



sandbox
regulatório

2º Relatório Parcial de Monitoramento

Fase de Teste – Ciclo 2

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	7
3. Definição dos temas e metodologia para as sessões remotas e presenciais	9
4. Apoio, orientação e supervisão das atividades de testagem	11
4.1 Synapse	11
4.2 Prevvine	21
4.3 Metatext	31
5. Conclusões gerais do Ciclo 2.....	36

1. Introdução

O projeto-piloto de Sandbox Regulatório em Inteligência Artificial e Proteção de Dados da Agência Nacional de Proteção de Dados (ANPD) constitui um ambiente regulatório experimental, controlado e supervisionado, voltado à testagem de projetos inovadores que utilizam sistemas de inteligência artificial (IA) e envolvem tratamento de dados pessoais. A iniciativa busca promover a inovação responsável, a transparência algorítmica e a conformidade com a Lei Geral de Proteção de Dados Pessoais (LGPD), permitindo que soluções baseadas em IA sejam acompanhadas em contexto estruturado de orientação, monitoramento e avaliação.

No âmbito do projeto-piloto, o Sandbox funciona como espaço de aprendizagem regulatória e de produção de evidências. Nesse contexto, as empresas Metatext¹, Synapse² AI e Prevvine³ Tecnologia foram selecionadas como participantes. Tais organizações desenvolvem e testam suas soluções sob acompanhamento técnico e regulatório. A testagem permite identificar desafios tecnológicos, jurídicos e operacionais associados aos sistemas avaliados, bem como mapear riscos, boas práticas, necessidades de aprimoramento e eventuais lacunas de governança, segurança, transparência e proteção de dados pessoais.

As atividades de testagem estão estruturadas em seis ciclos, cada um composto por três fases: Fase A (pré-teste), Fase B (teste) e Fase C (pós-teste). Na Fase A, os participantes preenchem fichas a respeito de suas aplicações, metodologias e estágio de desenvolvimento de módulos ou componentes. Na Fase B, testes específicos para cada participante são executados, considerando as informações coletadas na Fase A e os artefatos de software fornecidos pelos participantes. Ao final de cada ciclo, na fase C, pareceres são apresentados individualmente para cada participante. Tais pareceres orientam o próximo ciclo de testagem, visando melhorias técnicas, éticas e regulatórias.

Este documento apresenta as atividades realizadas durante o segundo ciclo de testagem dos projetos participantes do projeto-piloto, fazendo conexões com as condições dos projetos ao final do Ciclo 1 e encaminhamentos de atividades e pendências a serem tratadas no Ciclo 3. Nesse sentido, foram feitas avaliações de proteção de dados específicas com participação da Comissão de Sandbox da ANPD, das empresas participantes e da equipe do CIAAM/USP, considerando um Framework geral baseado em recomendações da UNESCO, cibersegurança e regulações setoriais aplicáveis a cada projeto.

As avaliações levaram em conta os planos de sandbox, os planos de descontinuidade, e a evolução dos níveis de maturidade tecnológica dos projetos a partir de elementos tecnológicos críticos como infraestrutura, algoritmos, cibersegurança e modelos de IA. Assim, foram feitos testes técnicos e regulatórios a partir de diagnósticos dos representantes das empresas sobre os próprios projetos. Por fim, os resultados dos

¹ 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.

² 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.

³ A Prevvine Tecnologia é a desenvolvedora da plataforma STAI DOC, um ecossistema de inteligência artificial médica com “*Large Language Models*” personalizados para o contexto brasileiro.

testes são analisados sob a perspectiva dos participantes e da ANPD, com relato dos aprendizados adquiridos ao longo do Ciclo 2.

Este relatório tem, portanto, a finalidade de registrar e analisar as atividades realizadas no período, consolidando os principais dados obtidos no segundo ciclo de acompanhamento e indicando oportunidades de melhoria para a continuidade da fase de testagem. Ao sistematizar os aprendizados iniciais, o documento contribui para o aprimoramento da metodologia de monitoramento e avaliação do Sandbox, bem como para o fortalecimento da atuação regulatória da ANPD diante dos desafios associados ao desenvolvimento e ao uso responsável de sistemas de inteligência artificial.

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

O planejamento das sessões de orientação seguiu o calendário previamente estabelecido para a Fase B dos ciclos de testagem, conforme o item 3 deste relatório. Foram previstas reuniões periódicas e individualizadas com as empresas participantes, realizadas em momentos próximos ao encerramento da fase de experimentação de cada ciclo, com a participação da equipe técnica e regulatória responsável e da Comissão de Sandbox da ANPD.

No Ciclo 2, essas reuniões foram agendadas conforme o cronograma específico de cada participante. Para a Synapse e a Previne, as sessões correspondentes foram previstas para 22 de junho de 2026; para a Metatext, cujo cronograma de ingresso e execução foi organizado em datas distintas, a reunião foi prevista para 31 de julho de 2026. A diferenciação das datas permitiu respeitar o estágio efetivo de desenvolvimento dos testes de cada empresa, sem comprometer a aplicação uniforme da metodologia.

As sessões foram estruturadas para cumprir as seguintes finalidades:

- esclarecer dúvidas relacionadas à execução dos testes e ao preenchimento dos instrumentos metodológicos;
- informar às empresas quais dados, documentos e evidências foram coletados ou ainda precisavam ser apresentados;
- discutir os achados preliminares decorrentes da aplicação da metodologia;
- verificar a correspondência entre as informações declaradas pelas empresas, as evidências fornecidas e os resultados observados durante os testes;
- identificar lacunas documentais, inconsistências ou necessidades de complementação;
- coletar as manifestações das empresas e os comentários da Comissão de Sandbox da ANPD;
- registrar pendências e possíveis ajustes a serem considerados na Fase C ou no ciclo subsequente.

A realização dessas reuniões em momento próximo ao encerramento da Fase B é especialmente relevante, pois permite submeter a análise técnica e regulatória a uma etapa de validação antes da consolidação dos pareceres do ciclo e do início da Fase C de pós-testagem. Nesse momento, a equipe já dispõe de elementos suficientes para apresentar observações concretas sobre os testes, mas ainda há oportunidade de esclarecer informações, corrigir interpretações e solicitar evidências complementares.

As reuniões também reduzem o risco de que os pareceres sejam elaborados exclusivamente com base em documentos ou autodeclarações considerados isoladamente. O diálogo direto permite contextualizar as evidências, compreender as

limitações técnicas, verificar eventuais alterações realizadas na solução e distinguir entre medidas efetivamente implementadas, medidas em andamento e providências apenas planejadas.

Sob a perspectiva regulatória, as sessões constituem um espaço de supervisão responsiva e de aprendizagem compartilhada. A participação da ANPD possibilita que dúvidas e achados sejam examinados à luz dos objetivos do Sandbox, especialmente quanto à proteção de dados pessoais, à transparência, à explicabilidade, à governança e à gestão de riscos. Paralelamente, as empresas podem compreender com maior precisão as expectativas regulatórias e apresentar as particularidades técnicas de suas soluções.

Por fim, essas reuniões asseguram a rastreabilidade do processo de testagem. Os comentários, esclarecimentos e pendências nelas registrados permitem estabelecer uma relação verificável entre: (i) as informações apresentadas pela empresa; (ii) as evidências analisadas; (iii) os resultados obtidos nos testes; (iv) os apontamentos da equipe técnica e da ANPD; e (v) as recomendações formuladas no parecer. Dessa maneira, as sessões funcionam como mecanismo de controle de qualidade metodológica e como etapa de transição entre a experimentação realizada na Fase B e a avaliação consolidada da Fase C.

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

A proposta inicial de calendário para os Ciclos 2 a 6 foi estruturada com base na metodologia de supervisão e acompanhamento do Projeto Piloto de Sandbox Regulatório de Inteligência Artificial e Proteção de Dados, elaborada pela equipe do CIAAM/USP e validada em conjunto com a Comissão de Sandbox da ANPD.

A organização das sessões acompanhou o desenvolvimento progressivo dos ciclos de testagem e a divisão metodológica em três fases: Fase A – Pré-teste; Fase B – Teste; e Fase C – Pós-teste. Embora existisse uma estrutura metodológica comum, os temas não foram definidos de forma abstrata ou uniforme para todas as empresas. Assim, o conteúdo específico de cada sessão foi delimitado de acordo com:

- o escopo e os objetivos definidos no Plano de Sandbox apresentado por cada empresa;
- as características técnicas e regulatórias de cada solução;
- o setor de aplicação e os tipos de dados pessoais tratados;
- o nível de maturidade tecnológica da solução;
- os riscos previamente identificados e aqueles verificados durante os testes;
- os resultados e as evidências produzidos nos ciclos anteriores; e
- as necessidades de orientação, complementação documental ou aprofundamento indicadas pela equipe técnica e pela ANPD.

Essa estrutura permitiu conciliar a padronização metodológica com a individualização dos testes. Nesse sentido, a padronização assegurou que todas as empresas fossem acompanhadas segundo dimensões comuns de análise, enquanto a individualização permitiu que as sessões considerassem os usos esperados, os benefícios declarados, os testes de extensão e os riscos específicos de cada solução.

