no projeto piloto de sandbox de inteligência artificial e proteção de dados ao ampliar de cinco das principais normas indicadas pela empresa para 58 diplomas específicos que deverão ser analisados no atual e nos próximos ciclos com o objetivo de possibilitar a identificação das possíveis lacunas regulatórias para o fortalecimento do ambiente de inovação com segurança técnica e jurídica.

Nesse passo, foram realizadas sucessivas reuniões internas, bem como com a participante e a ANPD com o foco de compreender a profundidade de análise necessária para tornar proveitosa a experiência da análise técnica e regulatória sobre a viabilidade dos requisitos necessários para recepcionar a inovação proporcionada pela participante diante da sua complexidade e abstração em relação às demais selecionadas neste projeto piloto de sandbox.

Dessa forma, foi realizada reunião conjunta entre as equipes do CIAAM/USP, da ANPD e da Metatext, destinada à apresentação dos avanços realizados até então e à definição dos pontos do sistema GuardianAI que seriam submetidos a testes. Ficou definido que seriam realizados dois testes independentes: um relativo à proteção dos dados pessoais (pseudonimização) e outro referente ao envio de alertas à equipe de segurança. A partir desse alinhamento, as equipes regulatória e técnica iniciaram a estruturação da plataforma HarpIA para a testagem da pseudonimização e a elaboração dos *prompts* que serão utilizados no processo agendado para o Ciclo 4.

Assim, em seguida, diante da análise da documentação disponibilizada pela participante, a opção realizada foi organizar a pesquisa sobre a legislação aplicável focada em quatro núcleos fundamentais:

- principal (infraestrutura, aplicação e segurança cibernética); produtos e serviços (seguros, capitalização, consórcios, previdência, câmbio, valores mobiliários etc.);
- pessoas em situação de vulnerabilidade (idosos, crianças, adolescentes, pessoas com deficiência e pessoas superendividadas); e

- startups (pequenas empresas e empreendedorismo focado em inovação).

Tabela: Legislação aplicável à solução GuardionAI para o Núcleo Principal

Perspectiva regulatória		Norma	Autoridade
1	Lei Geral de Proteção de Dados	Lei nº 13.709/2018 (LGPD)	ANPD
2	Regulamento de Comunicação de Incidente de Segurança	Resolução CD/ANPD nº 15/2024 — Regulamento de Comunicação de Incidente de Segurança	ANPD
3	Regulamento de Transferência Internacional de Dados	Resolução CD/ANPD nº 19/2024	ANPD
4	Segurança cibernética das instituições financeiras	Resolução CMN nº 4.893/2021	CMN ¹² e BCB ¹³
5	Altera a Resolução CMN nº 4.893/2021	Resolução CMN nº 5.274/2025	CMN e BCB
6	Política de segurança cibernética das instituições de pagamento e demais entidades abrangidas	Resolução BCB nº 85/2021	BCB
7	Gestão de riscos operacionais e tecnológicos	Resolução CMN nº 4.557/2017	CMN e BCB
8	Prestadores de Serviços de Tecnologia da Informação e acesso à RSFN	Resolução BCB nº 498/2025 ; Resolução BCB nº 547/2026	BCB
9	Estabelece prazos para o PSTI conforme Resolução BCB nº 498/2025	Instrução Normativa BCB nº 664 de 11/9/2025	BCB
10	Sigilo das operações e dos serviços financeiros	Lei Complementar nº 105/2001	BCB

