REVISTA TÓPICOS
ISSN 2965-6672 | Qualis A2

ALGORITMOS ADAPTATIVOS PARA DETECÇÃO DE ANOMALIAS EM AMBIENTES COMPUTACIONAIS DISTRIBUÍDOS: ESTRATÉGIAS DE APRENDIZADO FEDERADO, PRESERVAÇÃO DA PRIVACIDADE E REDUÇÃO DE FALSOS POSITIVOS

ADAPTIVE ALGORITHMS FOR ANOMALY DETECTION IN DISTRIBUTED COMPUTING ENVIRONMENTS: FEDERATED LEARNING, PRIVACY PRESERVATION, AND FALSE-POSITIVE REDUCTION STRATEGIES

Ciências Exatas e da Terra, Engenharias • 22/11/2026

REGISTRO DOI: 10.70773/revistatopicos/795175204

Márcio Rodrigo Elias Carvalho[1]

Natália José Pereira Rodrigues[2]

Karinne Rocha Queiroz Menezes[3]

Jailson da Silva dos Santos[4]

Yklys Falcão Rodrigues[5]

Frederico Augusto Siqueira Gentil[6]

Carlos Alberto Neris de Jesus[7]

RESUMO

A detecção de anomalias em ambientes computacionais distribuídos enfrenta heterogeneidade dos dados, mudanças do comportamento operacional, restrições de compartilhamento e elevado custo dos falsos positivos. Objetiva-se analisar criticamente a integração de algoritmos adaptativos, aprendizagem federada e mecanismos de preservação de privacidade para apoiar monitoramento confiável. Realizou-se revisão narrativa qualitativa de estudos fundamentais de detecção de anomalias, otimização federada, segurança de atualizações, avaliação de classificadores e normas de proteção de dados. A literatura aponta que manter registros nos participantes reduz a centralização, mas não garante confidencialidade das atualizações; também evidencia que modelos globais podem perder desempenho diante de distribuições não idênticas. Como síntese conceitual, propõe-se arquitetura em camadas que articula coleta mínima, aprendizagem local, coordenação federada, calibração contextual e auditoria humana. A redução de alarmes indevidos deve ser examinada em conjunto com sensibilidade, precisão, custo de triagem e robustez temporal, sem inferir ganhos experimentais não realizados. Conclui-se que a integração é promissora quando acompanhada de validação por participante, gestão de riscos e governança contínua.

Palavras-chave: Detecção de anomalias; Aprendizado federado; Privacidade de dados; Sistemas distribuídos; Falsos positivos.

ABSTRACT

Anomaly detection in distributed computing environments faces heterogeneous data, changing operational behavior, data-sharing restrictions, and high false-positive costs. This article critically examines the integration of adaptive algorithms, federated learning, and privacy-preserving mechanisms for dependable monitoring. A

qualitative narrative review was conducted using foundational studies on anomaly detection, federated optimization, update security, classifier evaluation, and data protection frameworks. The literature indicates that keeping records at participating sites reduces centralization but does not ensure confidentiality of model updates. Global models may also underperform when client distributions are markedly different. As a conceptual synthesis, a layered architecture is proposed, comprising data minimization, local learning, federated coordination, contextual threshold calibration, and human auditing. False-positive reduction should be assessed jointly with sensitivity, precision, analyst workload, and temporal robustness, without claiming experimental improvements that were not measured. Federated approaches are promising when supported by per-client validation, explicit risk management, and continuous governance.
Keywords: Anomaly detection; Federated learning; Data privacy; Distributed systems; False positives.

1. INTRODUÇÃO

A crescente descentralização da computação, caracterizada por aplicações em nuvem, borda, contêineres, dispositivos conectados e serviços distribuídos, alterou o alcance da observabilidade operacional. O comportamento esperado de uma aplicação deixa de corresponder a um único servidor e passa a resultar de interações entre componentes com cargas, configurações e ciclos de atualização distintos. Nesse contexto, uma anomalia representa um desvio estatístico ou comportamental que merece investigação, mas não equivale necessariamente a ataque ou falha. Mudanças legítimas de volume, implantação de novas versões e rotinas administrativas podem produzir sinais semelhantes aos de

incidentes. A literatura clássica demonstra que a validade de um método de detecção depende das hipóteses sobre normalidade e do domínio de aplicação (Chandola; Banerjee; Kumar, 2009).