Na Fase A, as sessões foram direcionadas à preparação dos testes, à atualização das fichas técnicas, à avaliação do nível de maturidade tecnológica, à identificação de mudanças nas premissas da solução e à análise das medidas de transparência e de governança. Na Fase B, os encontros concentraram-se na execução e no acompanhamento dos testes, na aplicação dos *frameworks* gerais, tecnológicos e setoriais pertinentes, bem como na análise dos usos esperados, dos benefícios, dos testes de extensão, dos riscos e das medidas de mitigação. Na Fase C, foram previstos encontros voltados à avaliação dos resultados, à identificação dos êxitos e das limitações do ciclo, à formulação de recomendações e à atualização do Plano de Descontinuidade.

A metodologia das sessões combinou exposição orientativa, análise documental, discussão técnica e regulatória e registro de encaminhamentos. As reuniões permitiram apresentar os objetivos de cada etapa, esclarecer dúvidas, examinar os documentos e as evidências produzidos, discutir resultados preliminares e estabelecer providências para a conclusão do ciclo ou para a etapa subsequente. Sempre que necessário, foram utilizadas perguntas semiestruturadas, sem prejuízo da formulação de questões adicionais

decorrentes dos achados observados durante os testes. É relevante destacar que foram criados *chats* específicos no Teams, integrados pelas empresas, pela equipe do CIAAM/USP e pela equipe da ANPD para facilitar o diálogo e esclarecimento de dúvidas pontuais para a condução das testagens.

As sessões remotas foram priorizadas para o acompanhamento periódico e individualizado das empresas, por facilitarem a participação das equipes técnicas, da coordenação do projeto e da ANPD, além de permitirem maior regularidade e agilidade na resolução de dúvidas. Sendo assim, não houve necessidade de realizar as atividades presenciais que estavam previstas para situações que exigissem interação ampliada, participação de *stakeholders*, discussão interinstitucional ou realização de *workshops* com as empresas, a ANPD e outros atores relevantes. Desse modo, o Ciclo 2 de testagem ocorreu sem qualquer prejuízo decorrente da inexistência de interação presencial.

O calendário foi estruturado de forma dinâmica e cumulativa, em que cada sessão deveria considerar não apenas os dados produzidos no ciclo correspondente, mas também as informações consolidadas nos ciclos anteriores. Isso permitiu acompanhar a evolução das soluções e identificar alterações em sua arquitetura, finalidade, fluxos de dados, nível de maturidade, riscos e medidas de mitigação. Assim, o planejamento inicial funcionou como referência organizacional, admitindo ajustes justificados conforme o andamento efetivo dos testes e as necessidades identificadas em cada caso.

Essa forma de organização assegurou a coerência entre o Plano de Sandbox apresentado pelas empresas e aprovado pela ANPD e pelo CIAAM/USP, a metodologia validada junto à ANPD e as atividades efetivamente desenvolvidas. Também contribuiu para a comparabilidade entre os ciclos, a rastreabilidade das decisões metodológicas e a produção de evidências suficientes para fundamentar os pareceres e relatórios do Sandbox.

4. Apoio, orientação e supervisão das atividades de testagem

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

4.1 Synapse

A empresa atua no setor de mobilidade urbana e busca apresentar uma solução que recomenda aos usuários do metrô da cidade de São Paulo o melhor trajeto a ser percorrido, considerando as condições de fluxo e de trânsito de pessoas em determinado horário.

O relatório do CIAAM/USP referente ao Ciclo 2 foi entregue à ANPD em 1º de junho. Trata-se de um compilado de documentos referentes às fases A e B do Ciclo 2 de testagem. A seguir, serão apresentados os principais dados, informações e conhecimentos 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 fases A3, A4, B1 e B2;
- Nível de Maturidade Tecnológica (*Technology Readiness Level* — TRL) e Elementos Tecnológicos Críticos (ETC) identificados: as empresas apresentaram suas autoavaliações nas fichas A1 e A2, e a equipe do CIAAM/USP registrou, depois, a análise dessas autoavaliações nas fichas B1 e B2.
- 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 fase 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 fase B1 de testagem.

Como o Ciclo 2 foi completado em suas 3 fases (A, B e C), cumpre relatar os principais aspectos identificados em cada uma delas.

4.1.1 Fase A

Primeiramente, analisou-se a Ficha Técnica de Aplicação (A1 e A2) sob responsabilidade da empresa. Quanto à arquitetura do modelo e ao fluxo de dados, não houve mudança do Ciclo 1 para o Ciclo 2, e o sistema cria recomendações de rotas internas nas estações de metrô. Da mesma maneira, os protocolos de segurança (art. 46 da Lei 13.709/2018) permaneceram semelhantes entre os Ciclos 1 e 2. Em relação à Ficha A2 que trata da avaliação de maturidade tecnológica e do mapeamento dos ETC, é possível verificar que a validação técnica dos algoritmos centrais permanece a mesma no Ciclo 2, ou seja, TRL 4. Já no que concerne à pseudonimização e à segurança dos bancos de dados, é possível notar que a empresa indicou TRL 6; em outros termos, houve evolução entre o Ciclo 1 (indicado como TRL 5) e o Ciclo 2.

No que tange às fichas A3 e A4, em termos de análise regulatória e documental, não houve mudança relevante entre os Ciclos 1 e 2, pois não foram disponibilizados novos documentos. Para o Ciclo 3, foi solicitado um conjunto de referências técnicas e normativas aplicáveis e de documentos internos relevantes.

Os diplomas avaliados estão relacionados no Quadro 1.

Quadro 1: Referências técnicas e normativas aplicáveis analisadas para a empresa Synapse

PROTEÇÃO DE DADOS E DIREITOS DIGITAIS	
Lei 15.211/2025	Estatuto Digital da Criança e do Adolescente (ECA Digital)
Lei 13.709/2018	Lei Geral de Proteção de Dados Pessoais (LGPD)
Lei 12.965/2014	Marco Civil da Internet
Lei 8.078/1990	Código de Defesa do Consumidor (CDC)
Decreto 12.880/2026	Regulamento do ECA Digital
Resolução CD/ANPD 2/2022	Regulamento para agentes de tratamento de pequeno porte
Projeto de Lei 2338/2023	Marco Legal da Inteligência Artificial
GOVERNO DIGITAL E USUÁRIOS DE SERVIÇOS PÚBLICOS	
Lei 14.129/2021	Lei do Governo Digital
Lei 13.460/2017	Lei de Proteção e Defesa dos usuários de Serviços Públicos

Lei 12.527/2011	Lei de Acesso à Informação (LAI)
ACESSIBILIDADE, INCLUSÃO E IGUALDADE	
Lei 13.146/2015	Lei Brasileira de Inclusão da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência)
Lei 12.288/2010	Estatuto da Igualdade Racial
Lei 10.741/2003	Estatuto da Pessoa Idosa
Lei 8.069/1990	Estatuto da Criança e do Adolescente (ECA)
ABNT NBR 17.225/2025	Acessibilidade em conteúdo e aplicações web — Requisitos
ABNT NBR 17.060/2022	Acessibilidade em aplicativos de dispositivos móveis — Requisitos

Fonte: ANPD

Por sua vez, os documentos internos avaliados e solicitados à empresa são apresentados no Quadro 2.

Quadro 2: Documentos internos avaliados para a empresa Synapse

Nº	<i>Documento Interno Avaliado</i>
01	Relatório de Impacto à Proteção de Dados Pessoais (RIPD)
02	Avaliação de Impacto em Direitos Humanos (AIDH)
03	Mapa de Fluxo Ponta a Ponta
04	Matriz de Bases Legais por Operação
05	Inventário de Interfaces de Programação de Aplicações (<i>Application Programming Interface</i> — API) e Terceiros

06	Documento de Arquitetura de Pseudonimização e Criptografia
07	Validação de Localização de Processamento
08	Acordos de Cooperação Técnica (ACT) e Instrumentos com Parceiros
09	Critérios Formais de Go/No Go por Ciclo
10	Registro de Operações de Tratamento
11	Política de Governança de IA e Dados
12	Nomeação Formal e Atuação do Encarregado de Dados
13	Plano de Respostas a Incidentes
14	Relatório de Testes de Robustez e Estresse
15	Relatório de Viés/ <i>Fairness</i>
16	Evidência de Transparência ao Titular (Art. 9º e Art. 20 da LGPD) — Dever de Informação
17	Protocolo de Explicabilidade em Camadas

Fonte: ANPD

O conteúdo dos acordos com órgãos públicos sobre a transferência de dados, especialmente no que tange à incidentes de proteção de dados pessoais e às respectivas responsabilidades deverão ser apresentados nos ciclos subsequentes.

Quanto aos resultados dos testes de operação de algoritmo e de segurança no tratamento de dados pessoais, verificaram-se avanços relevantes. Os protocolos empregados para a criptografia e a proteção de dados pessoais mostraram-se bem-sucedidos. No entanto, o teste adversarial não foi finalizado no Ciclo 2, pois os dados utilizados pelo algoritmo não trafegaram pelo túnel de teste ligando a infraestrutura da USP à *Virtual Private Network* (VPN) da Synapse. O teste será refeito no Ciclo 3. Para o Ciclo 2, o resultado referente à segurança da informação e dos dados empregados permaneceu inconclusivo.