¹² CMN: Conselho Monetário Nacional

¹³ BCB: Banco Central do Brasil

11	Prevenção e compartilhamento de informações sobre fraudes	Resolução Conjunta nº 6/2023	CMN e BCB
12	Estabelece as medidas necessárias à execução do compartilhamento de dados conforme Resolução Conjunta nº 6/2023	Resolução BCB nº 343 de 4/10/2023	CMN e BCB
13	Altera a Resolução BCB nº 343/2023	Resolução BCB nº 569 de 19/5/2026	CMN e BCB
14	Prevenção de fraudes na prestação de serviços de pagamento	Resolução BCB nº 142/2021	BCB
15	Prevenção à lavagem de dinheiro e ao financiamento do terrorismo	Lei nº 9.613/1998	BCB e Coaf ¹⁴
16	Prevenção à lavagem de dinheiro e ao financiamento do terrorismo	Circular BCB nº 3.978/2020	BCB e Coaf
17	Prevenção à lavagem de dinheiro e ao financiamento do terrorismo	Carta Circular BCB nº 4.001/2020	BCB e Coaf
18	Código de Defesa do Consumidor	Lei nº 8.078/1990	Senacon ¹⁵ e Procons ¹⁶
19	Regulamentação do SAC	Decreto nº 11.034/2022	Senacon e Procons
20	Relacionamento entre instituições financeiras, clientes e usuários	Resolução CMN nº 4.949/2021	CMN e BCB
21	Marco Civil da Internet (segurança)	Lei nº 12.965/2014	ANPD
22	Regulamenta a Lei nº 12.965/2014	Decreto nº 8.771/2016	ANPD

¹⁴ Coaf: Conselho de Controle de Atividades Financeiras

¹⁵ Senacon: Secretaria Nacional do Consumidor.

¹⁶ Procons: órgãos estaduais e municipais de proteção e defesa do consumidor.

23	Política Nacional de Segurança de Infraestruturas Críticas	Decreto nº 9.573/2018	GSI/PR ¹⁷
24	Aprova a Estratégia Nacional de Segurança de Infraestruturas Críticas	Decreto nº 10.569/2020	GSI/PR
25	Estratégia nacional de cibersegurança	Decreto nº 12.573/2025	GSI/PR

Fonte: ANPD

Tabela: Legislação aplicável à solução GuardionAI para o Núcleo Produtos e Serviços

Produto ou atividade		Norma	Autoridade	Hipótese de aplicação
1	Investimentos e valores mobiliários	Lei nº 6.385/1976	CVM ¹⁸	Incide quando o <i>chatbot</i> apresentar investimentos, coletar o perfil do investidor, fornecer recomendações, participar da intermediação de valores mobiliários ou apoiar procedimentos relacionados ao mercado de capitais.
2	Investimentos e valores mobiliários	Resolução CVM nº 30/2021	CVM	Incide quando o <i>chatbot</i> participar da verificação da adequação de produtos, serviços e operações ao perfil do cliente.

¹⁷ GSI/PR: Gabinete de Segurança Institucional da Presidência da República.

¹⁸ CVM: Comissão de Valores Mobiliários

3	Investimentos e valores mobiliários	Resolução CVM nº 35/2021	CVM	Incide quando a solução integrar ou apoiar a intermediação de operações realizadas em mercados regulamentados de valores mobiliários, inclusive quanto à segurança cibernética e à gestão de prestadores.
4	Investimentos e valores mobiliários	Resolução CVM nº 50/2021	CVM	Incide quando a solução participar de procedimentos de prevenção à lavagem de dinheiro, ao financiamento do terrorismo e ao financiamento da proliferação de armas de destruição em massa.
5	Investimentos e valores mobiliários	Resolução CVM nº 135/2022	CVM	Incide quando a solução integrar sistemas críticos de mercados organizados, infraestruturas do mercado de capitais ou entidades administradoras de mercado.
6	Seguros	Decreto-Lei nº 73/1966	CNSP ¹⁹ e SUSEP ²⁰	Incide quando o <i>chatbot</i> ofertar, apresentar, explicar, recomendar ou apoiar a contratação de seguros ou quando a instituição contratante estiver sujeita à supervisão da Susep.