O problema torna-se mais sensível quando os registros necessários à aprendizagem estão distribuídos entre organizações ou ambientes que não podem transferir dados brutos para uma base central. Informações operacionais podem incluir identificadores pessoais, trajetórias de usuários, nomes de dispositivos e padrões comerciais confidenciais. A aprendizagem federada propõe coordenar o treinamento entre participantes que preservam seus conjuntos locais, compartilhando atualizações de modelos em vez dos registros originais (McMahan et al., 2017). Entretanto, descentralizar os dados não elimina riscos de inferência, vazamento ou manipulação das atualizações, e a diversidade entre participantes pode prejudicar a generalização.

Este artigo examina como algoritmos adaptativos, treinamento federado e mecanismos complementares de privacidade podem contribuir para identificar desvios com menos alarmes improcedentes. O objetivo não é anunciar um detector experimental nem afirmar taxas de desempenho não medidas, mas organizar criticamente evidências e critérios de projeto. A contribuição consiste em propor uma arquitetura conceitual auditável, com validação temporal, limiares contextualizados e supervisão humana, reconhecendo que reduzir falsos positivos sem manter sensibilidade suficiente pode ocultar acontecimentos relevantes.

2. METODOLOGIA

Foi conduzida revisão narrativa crítica, de abordagem qualitativa, orientada à integração de literatura sobre detecção de anomalias, sistemas distribuídos, aprendizagem federada, proteção de atualizações e avaliação de classificadores desbalanceados. A seleção bibliográfica priorizou artigos fundamentais de periódicos e conferências, trabalhos metodológicos de acesso verificável, recomendações institucionais e legislação brasileira pertinente. Foram considerados estudos de referência sobre detecção de anomalias, FedAvg, FedProx, agregação segura, métricas de precisão e resposta a incidentes, além de documentos normativos do NIST e a Lei Geral de Proteção de Dados.

A análise foi organizada em eixos: definição do evento anômalo, heterogeneidade e mudança de distribuição; modelos de aprendizagem local e agregação; confidencialidade e ameaça à integridade; estratégias de calibração e avaliação; governança e implantação. Para cada eixo, distinguiram-se mecanismos propostos, demonstrações experimentais reportadas pelos autores originais e recomendações de natureza conceitual. A pesquisa não realizou coleta de logs, simulação, treinamento, testes de penetração ou metanálise. Por isso, nenhuma melhoria percentual apresentada em fonte externa deve ser atribuída à proposta aqui descrita.

Não foi empregado protocolo registrado de revisão sistemática, triagem independente por múltiplos revisores ou avaliação formal de risco de viés. Essa limitação afeta a possibilidade de exaustividade e impede caracterizar o trabalho como revisão sistemática. O uso da literatura tem finalidade analítica e de desenho de arquitetura, preservando a distinção entre inferência teórica e resultado empiricamente demonstrado. Estudos futuros deverão transformar as recomendações em hipóteses verificáveis mediante ambientes

controlados, conjuntos de teste descritos e procedimentos reprodutíveis.

3. FUNDAMENTOS DA DETECÇÃO DE ANOMALIAS

Detecção de anomalias envolve identificar observações, sequências ou relações que divergem de padrões estabelecidos. As técnicas podem empregar regras, estatística robusta, agrupamento, reconstrução, densidade, previsão ou classificação de uma classe. A escolha depende de disponibilidade de rótulos, frequência dos eventos críticos, estrutura temporal e explicabilidade esperada. Um detector por reconstrução, por exemplo, sinaliza amostras que um modelo não representa bem; isso não prova que sejam maliciosas. Ruff et al. (2021) organizam métodos rasos e profundos sob hipóteses comparáveis e demonstram a necessidade de esclarecer o conceito de normalidade adotado.