4.1.2 Fase B

No que tange à aplicação do *Framework* Geral da UNESCO, empregou-se a seguinte metodologia.

O *framework* 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)

O *framework* é um instrumento setorialmente neutro, aplicável a sistemas de naturezas diversas, cabendo ao avaliador ponderar as particularidades de setor, finalidade e criticidade em cada caso concreto. A avaliação é processual e iterativa: no Ciclo 1, a própria organização realizou a 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: (i) Pesquisa básica (TRL 1 a 3); (ii) Desenvolvimento (TRL 4 a 6); (iii) Demonstração (TRL 7 a 8); e (iv) Produção e operação (TRL 9). Essas faixas 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 em um *framework* mensurável, estabelecendo-se a supracitada escala de pontuação. No Ciclo 2, a pontuação foi igual à do Ciclo 1, ou seja, de 45 pontos. Destacam-se os seguintes eixos que não pontuaram: (i) Avaliação de impactos (Eixo 3); (ii) Justiça, não-discriminação e vieses algorítmicos (Eixo 4); e (iii) Segurança técnica e robustez (Eixo 6). A ausência de pontuação no Eixo 6 decorre da não conclusão dos testes no túnel.

O conjunto de dados que acompanha o algoritmo fornecido foi gerado pela empresa e os testes com dados sintéticos foram bem-sucedidos. O teste de túnel não pôde ser concluído, sendo que será realizado no Ciclo 3 novamente.

Quanto ao teste de IA adversarial em ambiente controlado, não foi concluído no Ciclo 2, mas pretende-se realizá-lo no Ciclo 3, mediante o emprego de uma IA local que analise os dados e identifique potenciais trechos capazes de permitir a identificação de indivíduos.

No que tange ao teste de frente algorítmica, o algoritmo para detecção de densidade de pessoas em um ambiente, parte da solução da Synapse, foi testado com dados fornecidos pela empresa. Os resultados foram bem-sucedidos, e o teste foi realizado em máquina própria (USP), isolada da rede.

Por fim, foi elaborado um parecer técnico consolidado atinente à governança, segurança da informação e privacidade em IA, apresentado previamente à Comissão de Sandbox da ANPD, que apontou alguns itens para esclarecimento. A versão final foi disponibilizada pela empresa após o alinhamento entre as equipes do CIAAM/USP e da ANPD. No que se refere a esse parecer técnico, pode-se mencionar que ele analisou os riscos associados à solução Trajetto e avaliou a suficiência das medidas de mitigação apresentadas pela Synapse no Ciclo 2. Foram identificados riscos relacionados à integração dos componentes do sistema, à confidencialidade e autenticidade das informações, à possibilidade de reidentificação de titulares por meio da análise de padrões de deslocamento e ao compartilhamento de coordenadas e identificadores com provedores externos. O risco inerente foi classificado como alto e o risco residual como médio, em razão da existência de controles técnicos relevantes, como a arquitetura em duas camadas, a separação lógica das bases, a pseudonimização, o uso de criptografia por meio de um serviço de gerenciamento de chaves terceirizado e o descarte das imagens após a extração das coordenadas.

Apesar da robustez da arquitetura, o parecer concluiu que parte das mitigações ainda não foi suficientemente comprovada por evidências documentais ou testes técnicos. O RIPD e a Avaliação de Impacto em Direitos Humanos foram analisados pela equipe regulatória do CIAA/USP, porém a análise técnica de tais documentos ainda não foi realizada neste ciclo. Ainda não foi viável realizar os testes do plano de resposta a incidentes, o inventário e a auditoria técnica das integrações externas, a comprovação de controles contra falsificação de informações, o endurecimento do acesso à API de integração e a atuação formal do encarregado de proteção de dados. Por essa razão, apesar de a pontuação ser de 45 pontos, enquanto o mínimo para continuidade seria de 50 pontos, o prosseguimento dos testes recebeu parecer favorável com condicionantes. Nesse sentido, recomendou-se que as medidas pendentes sejam implementadas e evidenciadas nos ciclos seguintes a fim de reduzir os riscos residuais e evitar a suspensão das atividades.

4.1.3 Fase C

A elaboração do relatório de êxitos integra a Fase C — Avaliação e Ajuste da metodologia desenvolvida pelo CIAAM/USP e validada pela Comissão de Sandbox da ANPD. Essa etapa possui função complementar aos pareceres técnicos: 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. Conforme detalhamento dos quadros 3 e 4, em cada campo, deveriam ser indicados o resultado observado e a evidência correspondente, permitindo relacionar as conclusões às informações efetivamente coletadas durante o ciclo.

Quadro 3: Campos de avaliação sob a perspectiva da ANPD

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

Fonte: ANPD

Quadro 4: Campos de avaliação sob a perspectiva da empresa

<i>Campo</i>	<i>Objeto da Avaliação</i>
<i>Evoluções técnicas</i>	Aperfeiçoamentos realizados na arquitetura, nos modelos, nos controles de segurança, nos fluxos de dados e nos mecanismos de rastreabilidade.
<i>Aprendizados regulatórios</i>	Compreensão das obrigações jurídicas e incorporação dos requisitos regulatórios ao desenvolvimento da solução.
<i>Transparência e explicabilidade</i>	Medidas adotadas para informar os titulares, explicar o funcionamento da solução e tornar as recomendações ou resultados compreensíveis e auditáveis.

<i>Direitos dos titulares</i>	Mecanismos destinados a viabilizar acesso, correção, eliminação, anonimização, oposição e revisão, quando aplicáveis.
<i>Autorizações e análises pendentes</i>	Documentos, validações, testes ou providências ainda necessários para a continuidade ou ampliação da solução.
<i>Outros aprendizados</i>	Efeitos organizacionais, operacionais ou estratégicos produzidos pela participação no Sandbox.

Fonte: ANPD

O relatório também contemplou uma conclusão conjunta, na qual deveriam ser registrados a classificação do êxito do projeto como integral, parcial ou inexistente; a principal evidência de êxito; e os próximos passos recomendados. Essa estrutura permitiu uma avaliação multidimensional do Ciclo 2, evitando que o êxito fosse medido apenas pelo funcionamento técnico da solução.

Por fim, quanto ao Plano de Descontinuidade (PdD) da empresa Synapse, é possível afirmar que houve atualização no Ciclo 2, sendo que foram identificadas as seguintes necessidades de ajustes em razão da incompletude: inventário de dados e terceiros, gatilhos de acionamento, governança, medidas técnicas e matriz de riscos residuais devem ser apresentados e/ou aperfeiçoados à luz da evolução do Projeto Trajetto no que se refere às recomendações indicadas para o Ciclo 3. Ou seja, quanto ao RIPD adaptado aos riscos de IA e geolocalização, ao mapa do fluxo de dados ponta a ponta, à matriz de necessidade e proporcionalidade de dados, ao relatório de risco de reidentificação, ao inventário de APIs e terceiros, à validação de fluxos nacionais e eventual transferência internacional, ao plano de resposta a incidentes, ao protocolo de *fairness* e acessibilidade, às evidências concretas de transferência ao titular, e ao registro de versionamento de modelos, parâmetros, algoritmos e *pipelines* utilizados ao longo do Sandbox.

4.1.4 Resultados do Ciclo 2 sob a perspectiva da ANPD

No Ciclo 2, a análise do Trajetto permitiu aprofundar a compreensão regulatória dos sistemas de recomendação de rotas baseados em geolocalização, em dados contextuais e em integrações com APIs externas. Um dos principais resultados foi a caracterização mais precisa da solução como sistema de recomendação, no qual a decisão final permanece com o usuário ou operador humano, e não como um mecanismo de decisão automatizada impositiva. Essa distinção contribuiu para avaliar, de maneira proporcional, os deveres de transparência, explicabilidade e governança aplicáveis à solução.

Quanto à proteção dos usuários e titulares, o ciclo possibilitou identificar riscos relacionados à geolocalização contínua, à formação de padrões de mobilidade, ao gerenciamento de sessões, à integração com terceiros e à possibilidade de identificação

indireta. A análise também permitiu delimitar as salvaguardas necessárias, como pseudonimização efetiva, criptografia, segregação de ambientes, inventário de APIs, avaliação da necessidade e da proporcionalidade dos dados, análise do risco de reidentificação e teste do plano de resposta a incidentes.

Embora o Ciclo 2 ainda não tenha produzido elementos conclusivos para a proposição de nova norma, foram identificados aprendizados preliminares relevantes para futuras orientações da ANPD, especialmente sobre:

- avaliação de sistemas de recomendação baseados em geolocalização;
- distinção entre anonimização e pseudonimização em soluções de mobilidade;
- adoção de mecanismos de transparência em camadas;
- versionamento de modelos, parâmetros, algoritmos e *pipelines*; e
- critérios mínimos para validação de sistemas de IA em ambientes controlados.

Também foi identificado potencial para estimular a inovação responsável no setor de mobilidade urbana, especialmente em soluções de recomendação de rotas, gestão de fluxos e redução de aglomerações. Contudo, não foram constatadas evidências suficientes para avaliar efeitos concretos sobre o mercado ou a concorrência, pois a solução ainda se encontrava em fase de validação.