¹⁹ CNSP: Conselho Nacional de Seguros Privados

²⁰ SUSEP: Superintendência de Seguros Privados

7	Lei do Contrato de Seguro	Lei nº 15.040/2024	CNSP, SUSEP e Poder Judiciário	Incide sobre as relações contratuais de seguro, inclusive deveres de informação, formação do contrato, cobertura, sinistro e proteção do segurado.
8	Seguros	Resolução CNSP nº 416/2021	CNSP e SUSEP	Incide sobre sistemas de controles internos e gestão de riscos das entidades supervisionadas.
9	Seguros	Resolução CNSP nº 491/2026	CNSP e SUSEP	Incide quando a solução for utilizada por administradoras de operações de proteção patrimonial mutualista ou em atividades relacionadas a esse segmento.
10	Seguros	Resolução CNSP nº 492/2026	CNSP e SUSEP	Incide quando a solução for utilizada por sociedades cooperativas de seguros ou em atividades relacionadas ao cooperativismo de seguros.
11	Seguros	Circular Susep nº 638/2021	SUSEP	Incide sobre os requisitos de segurança cibernética aplicáveis às entidades supervisionadas, incluindo política de segurança, gestão de incidentes, terceirização e computação em nuvem.
12	Seguros	Circular Susep nº 700/2024	SUSEP	Incide sobre procedimentos de autorização, funcionamento e organização das

				entidades supervisionadas, inclusive sociedades participantes de sandbox regulatório.
13	Previdência complementar aberta	Lei Complementar nº 109/2001	CNSP e SUSEP	Incide quando o <i>chatbot</i> ofertar, apresentar, recomendar ou prestar informações sobre planos de previdência complementar aberta.
14	Títulos de capitalização	Decreto-Lei nº 261/1967	CNSP e SUSEP	Incide quando houver oferta, apresentação, atendimento, recomendação ou suporte relacionado a títulos de capitalização.
15	Títulos de capitalização	Lei Complementar nº 213/2025	CNSP e SUSEP	Incide sobre a organização e a atuação de cooperativas de seguros e operações de proteção patrimonial mutualista, conforme o produto oferecido e o enquadramento da entidade.
16	Lei dos Consórcios	Lei nº 11.795/2008	BCB	Incide quando o <i>chatbot</i> ofertar, apresentar, explicar, recomendar ou encaminhar a contratação de cotas de consórcio.
17	Marco Legal do Mercado de Câmbio	Lei nº 14.286/2021	BCB	Incide quando o atendimento abranger operações de câmbio, compra ou venda de moeda estrangeira, remessas e recebimentos internacionais ou movimentação de capitais.

18	Pagamentos e contas de pagamento	Resolução BCB nº 80/2021	BCB	Incide quando a solução for utilizada por instituição de pagamento ou apoiar serviços sujeitos ao regime de autorização e funcionamento dessas instituições.
19	Organização e funcionamento do Sistema Financeiro Nacional	Lei nº 4.595/1964	CMN e BCB	Estabelece a estrutura do Sistema Financeiro Nacional e as competências do CMN e do BCB, servindo de fundamento para a identificação da autoridade competente e do regime aplicável.
20	Tributação (prestação de informações às autoridades)	Lei nº 5.172/1966 — Código Tributário Nacional	Receita Federal, Estados, Distrito Federal e Municípios	Fornece as normas gerais aplicáveis à obrigação tributária, aos tributos e às competências dos entes federativos (artigo 197 do CTN ²¹).
21	Tributação de ISS	Lei Complementar nº 116/2003	Municípios e Distrito Federal	Incide sobre a tributação pelo ISS ²² dos serviços de <i>software</i> , tecnologia, processamento de dados e cibersegurança, conforme o enquadramento da atividade.
22	Tributação (rol)	Lei Complementar nº 214/2025	União, Estados, Distrito Federal e Municípios	Disciplina aspectos da reforma da tributação do consumo, com repercussões sobre a prestação de serviços tecnológicos e digitais (artigo 183)