Em sistemas distribuídos, a unidade de análise não precisa ser uma linha de log. Pode ser um episódio que reúne aumento de latência, falhas de autenticação, alteração de topologia e variação de consumo de recursos. A correlação multivariada ajuda a distinguir um pico de tráfego esperado de um conjunto de eventos incompatível com o histórico. Contudo, a agregação exagerada perde sinais localizados, enquanto granularidade excessiva amplia ruído e custo de processamento. É necessário definir janelas temporais, fontes de observação e regras de deduplicação antes do treinamento.

Outra distinção essencial envolve novidade e gravidade. Uma alteração sem precedentes pode representar inovação legítima de processo, enquanto um incidente repetitivo pode parecer

estatisticamente comum. Portanto, a pontuação anômala não deve determinar automaticamente prioridade máxima. O modelo técnico precisa dialogar com contexto de ativos, criticidade de processos e verificações posteriores. Em ambiente de segurança, anomalia é uma evidência para investigação, e a classificação de incidente exige elementos adicionais.

4. TOPOLOGIA DOS AMBIENTES DISTRIBUÍDOS

Aplicações modernas atravessam diferentes domínios administrativos: redes corporativas, provedores externos, serviços de nuvem, equipamentos industriais e dispositivos na borda. Cada segmento possui latência, disponibilidade, formatos e limitações de retenção próprios. O monitoramento de uma topologia assim não pode depender exclusivamente de uma captura central de pacotes. Logs de aplicação, métricas de infraestrutura, rastreamentos distribuídos e eventos de identidade oferecem perspectivas complementares, embora apresentem lacunas e relógios nem sempre sincronizados.

A arquitetura de observabilidade deve identificar quais componentes produzem sinais confiáveis e quais dados são desnecessários ao objetivo de detecção. A coleta seletiva reduz armazenamento e exposição indevida. Por outro lado, coletar apenas eventos já classificados por regras antigas cria circularidade: o algoritmo aprenderá o que o sistema anterior considerou relevante e pode ignorar padrões desconhecidos. A combinação de amostragem, telemetria contextual e retenção proporcional é mais defensável que o acúmulo indiscriminado de dados.

Falhas de comunicação, reinicializações e políticas de escalonamento automático afetam simultaneamente os sinais e os próprios mecanismos de coleta. Por isso, a ausência de telemetria deve ser representada explicitamente e não confundida com ausência de anomalias. Um modelo federado precisa conhecer a qualidade de cada participante, registrar intervalos de indisponibilidade e evitar que participantes com observação incompleta determinem o comportamento esperado para toda a rede.

5. HETEROGENEIDADE ESTATÍSTICA ENTRE PARTICIPANTES

O aprendizado federado distingue-se do treinamento central porque os dados locais frequentemente não seguem distribuições idênticas. Um hospital, um campus universitário e um centro de processamento podem ter combinações radicalmente diferentes de horários, serviços e protocolos. Mesmo dentro da mesma organização, grupos de servidores variam conforme função, localização e carga. Atualizações enviadas por esses clientes podem apontar em direções conflitantes, reduzindo a estabilidade de um modelo global único. Li et al. (2020) enfatizam heterogeneidade estatística e de sistemas como dificuldades estruturais da otimização federada.

O desbalanceamento também ocorre no número de registros disponíveis. Um grande participante tende a influenciar agregações ponderadas pelo tamanho de amostra, embora seu padrão possa ser pouco representativo dos demais. Já atribuir pesos iguais pode ampliar a contribuição de conjuntos pequenos e ruidosos. A escolha da ponderação precisa estar vinculada ao objetivo: desempenho global médio, garantia de qualidade em pequenos clientes ou

proteção de participantes críticos. Nenhuma dessas métricas é neutra em termos organizacionais.

Uma resposta técnica envolve agrupamento por afinidade operacional, personalização das últimas camadas do detector e avaliação por cliente. O modelo compartilhado pode aprender representações comuns, enquanto decisões locais de alerta consideram a distribuição de cada segmento. Esse arranjo evita impor um limiar global que poderia ser excessivamente permissivo em uma unidade e exageradamente sensível em outra. Sua eficácia, entretanto, deve ser testada, não presumida.