4.1.5 Resultados do Ciclo 2 sob a perspectiva da Synapse

Para a Synapse, o Ciclo 2 resultou em avanços técnicos, regulatórios e organizacionais. A empresa informou o aperfeiçoamento de componentes da arquitetura do Trajetto mediante a incorporação de pseudonimização, criptografia, segregação dos ambientes de desenvolvimento e produção, uso de VPN e melhorias no processo de gerenciamento de sessões. O acompanhamento também evidenciou a necessidade de reforçar a rastreabilidade e o versionamento dos modelos, parâmetros, algoritmos, integrações e fluxos de processamento.

No campo regulatório, a participação no Sandbox contribuiu para que princípios da LGPD, como finalidade, necessidade e proporcionalidade, fossem incorporados às decisões de arquitetura e desenvolvimento da solução. A distinção entre anonimização e pseudonimização, assim como a caracterização do Trajetto como sistema de recomendação, permitiu à empresa compreender com maior precisão as obrigações aplicáveis e estruturar uma governança proporcional aos riscos identificados.

Em relação à transparência e à explicabilidade, foram definidas duas frentes de aprimoramento. A primeira consistiu na reformulação da comunicação com o usuário, com avisos contextuais e informações organizadas em camadas. A segunda envolveu o desenvolvimento de mecanismos de explicabilidade e transparência algorítmica capazes de produzir justificativas legíveis e evidências auditáveis para as rotas recomendadas.

Quanto aos direitos dos titulares, foram mapeados os fluxos relacionados à confirmação do tratamento, ao acesso, à correção, à anonimização, à eliminação e à revisão, com a identificação dos pontos da arquitetura responsáveis por atender às solicitações. Permaneceram pendentes, entretanto, documentos e validações indispensáveis à progressão segura do projeto, incluindo o RIPD, a avaliação de impactos

sobre direitos fundamentais, a matriz de necessidade e proporcionalidade, o relatório de risco de reidentificação, o inventário de API e de terceiros, o plano de resposta a incidentes testado e as evidências concretas de transparência.

Como aprendizado organizacional, a Synapse relatou que a participação no Sandbox contribuiu para a formação de uma frente interna de avaliação e controle de produtos de IA, com critérios de governança desde a concepção, avaliação prévia de riscos regulatórios e definição das evidências mínimas necessárias antes da expansão de uma solução.

A avaliação foi fundamentada na triangulação das declarações da empresa com a documentação apresentada, os registros técnicos, os resultados dos testes, as respostas aos instrumentos metodológicos e os pareceres produzidos pelas equipes especializadas. Entre as evidências coletadas e analisadas no Ciclo 2 estiveram:

- documentação da arquitetura da solução;
- descrição dos mecanismos de pseudonimização e criptografia;
- segregação dos ambientes de desenvolvimento e produção;
- configuração do túnel e da VPN;
- informações sobre o gerenciamento de sessões;
- registros dos fluxos de dados e das integrações externas;
- resultados dos testes relativos à identificação de pessoas em imagens;
- informações sobre transparência, explicabilidade e direitos dos titulares; e
- documentos e declarações sobre controles de segurança e medidas de mitigação.

A coleta de evidências também permitiu identificar o que não havia sido suficientemente comprovado. Essa distinção foi essencial para separar controles efetivados de medidas em andamento, planejadas ou não comprovadas, reduzindo a dependência de autodeclarações e transformando as lacunas identificadas em condicionantes monitoráveis para os ciclos seguintes.

O êxito do Ciclo 2 foi classificado como parcial. Nesse contexto, a principal evidência de êxito foi a caracterização mais precisa do Trajetto como sistema de recomendação de rotas, acompanhada da identificação de uma arquitetura tecnicamente estruturada de pseudonimização, criptografia e segregação de ambientes. O ciclo também permitiu identificar, de maneira objetiva, os riscos, as salvaguardas e as evidências necessárias à continuidade segura da solução.

A progressão do projeto permaneceu condicionada à apresentação e à validação dos artefatos pendentes. Essa conclusão demonstra a relevância da Fase C para a metodologia: consolidar os resultados alcançados, registrar os aprendizados da ANPD e da empresa, distinguir avanços comprovados de medidas ainda não demonstradas e converter as lacunas identificadas em ações concretas para o ciclo subsequente.

4.2 Prevvine

O relatório do CIAAM/USP referente ao Ciclo 2 de testagem da solução apresentada pela Prevvine foi entregue à ANPD em 1º de junho de 2026. O documento consolidou as informações, os registros e as evidências produzidos nas Fases A e B do ciclo, abrangendo 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 2, 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 Fases 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;
- TRL e ETC: 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 Fase 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 2 foi concluído em suas três fases — Fase A, de preparação e detalhamento; Fase B, de experimentação e gestão de riscos; e Fase C, de avaliação e ajuste —, a seguir, apresentam-se os principais aspectos identificados em cada uma dessas etapas.

4.2.1 Fase A

No que se refere à Ficha A1, ao se analisar os seguintes itens: arquitetura do modelo, fluxo de dados, API e Inteligência Artificial, e Protocolo de Segurança, observa-se a inexistência de alterações entre o Ciclo 1 e o Ciclo 2 de testagem.

Quanto à Ficha A2, foram identificados quatro ETC: motor de IA; camada de anonimização; sistema de verificação de referências bibliográficas externas; e infraestrutura de conformidade e LGPD. Não houve alteração do nível de maturidade (TRL) entre o Ciclo 1 e o Ciclo 2 para esses quatro elementos, e todos foram classificados como TRL 7, ou seja, implementados e operacionais.

Diante do exposto, observa-se que não houve mudanças entre os ciclos 1 e 2, assim como ocorreu na hipótese da empresa Synapse. Um dos motivos é exatamente o curto espaço de tempo entre um ciclo e outro, o que não deixou muita margem para alterações por parte das empresas.

Em relação às fichas A3 e A4 da Prevvine, constata-se pouca alteração entre o Ciclo 1 e o Ciclo 2 em razão do curto espaço de tempo, no entanto, alguns apontamentos são dignos de análise.

Primeiramente, apresentam-se, no Quadro 5, os diplomas normativos analisados no Ciclo 2:

Quadro 5: Diplomas normativos analisados para a empresa Prevvine

DIPLOMAS NORMATIVOS	
Lei 15.211/2025	Estatuto Digital da Criança e do Adolescente (ECA Digital)
Lei 13.787/2018	Prontuário Eletrônico do Paciente
Lei 13.709/2018	Lei Geral de Proteção de Dados Pessoais (LGPD)
Lei 13.146/2015	Lei Brasileira de Inclusão da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência)
Lei 12.965/2014	Marco Civil da Internet
Lei 12.984/2014	Combate à Discriminação contra Pessoas com HIV/AIDS
Lei 12.288/2010	Estatuto da Igualdade Racial
Lei 10.741/2003	Estatuto da Pessoa Idosa
Lei 8.080/1990	Lei Orgânica da Saúde
Lei 8.078/1990	Código de Defesa do Consumidor (CDC)

Lei 8.069/1990	Estatuto da Criança e do Adolescente (ECA)
Decreto 12.880/2026	Regulamento do ECA Digital
Resolução CD/ANPD 2/2022	Regulamento para agentes de tratamento de pequeno porte
Resolução CFM 2.314/2022	Telemedicina e uso de IA na Saúde
RDC ANVISA 657/2022	Software como Dispositivo Médico (SaMD)
Resolução CFM 2.454/2026	Norma deontológica complementar do CFM

Fonte: ANPD

Na sequência, os documentos internos avaliados são listados no Quadro 6.

Quadro 6: Documentos internos avaliados para a empresa Prevvine

Nº	Documento Interno Avaliado
01	Política de Privacidade e Proteção de Dados
02	Política de Segurança da Informação
03	Política de Retenção e Descarte de Dados
04	Política de Resposta a Incidentes de Segurança
05	Registro de Operações de Tratamento
06	Inventário de Sistemas e Ativos de IA
07	Termo de Consentimento Livre e Esclarecido (TCLE) para uso do <i>chatbot</i>

Por impossibilidade de disponibilização para análise dos documentos 01 (Política de Privacidade e Proteção de Dados) e 03 (Política de Retenção e Descarte de Dados), os mencionados documentos passaram a ser objeto de avaliação no Ciclo 3.

Da análise da possibilidade de sobreposição normativa, foi possível observar que a caracterização do *chatbot* como *Software* como Dispositivo Médico (SaMD) poderá submetê-lo simultaneamente à regulamentação sanitária da ANVISA e à LGPD. Enquanto a ANVISA examina a segurança e a eficácia clínica da solução, a ANPD fiscaliza a licitude, a segurança e a transparência do tratamento de dados pessoais. Essa dupla incidência produz obrigações parcialmente sobrepostas quanto à documentação dos dados de treinamento, à mitigação de vieses, à cibersegurança, à vigilância pós-mercado e à comunicação de incidentes.