Fonte: ANPD

²¹ CTN: Código Tributário Nacional

²² ISS: Imposto sobre Serviços

Tabela: Legislação aplicável à solução GuardionAI para o Núcleo de Pessoas em Situação de Vulnerabilidade

<i>Grupo ou risco</i>		<i>Norma</i>	<i>Aplicação</i>
1	Crianças e adolescentes	Lei nº 8.069/1990 (ECA²³)	Estabelece o princípio da proteção integral e determina a consideração prioritária dos direitos e do melhor interesse de crianças e adolescentes.
2	Crianças e adolescentes	Lei nº 13.709/2018 (LGPD - artigo 14)	Disciplina o tratamento de dados pessoais de crianças e adolescentes, que deve ser realizado em seu melhor interesse e com salvaguardas adequadas.
3	Crianças e adolescentes	Lei nº 15.211/2025 (Estatuto Digital da Criança e do Adolescente)	Estabelece direitos, deveres e salvaguardas aplicáveis a produtos e serviços digitais utilizados ou acessíveis a crianças e adolescentes.
4	Crianças e adolescentes	Decreto nº 12.880/2026	Regulamenta medidas de proteção relacionadas ao ambiente digital, incluindo requisitos de segurança, informação, desenho adequado e proteção contra riscos.
5	Pessoas idosas	Lei nº 10.741/2003 (Estatuto da Pessoa Idosa)	Exige atendimento adequado e acessível, preservação da autonomia, informação compreensível e proteção contra fraude, exploração patrimonial e indução indevida.

²³ ECA: Estatuto da Criança e do Adolescente

6	Pessoas com deficiência	Lei nº 13.146/2015 (Lei Brasileira de Inclusão da Pessoa com Deficiência)	Exige acessibilidade, igualdade de oportunidades, desenho inclusivo, não discriminação e adaptação dos canais digitais e conversacionais.
7	Consumidores vulneráveis ou superendividados	Lei nº 14.181/2021 (Lei do Superendividamento)	Reforça os deveres de informação, prevenção e tratamento responsável do consumidor, especialmente na oferta de crédito e em operações que possam comprometer o mínimo existencial.
8	Consumidores vulneráveis ou superendividados	Decreto nº 11.150/2022	Regulamenta aspectos relacionados à prevenção e ao tratamento do superendividamento e à preservação do mínimo existencial do consumidor.

Fonte: ANPD

Tabela: Legislação aplicável à solução GuardionAI para o Núcleo de Startups

	<i>Tema</i>	<i>Norma</i>	<i>Aplicação</i>
1	Enquadramento e estímulo à inovação	Lei Complementar nº 182/2021 (Marco Legal das Startups e do Empreendedorismo Inovador)	Após análise do caso concreto da participante, é possível fundamentar tratamento regulatório experimental e proporcional. Porém, não afasta a LGPD, as normas de segurança cibernética, a legislação financeira, as regras de proteção do consumidor ou os deveres contratuais.

2	Agentes de tratamento de pequeno porte	Resolução CD/ANPD nº 2/2022	A aplicação das flexibilizações não é automática. Por exemplo, a <i>startup</i> poderá ser excluída do regime diferenciado quando realizar tratamento de alto risco, em grande escala ou envolvendo monitoramento sistemático, dados financeiros, biometria ou tecnologias emergentes.
3	Microempresas e empresas de pequeno porte	Lei Complementar nº 123/2006 (Estatuto Nacional da Microempresa e da Empresa de Pequeno Porte)	A incidência depende do porte, da receita e da atividade econômica da participante. O enquadramento não reduz as exigências prudenciais de segurança e continuidade impostas pela instituição financeira contratante.