6. HETEROGENEIDADE COMPUTACIONAL E PARTICIPAÇÃO PARCIAL

Os participantes de uma federação não dispõem de capacidade homogênea de CPU, memória, energia ou conectividade. Equipamentos de borda podem trabalhar com orçamento de processamento reduzido e longos períodos de desconexão, enquanto centros de dados executam ciclos locais mais extensos. A sincronização estrita aumenta o tempo de espera pelos clientes lentos e favorece exclusão involuntária de ambientes menos equipados. Estratégias de participação parcial devem ponderar custo de comunicação, representatividade e idade das atualizações.

Uma atualização muito antiga pode refletir uma topologia já desativada. Por isso, modelos assíncronos precisam registrar versão, período de treinamento e compatibilidade de atributos antes da agregação. A tolerância à interrupção deve ser acompanhada de critérios para prevenir incorporação repetida de gradientes obsoletos. O FedProx foi proposto para tornar o processo mais

robusto diante de variabilidade de trabalho local e de distribuição dos dados (LI et al., 2020), mas sua aplicação à detecção de anomalias requer adequação do objetivo e validação específica.

Em termos operacionais, o custo não se resume ao tráfego de parâmetros. Há consumo de energia, ciclos de inferência, armazenamento de versões, rotinas de recuperação e carga de revisão de alertas. Uma arquitetura teoricamente privada que paralise dispositivos ou produza atrasos incompatíveis com a resposta operacional pode falhar em seus objetivos de segurança. A análise de viabilidade precisa contemplar infraestrutura, pessoal e continuidade do serviço.

7. APRENDIZAGEM FEDERADA E AGREGAÇÃO DE MODELOS

O algoritmo FedAvg, descrito por McMahan et al. (2017), coordena rodadas em que clientes treinam cópias locais e enviam atualizações para composição de um modelo compartilhado. O procedimento reduz a necessidade de centralizar dados brutos, mas não define automaticamente a qualidade dos rótulos nem resolve diferenças semânticas entre ambientes. Se cada cliente interpreta de modo distinto uma falha de autenticação, a agregação de parâmetros pode combinar objetivos incompatíveis.

A construção da federação deve começar por um acordo sobre o espaço de atributos, unidades de medida, tratamento de valores ausentes e definição de evento. Em seguida, devem ser definidas as regras de seleção de clientes, limites de atualização, validação das versões e mecanismos de rollback. Uma atualização de modelo é um artefato técnico sensível: deve ser autenticada, vinculada a um cliente autorizado e examinada antes de integrar a versão global.

Convém diferenciar compartilhamento de representação e compartilhamento de decisão. A rede pode distribuir parâmetros de um extrator de características e manter classificadores finais locais; alternativamente, pode agregar modelos completos. O primeiro caminho tende a preservar particularidades operacionais, enquanto o segundo facilita administração central. A seleção depende da heterogeneidade, da necessidade de explicação e do risco de contaminação de parâmetros. Em ambos, o sucesso exige medir resultados separados por participante.

8. MODELOS ADAPTATIVOS E MUDANÇA DE CONCEITO

A adaptatividade responde ao fato de que o comportamento normal de um sistema varia com novos serviços, alterações de acesso, sazonalidade e crescimento de tráfego. O chamado concept drift ocorre quando a relação entre atributos e classes muda, enquanto mudanças de distribuição podem ocorrer mesmo sem alteração de significado. Um detector treinado sobre um período estável pode produzir grande quantidade de alertas após uma migração legítima de infraestrutura. Atualizar o modelo indiscriminadamente, porém, pode absorver comportamento anômalo persistente como novo padrão.

Um mecanismo responsável de adaptação deve combinar detecção de deriva, revisão de alterações planejadas e validação contra eventos anteriormente conhecidos. Janelas móveis, pesos temporais e modelos de referência congelados podem fornecer sinais complementares. Quando uma mudança abrupta coincide com implantação documentada, é possível recalibrar temporariamente o limiar e monitorar sua estabilização. Se a causa é desconhecida, a

mesma alteração deve desencadear investigação antes de alimentar o treinamento.