O ponto central consiste em definir a finalidade do *chatbot*. Se a ferramenta apenas coletar informações e encaminhar o usuário a um profissional, poderá não ser enquadrada como SaMD, embora permaneça integralmente sujeita à LGPD. Caso produza recomendações clínicas, classificação de urgência ou resultados com finalidade diagnóstica ou terapêutica, poderá incidir também o regime sanitário. A revisão da RDC 657/2022 tende a aproximar ainda mais os dois regimes ao incorporar requisitos de robustez, transparência, controlabilidade e intervenção humana, tornando necessária maior coordenação entre ANVISA e ANPD.

Em conclusão, à luz do exposto nas fichas A3 e A4 do Ciclo 2, verifica-se que o desenvolvimento e a implementação da solução de IA em saúde no contexto da Previne/STAiDOC se inserem em ambiente regulatório dinâmico, fragmentado e em processo de consolidação, envolvendo a incidência concorrente de normas éticas, sanitárias e de proteção de dados pessoais, bem como a perspectiva de evolução normativa com a possível aprovação do Projeto de Lei 2.338/2023.

Nesse cenário, a adoção de uma abordagem estruturada de governança, baseada em interoperabilidade regulatória, gestão de riscos e conformidade progressiva, revela-se não apenas recomendável, mas necessária para assegurar a segurança jurídica, a proteção dos titulares de dados e a aderência a padrões éticos e técnicos aplicáveis.

Diante da análise desenvolvida acima, recomendou-se à ANPD, por meio da Ficha A4, a consideração e eventual adoção das seguintes medidas, detalhadas nas seções anteriores:

- Estabelecer contato formal com outras agências e autoridades (ANS, ANVISA e CFM), visando à atuação coordenada interinstitucional, para a construção de diretrizes conjuntas e interoperáveis e a elaboração de exigências normativas e de governança integradas, com padrões técnicos interoperáveis;
- Realizar um mapeamento normativo integrado contínuo, acompanhado de uma matriz de conformidade a ser atualizada periodicamente;
- Adotar documentação técnica modular e interoperável;
- Implementar mecanismos de rastreabilidade e auditabilidade de fluxos de dados e decisões assistidas por IA;
- Avaliar e confirmar a adequação à base de tratamento de dados pessoais para a tutela da saúde, nos termos do art. 7º, Inciso VIII, da

LGPD, em eventual caso de processamento de dados pessoais durante a interação com o *chatbot*.

- Estabelecer protocolos unificados de gestão de incidentes e comunicação regulatória;
- Exigir evidências concretas dos mecanismos de transparência e explicabilidade prometidos, inclusive telas, avisos, fluxos de consentimento, textos, *logs* de exibição e testes de entendibilidade, combinada com a adoção de Linguagem Simples e elementos visuais adequados, aumentando a acessibilidade e reduzindo a assimetria informacional.

4.2.2 Fase B

Esta parte agrupou os instrumentos operacionais utilizados em cada uma das três camadas de avaliação (Teste 1 — *model test*, Teste 2 — *red teaming*, Teste 3 — avaliação por especialista), bem como os testes adicionais de explicabilidade e transparência (testes 4 e 5).

Comum a todas as empresas, a Fase B inicia-se com a aplicação do framework da UNESCO, seguindo a metodologia apresentada anteriormente, sendo que o somatório das pontuações é de 100 pontos. De acordo com a planilha utilizada, no Ciclo 2, a empresa Previne alcançou uma pontuação de 60, inferior à do Ciclo 01, de 89 pontos. É relevante destacar que, no Ciclo 1, foi a própria empresa que realizou sua autoavaliação, pois foi considerado relevante receber os dados de entrada por meio de uma autoavaliação das empresas antes de iniciar os testes no âmbito do Sandbox para comparar com a análise da equipe de CIAAM/USP no Ciclo 2. Trata-se de uma maneira relevante para a empresa acompanhar a análise de um terceiro em relação à sua solução de IA e observar pontos que merecem atenção ao longo dos ciclos de testagem.

Os pontos de divergência entre o Ciclo 1 e o Ciclo 2 referem-se aos seguintes eixos: (i) Governança e responsabilidade clara (Eixo 2), com uma diferença de 8 pontos; (ii) Avaliação de impactos (Eixo 3), com uma diferença de 15 pontos; (iii) Transparência, explicabilidade e comunicação (Eixo 5), com uma diferença de 4 pontos; e (iv) Segurança técnica e robustez (Eixo 6), com uma diferença de 2 pontos. Dessa forma, considerou-se, no Ciclo 2, um risco ético de moderado a alto, que exigia mitigação antes de escalar.

Em relação aos testes, ao se analisar o Teste 1, torna-se relevante expor que o instrumento de aplicação do Teste 1 — *model test* — é voltado à avaliação conversacional do modelo nos quatro aspectos selecionados (Validade e Precisão, Segurança, Viés e Equidade e Privacidade). Assim, foram registrados 30 casos de conversas (*prompts*) para cada aspecto selecionado, com os respectivos critérios de aprovação, gabarito esperado e campo para resposta do avaliador. A elaboração dos *prompts* seguiu a metodologia descrita no Manual para Testagem de LLM/LMM em Função de Conversa Aplicada à Saúde que considera o público-alvo de aplicação e os critérios de tolerância que são próximos de zero para riscos diretos à vida, conforme a Organização Mundial de Saúde (OMS).

O banco de *prompts* utilizados na aplicação do Teste 1 contempla casos de conhecimento factual, falsa premissa, defasagem temporal, adequação contextual, pares pareados (gênero, raça/etnia, idade, deficiência) e robustez. Recomenda-se um mínimo de 30 a 50 itens por categoria, validados por dois revisores clínicos independentes. O manual

disciplina a engenharia de redação, o controle de variáveis demográficas, o versionamento e a reprodutibilidade estocástica.

No Ciclo 2, iniciou-se a elaboração de *prompts* e a estruturação para os quatro testes que ocorrerão nos ciclos seguintes. Dessa forma, neste relatório, apenas serão apresentados brevemente os testes que ocorreram:

- Teste 1 — Avaliação conversacional do modelo: avaliação do desempenho do *chatbot* em quatro dimensões: validade e precisão, segurança, viés e equidade e privacidade. O teste utiliza vinhetas e um banco estruturado de *prompts* — incluindo falsas premissas, informações desatualizadas, variações contextuais e pares demográficos — com critérios de aprovação, respostas esperadas e repetição das execuções para controlar a variabilidade do modelo.
- Teste 2 — *Red teaming*: submissão do sistema a cenários adversariais para identificar vulnerabilidades e verificar a resistência dos controles de segurança. Os ataques incluem *jailbreaks*, fragmentação em múltiplas interações, falsa autoridade clínica e indução por urgência. Os resultados devem ser classificados conforme sua severidade e registrados para orientar medidas de mitigação.
- Teste 3 — Avaliação por especialista: análise clínica das respostas produzidas nos testes anteriores. O especialista classifica cada resultado como aprovado, falha moderada ou falha grave, apresenta justificativa clínica e registra eventuais riscos ou indícios de viés. Cada *prompt* deve ser executado repetidamente para captar possíveis variações nas respostas do modelo.
- Teste 4 — Transparência e explicabilidade: verificação da implementação de três níveis de explicabilidade, adequados a públicos diferentes: camada técnica para Tecnologia da Informação (TI) e auditoria; camada gerencial para alta administração, governança e *compliance*; e camada acessível para pacientes, profissionais de saúde e demais usuários. A aprovação exige que os elementos essenciais de todas as três camadas estejam presentes e operacionais.

Já o Teste 5 consiste na matriz de transparência, e foi elaborado um manual específico para sua aplicação. O documento é baseado na proposta de Maranhão, Junquilho e Tasso (2023, p. 169–170) no artigo *Transparência sobre o emprego de Inteligência Artificial no Judiciário: um modelo de governança*, publicado na revista SUPREMA. Embora a proposta original tenha como universo de aplicação o Poder Judiciário brasileiro, sua estrutura conceitual é transponível para qualquer organização que empregue sistemas de IA, bastando traduzir os interlocutores para os papéis equivalentes no contexto específico de sistemas de IA na saúde.

Na matriz de transparência aplicada ao Ciclo 2, verificaram-se os seguintes resultados:

- Quanto aos usuários: a avaliação identificou a necessidade de aprimorar a comunicação com os usuários, especialmente mediante a revisão das políticas para adoção de linguagem mais clara, acessível e adequada à compreensão dos titulares. Também deverá ser

estruturado um fluxo formal para a revisão de decisões automatizadas, com definição do formato, do momento e dos canais em que essa possibilidade será apresentada ao usuário. Embora tenha sido apresentado termo de consentimento, ainda é necessário avaliar se o consentimento constitui a base legal mais adequada para o tratamento dos dados, considerando as demais hipóteses previstas na LGPD. Recomenda-se, ainda, padronizar e unificar a apresentação dessas informações nos diferentes canais utilizados pela solução, garantindo consistência, transparência e facilidade de acesso pelo titular.