Fonte: ANPD

Em seguida, verifica-se que a análise das medidas de transparência e conformidade regulatória concentra-se na aderência da solução ao artigo 20 da Lei Geral de Proteção de Dados (LGPD), considerando que a aplicação realiza decisões baseadas exclusivamente em tratamento automatizado de dados pessoais. O sistema tem como finalidade identificar interações potencialmente fraudulentas direcionadas às instituições financeiras, gerando alertas e bloqueios automáticos quando detecta comportamentos incompatíveis com perfis legítimos de clientes.

A avaliação preliminar constatou que a participante não mantém relação direta com os clientes finais das instituições financeiras. Quando uma operação é bloqueada em decorrência de um alerta gerado pela solução, o titular dos dados recebe apenas a negativa da operação, sem acesso imediato às justificativas técnicas que fundamentaram a decisão automatizada. Dessa forma, os esclarecimentos sobre a recusa dependem de solicitação posterior à instituição financeira responsável pelo atendimento ao usuário. Adicionalmente, cumpre destacar que a ausência de relação direta entre a Metatext e os clientes finais das instituições financeiras não é capaz de afastar a necessidade de

mecanismos de explicabilidade, contestação, revisão e atendimento aos direitos dos titulares, portanto, este é o motivo de serem tratadas estas dimensões no próximo ciclo.

Em observância ao §1º do artigo 20 da LGPD, o próximo ciclo do Sandbox concentrará sua análise em três dimensões de explicabilidade. A primeira refere-se à qualidade e suficiência das informações disponibilizadas aos titulares impactados pelas decisões automatizadas. A segunda envolve a capacidade da participante de fornecer justificativas técnicas e regulatórias adequadas em eventuais solicitações da autoridade supervisora ou das instituições financeiras contratantes. A terceira diz respeito à utilidade, completude e robustez das informações transmitidas à equipe de segurança cibernética responsável pela gestão dos alertas gerados pela solução.

Nesse passo, recorda-se que a metodologia de avaliação prevê a realização de testes operacionais e análises documentais fundamentadas exclusivamente no arcabouço normativo vigente à data de início dos testes. A documentação já fornecida pela participante servirá como base para verificar não apenas a conformidade formal do modelo, mas também a efetividade prática dos mecanismos de transparência, rastreabilidade, explicabilidade e prestação de contas relacionados ao processamento automatizado de dados pessoais. Por fim, a observância dos requisitos pertinentes ao plano de sandbox atualizado da participante para cada uma das 58 normas indicadas acima será avaliada sob a perspectiva regulatória diante da documentação recebida e dos testes realizados no próximo ciclo.

5. Conclusões gerais do Ciclo 3

O Ciclo 3 possibilitou a continuidade da metodologia de supervisão e testagem das soluções das participantes Synapse, Prevvine e Metatext, aprofundando a análise documental, a verificação de controles técnicos e a identificação de condicionantes para os ciclos subsequentes. Dessa forma, as atividades compreenderam a atualização das informações sobre as soluções, o exame de sua arquitetura e governança, a execução ou preparação de testes e a consolidação dos aprendizados técnicos e regulatórios. Da mesma forma que o Ciclo 2, a organização nas Fases A, B e C permaneceu adequada para relacionar a caracterização dos projetos, as evidências experimentais e os encaminhamentos decorrentes da avaliação.

Preliminarmente, é relevante destacar que a experiência confirmou a importância de combinar instrumentos comuns com procedimentos ajustados às características de cada participante. Assim, os resultados devem ser interpretados considerando o escopo efetivamente examinado, as condições de execução dos testes e as evidências disponíveis. Nesse sentido, a continuidade metodológica possibilitou acompanhar o aprimoramento das soluções, mas essa percepção não implica equivalência entre seus estágios de desenvolvimento, a abrangência dos testes realizados ou o grau de comprovação das salvaguardas apresentadas.

Dessa forma, os resultados do Ciclo 3 devem ser interpretados de acordo com o escopo de testagem efetivamente realizado por cada participante, o nível de maturidade tecnológica das soluções, a documentação apresentada, as evidências materialmente verificadas e os limites metodológicos de cada avaliação. Essa distinção é relevante para evitar comparações diretas ou conclusões uniformes entre projetos que se encontram em estágios distintos de desenvolvimento, comprovação técnica e amadurecimento regulatório.