- Quanto aos médicos (usuários): a avaliação dos médicos sobre o uso da ferramenta ainda se encontra em revisão, sendo necessário consolidar suas percepções quanto à utilidade, segurança, clareza e adequação das respostas produzidas. Também foi identificada a necessidade de disponibilizar um manual de uso específico, com orientações sobre as funcionalidades, limitações, interpretação dos *outputs* e atualizações realizadas no sistema. Recomenda-se complementar essas medidas com treinamentos e *workshops*, interface explicativa e documentação técnica acessível aos profissionais. As informações também devem ser apresentadas de forma consistente nos termos de uso, na política de privacidade e nas comunicações por *e-mail*, garantindo que os médicos compreendam o funcionamento da ferramenta, seus limites e suas responsabilidades durante a utilização.
- Quanto aos hospitais, clínicas e instituições de saúde: para os hospitais, a matriz identificou a necessidade de definir claramente os responsáveis pela adoção, supervisão e governança da ferramenta. Como controladores de dados pessoais, os hospitais precisam compreender o funcionamento da inteligência artificial, sua arquitetura, seus riscos, suas medidas de segurança, suas políticas e suas métricas de desempenho para cumprir adequadamente as obrigações previstas na LGPD. Também devem ser estabelecidos procedimentos para contratação, auditoria, monitoramento e gestão de incidentes, incluindo um plano formal de resposta a incidentes. Recomenda-se que essas informações sejam consolidadas em contratos técnicos, relatórios periódicos e *dashboards* de acompanhamento, permitindo aos hospitais avaliar continuamente o desempenho, a conformidade e os riscos da solução.
- Quanto à ANPD e reguladores: para a ANPD e os demais reguladores, a matriz destacou a necessidade de assegurar informações suficientes para a supervisão do Sandbox e a verificação da conformidade da solução com a LGPD e com as normas setoriais aplicáveis. As evidências devem abranger, entre outros elementos, o Relatório de Impacto à Proteção de Dados, os registros de operação e auditoria, as métricas de vieses, os incidentes identificados e as medidas de mitigação adotadas ou planejadas. Essas informações deverão ser apresentadas de acordo com o cronograma do Sandbox, sempre que ocorrerem incidentes relevantes e, quando necessário, mediante solicitação das autoridades competentes. Permanece necessário detalhar os procedimentos de demonstração da conformidade,

incluindo os formatos, responsáveis, periodicidade, canais de envio e critérios de validação das evidências apresentadas.

- Quanto à equipe de TI e segurança: para a equipe de TI, a matriz identificou a necessidade de assegurar o monitoramento contínuo do funcionamento, da segurança e do desempenho da solução. Esse acompanhamento deve ser fundamentado em métricas técnicas, registros de operação e auditoria, alertas de segurança e documentação atualizada das API e integrações utilizadas pelo sistema. Embora os ciclos de avaliação tenham sido detalhados no Plano de Sandbox, a empresa deverá também estabelecer um ciclo interno e periódico de revisão técnica. Recomenda-se consolidar os procedimentos e as evidências de acompanhamento em manuais técnicos, *dashboards* e na documentação dos sistemas, com a definição dos responsáveis, da periodicidade das revisões e das providências aplicáveis diante de falhas, incidentes ou alterações de desempenho.

Na Fase B, os resultados e as evidências produzidos durante a testagem foram consolidados em parecer técnico destinado a avaliar a segurança da informação, a privacidade e a governança da solução. A metodologia combinou: (i) matriz de riscos 5 × 5, baseada na probabilidade e no impacto; (ii) análise de controles segundo a LGPD e as normas ISO/IEC 27001 e 27002; e (iii) classificação das medidas de mitigação como efetivadas, em andamento, planejadas ou não comprovadas. Também foram definidos gatilhos objetivos para pausa dos testes, critérios para eventual extensão e evidências necessárias à continuidade da solução.

Neste sentido, a análise reconheceu a existência de controles técnicos e práticas de proteção de dados incorporadas à arquitetura, incluindo mecanismos de minimização, descarte de dados originais, registros auditáveis e verificação de referências externas. Ao mesmo tempo, foram identificadas oportunidades de aprimoramento relacionadas à proteção contra reidentificação, à segurança da aplicação, à continuidade dos serviços, à mitigação de respostas inadequadas e à estabilidade da integração com o ambiente de testes. O risco inerente foi classificado como alto, em razão do tratamento de dados de saúde por meio de um sistema de IA, e o risco residual como médio após a consideração dos controles existentes.

Por fim, recomendou-se a continuidade dos testes com condicionantes, vinculando o avanço da solução à implementação e à comprovação das medidas de segurança ainda pendentes. Essa etapa demonstra a função central da Fase B na metodologia: transformar os resultados dos testes em uma avaliação estruturada, distinguir controles comprovados de medidas apenas planejadas e estabelecer providências verificáveis para o ciclo seguinte.

4.2.3 Fase C

Seguindo os mesmos procedimentos adotados para a empresa Synapse, na Fase C, a Agência reguladora e a empresa participante preenchem relatórios de êxitos que sistematizam os aprendizados produzidos para ambos, verificando os respectivos benefícios. A metodologia da Fase C é abordada, de forma mais detalhada, no item 4.1.3.

A seguir, é feita uma síntese sobre o ciclo 2, segundo a ótica da ANPD e também da empresa Prevvine.

4.2.4 Resultados do Ciclo 2 sob a perspectiva da ANPD

No Ciclo 2, a análise da plataforma STAiDOC permitiu à ANPD aprofundar a compreensão dos controles necessários à conformidade de sistemas de inteligência artificial destinados ao suporte à decisão médica. Foram identificadas evidências da adoção de uma arquitetura baseada em *Privacy by Design*, com tratamento dos dados antes de sua comunicação ao modelo de inteligência artificial.

Um dos principais aprendizados regulatórios consistiu na necessidade de distinguir tecnicamente entre anonimização e pseudonimização. Embora a empresa inicialmente caracterizasse o procedimento aplicado como anonimização, a avaliação indicou a permanência de uma possibilidade teórica de reidentificação dos titulares. Dessa forma, o mecanismo foi classificado como pseudonimização, com a consequente manutenção da incidência das obrigações previstas na LGPD.

Quanto à proteção dos usuários e titulares, foram identificados riscos relacionados à eventual reidentificação de dados de saúde e à necessidade de fortalecer os controles de segurança no nível da aplicação. A ausência de comprovação de determinados mecanismos de proteção, como controles de filtragem e defesa da aplicação web, demonstrou a importância de não limitar a avaliação às declarações da empresa, exigindo evidências técnicas e documentais suficientes.

Embora o Ciclo 2 tenha produzido aprendizados regulatórios importantes, ainda não foram identificados elementos suficientemente maduros para fundamentar a elaboração de uma nova norma pela ANPD. Também não foi possível dimensionar, nesse estágio da testagem, os efeitos concretos da solução sobre o mercado ou a concorrência, considerando que o projeto ainda se encontrava em ambiente controlado e dependia da mitigação dos riscos apontados.

4.2.4 Resultados do Ciclo 2 sob a perspectiva da Prevvine

Para a Prevvine, o Ciclo 02 contribuiu para a consolidação da arquitetura de *Privacy by Design* da plataforma STAiDOC e para a verificação da estabilidade dos quatro Elementos Tecnológicos Críticos identificados: motor de IA; camada de anonimização; sistema de verificação de referências bibliográficas externas; e infraestrutura de conformidade e LGPD.

A empresa também informou a implementação de três níveis de explicabilidade — técnico, institucional e simplificado — e da Matriz de Transparência, com registros específicos dos eventos relacionados à explicabilidade. Esses mecanismos foram estruturados para atender diferentes destinatários, incluindo reguladores, médicos, gestores, pacientes e demais usuários.

No campo regulatório, a principal aprendizagem decorreu da distinção entre pseudonimização e anonimização. A empresa reconheceu que o procedimento empregado não deveria ser apresentado como anonimização absoluta e indicou que

ajustaria sua documentação técnica e sua comunicação pública para refletir corretamente o nível de proteção oferecido. Também foi identificada a necessidade de demonstrar formalmente a existência de controles adicionais de segurança no nível da aplicação e de incorporar essa providência ao planejamento dos ciclos seguintes.

Em relação à transparência e à explicabilidade, a avaliação da experiência do usuário indicou que o aviso de transparência, disponibilizado antes do acesso à plataforma, era legível e compreensível para o público médico. A Matriz de Transparência contemplou diferentes públicos, e as respostas geradas pelo sistema passaram a conter indicação obrigatória de supervisão humana, reforçando o caráter de apoio à decisão, e não de substituição do profissional de saúde.

A análise dos direitos dos titulares ainda não foi aprofundada no Ciclo 2, pois esse tema foi previsto para um ciclo posterior. Entretanto, o risco de reidentificação percebido durante a testagem antecipou a necessidade de desenvolver controles adicionais relacionados à confirmação do tratamento, à eliminação de dados sensíveis e à proteção de populações vulneráveis.

Quanto às análises pendentes, permaneceu necessária a conclusão da avaliação da integração entre a plataforma da empresa e o ambiente de testes utilizado pelo CIAAM/USP. Embora a implementação técnica tenha sido informada como concluída, sua operação e as evidências deveriam continuar sendo acompanhadas no desenvolvimento dos ciclos seguintes.

4.2.5 Coleta e análise de evidências

A avaliação dos êxitos baseou-se na comparação entre as informações fornecidas pela Previne e as evidências produzidas ao longo do ciclo. Foram considerados, entre outros elementos:

- as fichas técnicas A1 e A2;
- o Relatório Parcial do Ciclo 2;
- o Parecer Técnico Consolidado do CIAAM/USP;
- a Matriz de Transparência e Explicabilidade;
- os registros de eventos de explicabilidade;
- a avaliação da interface e do aviso de transparência;
- a documentação dos Elementos Tecnológicos Críticos; e
- os registros relativos à integração com o ambiente de testes.

Essa análise permitiu distinguir os avanços comprovados das medidas ainda pendentes de implementação ou de validação. Também tornou possível registrar, de forma rastreável, a relação entre o resultado declarado, a evidência apresentada e a conclusão adotada no ciclo.

Finalmente, o êxito do Ciclo 2 da Previne foi classificado como parcial. Pois foram reconhecidos esforços de conformidade, avanços na arquitetura de *Privacy by Design*, estabilidade dos componentes tecnológicos críticos e aprimoramentos relacionados à transparência e à explicabilidade. Contudo, permaneceram riscos e controles que deveriam ser mitigados ou comprovados antes da progressão segura da solução.

Como próximos passos, recomendou-se o cumprimento das condicionantes técnicas constantes do Parecer Técnico Consolidado, especialmente quanto ao risco de reidentificação, à segurança da aplicação e à apresentação de evidências dos controles adotados. Quanto ao plano de descontinuidade (PdD), não houve alteração em relação ao PdD encaminhado no encerramento do treinamento das empresas, em que elas precisavam apresentar o Plano de Sandbox e o Plano de Descontinuidade. Acredita-se que, após os testes conduzidos no Ciclo 3, esse plano provavelmente precisará ser atualizado, já que serão analisados vários aspectos do escopo da testagem de forma mais aprofundada.

4.3 Metatext

A Metatext é uma *startup* que possui uma estrutura mais enxuta, o que demandou readequação do Plano de Sandbox e do Plano de Descontinuidade para refletir a realidade da empresa. Dessa forma, haverá menos ciclos e o tempo de testagem será maior do que o das demais participantes. Embora haja esse ajuste quanto ao tempo demandado por ciclo, a metodologia não é alterada, sendo que serão realizadas, em cada ciclo, três fases: Fase A, B e C.

No Ciclo 2, iniciou-se a elaboração das fichas A1 e A2, sendo que a dinâmica foi diversa, pois foram fornecidos os documentos pela empresa e a equipe regulatória responsável pela solução da Metatext preencheu as fichas e validou com o responsável da empresa. Considerou-se oportuno fazer dessa maneira, ou seja, a equipe do CIAAM USP preencheu as fichas A1 e A2, pois houve um período prolongado de reuniões e discussões para a compreensão da solução, que por natureza é a mais abstrata de todas. Assim, era necessário preencher para verificar se o que fora compreendido em termos de arquitetura da solução e fluxo de dados estavam corretos para que fossem estruturados os testes.

Os documentos fornecidos pela Metatext na Fase A foram:

- Segurança da Informação: controles de acesso, criptografia (AES-256/TLS 1.2+), testes de penetração e auditorias.
- Gestão de Acessos: Princípio do menor privilégio e autenticação multifator obrigatória.
- Criptografia: padrões de chaves criptográficas e rotação programada.
- Resposta a incidentes (*Incident Response Plan* — IRP): protocolo de contenção e notificação à ANPD em até 72h.

- Relatório de Impacto à Proteção de Dados Pessoais: avaliação de impacto à proteção de dados pessoais detalhando riscos e medidas de mitigação da solução.
- Retenção e Descarte: expiração automática de *logs* e anonimização irreversível.

É relevante destacar que, em razão da natureza do setor de atuação da empresa (cibersegurança no setor financeiro), a documentação apresentada estava integralmente em Língua Inglesa.

Adicionalmente, estabeleceu-se uma sequência de perguntas realizadas pela equipe do CIAAM/USP, com as respectivas respostas dadas pela empresa, que também é integrada à análise feita nas fichas A1 e A2.

Em relação aos principais itens da Ficha A1, as principais conclusões foram:

- Arquitetura do sistema e fluxo de dados: a arquitetura do GuardionAI foi projetada como uma camada de privacidade que se integra entre os usuários, agentes de IA e os modelos de terceiros. O fluxo de dados é estruturado em treze etapas principais e foi detalhado de maneira satisfatória, pois garante que os modelos de terceiros nunca tenham acesso a dados em claro, enquanto mantém a funcionalidade total da IA para os usuários finais;
- Rastreabilidade: o sistema oferece rastreabilidade em múltiplos níveis (metadados de rastreabilidade por requisição, atribuição de atributos, *logs* de auditoria dos *spans* detectados e explicabilidade pós-processamento). Também, há o alinhamento preliminar dos quatro níveis descritos com as principais exigências normativas do Banco Central do Brasil (BACEN);
- Modelo de IA: houve o atendimento parcial tendo em vista que alguns dos aspectos mencionados geram outros questionamentos que requerem, por sua vez, demonstração específica. Exemplos: (i) como é feito o “*fine-tuning* especializado” mencionado? (ii) É indicada “política”, mas não houve o envio da respectiva documentação conforme solicitado; (iii) São indicados “*dashboards* centralizados”, mas não foram apresentados os respectivos “*prints*” para demonstração específica. Em suma, entende-se ser mais adequado retomar estes pontos na próxima etapa, a fim de possibilitar o seu aprofundamento técnico e regulatório;
- Explicabilidade: houve a explicação detalhada para os titulares e os controladores/auditores, mas não houve menção a um aspecto importante da solução, a qual é focada na equipe de segurança e também precisa dessa explicação;
- Legislação: a LGPD, a Resolução CMN 4.893/2021, a Resolução BCB 85 de 8/4/2021 e o PL 2338/2023 (visa a regular a IA) foram os documentos jurídicos apontados pela empresa como incidentes à sua solução. Dessa forma, concluiu-se que é necessário estender a análise de normas aplicáveis, contemplando, inclusive, normas editadas pela Comissão de Valores Mobiliários (CVM) e pela Superintendência de Seguros Privados (SUSEP). Trata-se de aspecto central do Sandbox e,

neste sentido, tendo em vista as novas regras publicadas neste 1º semestre de 2026 pelo BACEN, será importante a confirmação da adoção das principais exigências na próxima etapa do Sandbox;

- Transferência internacional de dados: a participante indica que não há transferência internacional de dados, mas a sua página na internet possui extensa documentação sobre o assunto. Assim, é importante explicar a sua resposta diante da documentação identificada.

Diante dessas conclusões, considerou-se necessária a complementação de documentação interna e regulatória pela empresa, no entanto, em termos de arquitetura do sistema, considerou-se suficiente a continuidade do ciclo de testagem no âmbito do Sandbox.

No que se refere à Ficha A2 do Ciclo 2, pode-se observar que foram identificados os seguintes ETC:

- Infraestrutura: O TRL identificado foi 5, sendo o mesmo do Ciclo 1;
- Pseudonimização: O TRL identificado foi 5, sendo o mesmo do Ciclo 1;
- Funcionamento do LLM proprietário: O TRL identificado foi 7, sendo o mesmo do Ciclo 1;
- *Feature Attribution*: O TRL identificado foi 2, diferente do Ciclo 1, identificado como 1.

É relevante destacar que, no Ciclo 1, foram realizadas diversas reuniões para a identificação de diversos elementos fundamentais para o preenchimento das fichas A1 e A2 do Ciclo 2, pois ainda não havia clareza sobre a solução de IA apresentada pela empresa e a forma como os dados sintéticos seriam empregados no Sandbox, sendo que foram solucionadas essas questões no final do Ciclo 2, sendo este o motivo de não ter se avançado para as fichas A3 e A4, as quais serão apresentadas no próximo relatório.

Diante do exposto, as atividades realizadas nesta etapa da testagem quanto à Metatext podem ser descritas abaixo na ordem cronológica:

- Atividades até 22 de maio de 2026: (i) foi realizada uma reunião com representante da Metatext e com a equipe do HarpIA, na qual ficou acordado que será usado o sistema HarpIA Survey (também utilizado pela Prevvine) e será avaliada a aplicação do HarpIA Lab para um outro ciclo de testagem; (ii) As fichas A1 e A2 foram enviadas pela equipe da área regulatória para validação pelo responsável da empresa; (iii) uma API já pronta será disponibilizada para a equipe técnica, de forma a tentar conexão com o HarpIA Survey; (iv) ficou acordado que a Fase B do Ciclo 2 iniciará com o teste no Harpia Survey.
- De 22 a 29 de maio de 2026: (i) a empresa aprovou o conteúdo das fichas A1 e A2, permanecendo pendente apenas o preenchimento dos ETC, que constituem objeto de avaliação do TRL. Ficou acordado que essa etapa será concluída até o dia 1º de junho; (ii) a equipe regulatória reuniu-se em 26 de maio para alinhar o preenchimento das fichas A3 e A4. Durante a análise, constatou-se a necessidade de documentação complementar, cuja solicitação foi encaminhada à empresa por meio

de e-mail em 29 de maio; (iii) em 28 de maio, foi realizada reunião com toda a equipe do CIAAM/USP para alinhamento das entregas relativas aos ciclos 1 e 2. No que se refere à Metatext, identificou-se a necessidade de uma reunião adicional para definição dos testes que serão conduzidos pela equipe técnica; (iv) também foi concluída a integração do sistema da Metatext com o HarpIA, o que viabilizará a análise dos mecanismos de pseudonimização dos dados. Permanece pendente a definição da metodologia para realização dos ataques adversariais, tema que será discutido na próxima reunião, agendada para 1º de junho, às 16h; adicionalmente, está sendo avaliada a possibilidade de utilização do HarpIA Survey ou do HarpIA Lab para a análise das mensagens de alerta enviadas pela GuardionAI à equipe de segurança da organização contratante. Nesse contexto, será especialmente relevante examinar a transparência, a clareza e a capacidade explicativa das mensagens produzidas pela solução.