Em primeiro lugar, sobre a participante Synapse, o ciclo permitiu aprofundar a avaliação da solução Trajetto, com avanços na documentação de governança e na verificação de mecanismos de proteção de dados e segurança. Destacaram-se a implementação de protocolo de segurança na camada de transporte e os resultados dos testes sobre o tráfego auditado e o processamento de imagens, nos quais não foi observada identificação individual nas condições examinadas. Ademais, a pontuação segundo o *framework* geral passou de 45 para 68 pontos, e o parecer técnico registrou risco inerente alto e risco residual médio, com avaliação suficiente com condicionantes. Contudo, os resultados não abrangem integralmente as conexões externas, especialmente os fluxos associados às API de terceiros. Ainda, permaneceram pontos de atenção relativos à possibilidade de identificação indireta, à acessibilidade, à transparência, à comprovação de controles de proteção da aplicação e à definição das responsabilidades dos parceiros envolvidos no tratamento de dados.

Em segundo lugar, no que se refere à participante Prevvine, o Ciclo 3 comprovou avanços operacionais na plataforma STAiDOC, particularmente o restabelecimento da

integração com a plataforma HarpIA, que viabilizou a aplicação das 30 vinhetas clínicas, e a correção da resposta disponibilizada pela API, com remoção de artefatos técnicos no servidor e inclusão de *links* diretos para referências no PubMed. Nesse passo, a pontuação no *framework* geral passou de 60 para 63 pontos, enquanto o parecer técnico classificou o risco inerente como crítico e o risco residual como médio, mantendo avaliação favorável com condicionantes. Em seguida, a Ficha C1 classificou o êxito do ciclo como parcial. Dessa forma, a retomada dos testes constitui avanço relevante, mas a aplicação das vinhetas, isoladamente, não demonstra o atendimento das condicionantes relatadas no parecer técnico. Assim, permaneceram necessárias melhorias na explicabilidade, na supressão de dados pessoais, nas rotinas criptográficas, na proteção da aplicação, na continuidade operacional e na apresentação das políticas de privacidade, retenção e descarte.

Em terceiro lugar, no que se refere à solução GuardionAI, o Ciclo 3 concentrou-se no aprofundamento da compreensão da arquitetura, dos fluxos de dados e da documentação técnica e regulatória, bem como na delimitação dos testes previstos para o Ciclo 4. Diante tanto da complexidade técnica e regulatória da inovação quanto da sua abstração em relação aos titulares de dados pessoais, a disponibilização da documentação confidencial da empresa e a ampliação do mapeamento normativo contribuíram para estruturar a avaliação, sem, por enquanto, encerrar todas as dúvidas sobre transferências internacionais, supervisão humana, anonimização, pseudonimização, governança e resposta a incidentes. Foram definidos dois testes independentes, relativos à proteção dos dados pessoais e ao envio de alertas à equipe de segurança. Assim, as informações documentais e as declarações da participante permanecem sujeitas à confirmação experimental, especialmente quanto à anonimização, pseudonimização, transferência internacional, supervisão humana, explicabilidade, governança, resposta a incidentes e suficiência das informações destinadas aos titulares e às instituições contratantes.

De modo geral, o Ciclo 3 reforçou a necessidade de distinguir controles declarados, documentados, implementados e efetivamente verificados. Nessa linha, verificou-se que a existência de políticas, a evolução da pontuação em instrumentos de avaliação e a correção de impedimentos operacionais constituem evidências relevantes, mas estas devem ser examinadas em conjunto com os resultados dos testes e os limites identificados (tanto operacionais quanto regulatórios). Do mesmo modo, a maturidade tecnológica, a avaliação ética e a classificação de riscos são dimensões complementares que contribuem para a melhor experiência do ambiente experimental. Finalmente, a continuidade da experimentação deve ser orientada pelo acompanhamento das condicionantes identificadas e pela demonstração progressiva da efetividade das medidas de mitigação.