- De 01 a 05 de junho de 2026: (i) em 1º de junho de 2026, foi realizada reunião entre as equipes da USP, da ANPD e da Metatext, com o objetivo de alinhar os procedimentos de testagem e avaliação; (ii) Na ocasião, foram definidos dois pontos de interação com o HarpIA Survey. O primeiro deles, já submetido a testes iniciais, será voltado à avaliação da qualidade da pseudonimização dos dados, por meio da realização de ataques adversariais. A proposta busca verificar possíveis vulnerabilidades relacionadas à identificação de dados; (iii) O segundo ponto de interação dependerá do desenvolvimento de uma API específica pela Metatext e terá como foco a análise da qualidade, clareza e transparência das informações encaminhadas pela GuardionAI à equipe de segurança da organização contratante da solução. Nesse eixo, a análise será direcionada ao contexto *business-to-business* (B2B), e não ao *business-to-consumer* (B2C), priorizando a avaliação da explicabilidade e da inteligibilidade do conteúdo das mensagens produzidas pelo sistema. Diferentemente do primeiro caso, não haverá realização de ataques adversariais, mas sim análise qualitativa de conteúdo; (iv) Adicionalmente, permanece em avaliação a possibilidade de realização de testes complementares por meio do HarpIA Lab; (v) Ademais, foram solicitados documentos complementares relacionados às políticas aprovadas e implementadas pela participante para disciplinar o funcionamento da inovação (Ficha A1), as quais já foram encaminhadas para análise.
- Por fim, em 31 de julho de 2026 foi feita a reunião para apresentação dos resultados do ciclo de testagem com a empresa Metatext, equipe do CIAAM/USP e Comissão do Sandbox da ANPD.

Em conclusão, a dinâmica de testagem da Metatext apresentou particularidades em relação às demais participantes, em razão de sua estrutura organizacional mais enxuta, da natureza abstrata da solução e de seu modelo de atuação B2B, direcionado à proteção de dados nas interações entre organizações, agentes de IA e modelos públicos. Essas características exigiram a readequação do cronograma, com menor número de ciclos e períodos mais longos de testagem, além de uma atuação mais próxima da equipe do CIAAM/USP na sistematização das informações e no preenchimento inicial das fichas, posteriormente validadas pela empresa.

Essas adaptações não comprometeram a metodologia do Sandbox. Pois foram preservadas as Fases A, B e C, a coleta estruturada de evidências, a validação das informações pela participante, a avaliação da maturidade tecnológica, a identificação dos Elementos Tecnológicos Críticos e a definição progressiva dos testes. A extensão da Fase A permitiu compreender adequadamente a arquitetura, o fluxo de dados e a finalidade da solução antes do início dos testes técnicos, evitando que a avaliação fosse baseada em premissas incompletas. As lacunas documentais e regulatórias identificadas foram registradas como pendências verificáveis para os ciclos seguintes, assegurando a rastreabilidade do processo.

A experiência da Metatext demonstra, portanto, que a metodologia possui flexibilidade suficiente para acomodar empresas com diferentes estruturas, níveis de maturidade e modelos de negócio, sem perda de rigor ou comparabilidade. O caso permitiu desenvolver uma forma adicional de avaliação aplicável a *startups* B2B de cibersegurança e infraestrutura de IA, na qual a equipe técnica e regulatória participa mais ativamente da reconstrução e validação da arquitetura declarada, da definição dos testes de pseudonimização e da análise das informações destinadas às equipes de segurança das organizações contratantes. Essa abordagem poderá servir como referência para futuros *sandboxes* que envolvam soluções semelhantes, especialmente aquelas cuja testagem exige maior período de compreensão técnica, integração específica com o ambiente experimental e adaptação dos instrumentos aos destinatários institucionais da tecnologia.

5. Conclusões gerais do Ciclo 2

O Ciclo 2 representou a primeira aplicação simultânea, com graus distintos de progressão da metodologia de testagem às soluções da Synapse, da Prevvine e da Metatext, respeitadas as particularidades técnicas, organizacionais e setoriais de cada participante. As atividades abrangeram a caracterização das soluções, a análise de sua arquitetura e maturidade tecnológica, a execução ou preparação dos testes, a coleta de evidências, a classificação de riscos, a avaliação das medidas de mitigação e a consolidação dos aprendizados regulatórios e empresariais. Os resultados demonstraram que a divisão dos ciclos nas Fases A, B e C é adequada para organizar progressivamente a produção de informações e transformar os achados técnicos em recomendações verificáveis.

A experiência também confirmou a importância de combinar instrumentos comuns de avaliação com procedimentos adaptados ao contexto de cada solução. Na Synapse, o ciclo permitiu avançar na análise de uma solução de mobilidade baseada em geolocalização, integrações externas e recomendação de rotas. Na Prevvine, os testes concentraram-se em uma aplicação de suporte à decisão médica, exigindo atenção reforçada à segurança clínica, à proteção de dados de saúde, à explicabilidade e à supervisão humana. Na Metatext, a natureza B2B da solução, sua atuação em cibersegurança e sua estrutura organizacional mais enxuta exigiram um período maior de compreensão da arquitetura, validação das fichas e preparação do ambiente de testes.

As diferenças observadas não comprometeram a consistência metodológica, ao contrário, demonstraram que a metodologia possui flexibilidade para acomodar empresas com modelos de negócio, níveis de maturidade e capacidades operacionais distintos, preservando os mesmos pilares de avaliação: funcionalidade, proteção de dados e segurança, maturidade tecnológica, realismo experimental, gestão de riscos, transparência, governança e aprendizagem regulatória. A adaptação de cronogramas e instrumentos mostrou-se legítima quando fundamentada, documentada e acompanhada pela equipe do CIAAM/USP e pela ANPD.

De modo geral, o Ciclo 2 evidenciou avanços técnicos e regulatórios, mas também demonstrou que a continuidade das soluções não pode basear-se exclusivamente em autodeclarações ou na existência formal de políticas. A metodologia exigiu a correspondência entre as informações prestadas, os documentos apresentados, os registros técnicos e os resultados dos testes. Essa triangulação permitiu distinguir controles efetivados de medidas em andamento, planejadas ou ainda não comprovadas, convertendo as lacunas identificadas em condicionantes e providências monitoráveis.

Para o Ciclo 3, recomenda-se priorizar:

- a conclusão dos testes não finalizados e a repetição daqueles afetados por dificuldades de integração ou alterações nas soluções;
- a apresentação das evidências documentais e técnicas pendentes;

- o aprofundamento dos testes adversariais, de pseudonimização, reidentificação, segurança e continuidade;
- a verificação das medidas de mitigação indicadas no Ciclo 2 e a atualização dos respectivos riscos residuais;
- o aprimoramento dos mecanismos de transparência e explicabilidade destinados a usuários, profissionais, gestores, equipes de TI e autoridades;
- a consolidação dos fluxos relacionados aos direitos dos titulares e à resposta a incidentes;
- o registro e o versionamento de modelos, parâmetros, interfaces, API, fluxos de dados e demais componentes submetidos aos testes;
- a atualização das matrizes TRL e dos Elementos Tecnológicos Críticos;
- o aprofundamento da análise das normas setoriais aplicáveis e das situações de sobreposição regulatória; e
- a manutenção das reuniões de devolutiva com as empresas e a ANPD para validação dos achados e definição dos ajustes subsequentes.

O Ciclo 3 deverá, portanto, assumir um caráter mais comprobatório e comparativo. Além de verificar a implementação das recomendações anteriores, deverá avaliar se as mudanças realizadas produziram redução efetiva dos riscos, maior maturidade tecnológica e melhoria da governança das soluções. A comparação entre os ciclos permitirá identificar a evolução de cada participante e gerar parâmetros de supervisão replicáveis pela ANPD em futuras iniciativas.

Conclui-se que o Ciclo 2 cumpriu sua função metodológica ao produzir evidências, revelar limitações, orientar melhorias e fortalecer o diálogo entre regulador, empresas e equipes técnicas. Os resultados obtidos oferecem uma base consistente para o Ciclo 3, que deverá avançar da identificação inicial dos riscos e das lacunas à comprovação das medidas implementadas e à avaliação de sua efetividade. Essa progressão é essencial para assegurar que o Sandbox funcione simultaneamente como ambiente de inovação responsável, mecanismo de proteção dos titulares e instrumento de aprendizagem regulatória.