Em razão do exposto, para o Ciclo 04, recomenda-se priorizar:

- A verificação das condicionantes e recomendações do Ciclo 3, com registro dos responsáveis, prazos, evidências exigidas e critérios de conclusão, seguida da atualização dos riscos residuais;

- No caso da Synapse, a apresentação e a verificação de evidências materiais de transparência na interface, acessibilidade, completude do inventário de API e terceiros, mitigação da reidentificação indireta e efetividade dos controles de segurança da aplicação. Esses elementos deverão permitir confrontar os instrumentos de governança apresentados com sua implementação operacional, reduzindo a distância entre controles documentados e efetivamente verificados;
- Sobre a empresa Previne, o exame dos resultados das vinhetas clínicas e o cumprimento das condicionantes sobre políticas documentais, explicabilidade, proteção criptográfica, segurança da aplicação, continuidade diante da (possível) indisponibilidade do PubMed, prazos de avaliação e aperfeiçoamento da supressão de dados pessoais;
- Na GuardionAI, a execução dos testes de proteção de dados e de geração de alertas, com avaliação da supervisão humana, da rastreabilidade e da explicabilidade das decisões, além do esclarecimento dos fluxos internacionais e das responsabilidades dos agentes envolvidos.

Em geral, também se verificou a necessidade de:

- Complementação ou aprofundamento do inventário de dados, integrações e terceiros, com identificação das finalidades, dos papéis dos agentes de tratamento e dos procedimentos de retenção, exclusão, atendimento aos titulares e resposta a incidentes;
- Registro e versionamento dos modelos, parâmetros, interfaces, API e ambientes examinados, assegurando que alterações relevantes sejam comunicadas e avaliadas quanto aos seus efeitos sobre a continuidade e a comparabilidade dos testes;
- Atualização das matrizes de maturidade tecnológica e dos Elementos Tecnológicos Críticos (ETC), mediante evidências, e a revisão dos Planos de Descontinuidade (PdD) quando mudanças na solução ou nos riscos justificarem ajustes;
- Aprofundamento da análise setorial e da coordenação institucional, acompanhado de reuniões de devolutiva entre as participantes, o CIAAM/USP e a ANPD para validação dos achados e definição dos encaminhamentos.

Destarte, o Ciclo 04 deverá aprofundar a comparação entre as medidas recomendadas, as alterações implementadas e os resultados efetivamente observados. Essa avaliação deverá permitir identificar quais controles reduziram os riscos, quais permanecem insuficientes e quais dependem de testes adicionais. Para tanto, será necessário preservar a rastreabilidade das evidências e explicitar as limitações que condicionam cada conclusão.

Finalmente, conclui-se que o principal resultado do Ciclo 3 foi o avanço qualitativo na produção, organização e análise das evidências regulatórias. Em contraste com os ciclos anteriores, dedicados predominantemente à estruturação metodológica, ao alinhamento inicial das participantes e à identificação de lacunas documentais e técnicas, o Ciclo 3 permitiu aprofundar a verificação dos controles apresentados e diferenciar, com maior precisão, controles declarados, documentados, implementados e efetivamente verificados.

Esse amadurecimento fortaleceu a interação entre avaliação técnica e análise regulatória, permitindo consolidar orientações para a continuidade da experimentação e definir encaminhamentos específicos para cada uma das três participantes. A progressão ao Ciclo 4 deverá, portanto, concentrar-se na verificação da efetividade desses controles, na atualização dos riscos residuais e na ampliação da base de aprendizado da ANPD, orientando o desenvolvimento das soluções em equilíbrio com a proteção dos titulares e a supervisão dos riscos.