Esse equilíbrio é relevante para evitar o paradoxo da aprendizagem contínua: quanto mais rapidamente o sistema aprende, maior o risco de normalizar condutas que deveria detectar. Uma política de atualização precisa estabelecer quem autoriza incorporação de novos padrões, como preservar versões anteriores e qual evidência de estabilidade é necessária. Adaptatividade não deve ser confundida com automação irrestrita.

9. DETECÇÃO NÃO SUPERVISIONADA E SEMISSUPERVISIONADA

Em ambientes operacionais, rótulos de incidentes são escassos e frequentemente incompletos. Técnicas não supervisionadas procuram regiões pouco prováveis no espaço de características, enquanto métodos semissupervisionados aprendem com um conjunto supostamente normal. Isolamento, estimativa de densidade, autoencoders e previsão de séries temporais oferecem estratégias diferentes, com sensibilidades próprias a escala e ruído. Ruff et al. (2021) mostram que métodos profundos não substituem automaticamente abordagens clássicas, sobretudo quando o volume de dados e a capacidade de validação são limitados.

O problema dos rótulos negativos é especialmente importante. Um registro sem confirmação de incidente não constitui evidência plena de benignidade. Se o conjunto de treinamento inclui ataques não detectados, o modelo pode incorporá-los à distribuição normal. Alternativamente, selecionar apenas períodos muito controlados produz um detector pouco tolerante às variações usuais. A

documentação da origem dos dados e dos critérios de inclusão é requisito indispensável para interpretar métricas posteriores.

Sempre que houver especialistas disponíveis, a aprendizagem ativa pode priorizar exemplos ambíguos para revisão humana. O objetivo é concentrar esforço em casos informativos, sem transformar todas as decisões do analista em verdade absoluta. Concordância entre revisores, registro de incerteza e recategorização posterior são procedimentos que melhoram a qualidade do conjunto rotulado e reduzem a cristalização de erros históricos.

10. DETECÇÃO TEMPORAL E CORRELAÇÃO DE EVENTOS

Alguns desvios só se tornam inteligíveis quando observados como sequência. Aumento de chamadas a serviço, repetidas negativas de autenticação e uma mudança de configuração podem ser aceitáveis isoladamente, mas adquirir significado quando ocorrem no mesmo intervalo e sobre o mesmo ativo. Modelos de série temporal, janelas deslizantes e grafos de dependência ajudam a preservar relações que seriam perdidas em classificadores baseados em amostras independentes.

O desafio consiste em escolher a extensão das janelas. Períodos curtos são adequados a picos súbitos e tendem a gerar mais ruído; períodos longos captam tendências graduais, mas podem atrasar a resposta e diluir sinais discretos. O desenho precisa relacionar janela, horizonte de previsão e gravidade do fenômeno monitorado. Operações de manutenção previamente programadas devem ser tratadas como contexto, e não simplesmente removidas do conjunto de observações.

Em uma federação, os relógios dos clientes e a ordem de chegada das atualizações podem divergir. Agregar características temporais exige normalização do fuso, tratamento de atrasos e identificação de eventos duplicados. Caso contrário, o modelo pode aprender uma falsa sequência causal criada por atrasos de transporte. Uma trilha de auditoria que preserve momento de ocorrência e momento de recebimento é valiosa para distinguir fenômeno observado de artefato da coleta.

11. LIMIARES ADAPTATIVOS E CALIBRAÇÃO LOCAL

A pontuação de anomalia precisa ser convertida em uma regra de decisão. Um único limiar aplicado indistintamente a todos os ambientes ignora diferenças de carga e perfil operacional. Limiar baixo favorece sensibilidade, mas aumenta alarmes; limiar alto reduz trabalho de triagem à custa de possíveis omissões. Em contextos críticos, o objetivo não deve ser minimizar falsos positivos isoladamente, mas controlar um equilíbrio explícito entre perdas decorrentes de alarmes e perdas associadas a incidentes não detectados.

A calibração pode utilizar quantis da distribuição local, validação sobre períodos conhecidos e restrições operacionais de volume de alertas. A decisão deve ser renovada após mudanças estruturais, mantendo uma linha de base comparável. Métodos de calibração que assumem estabilidade ou intercambialidade dos dados precisam ter seus pressupostos explicitados: a presença de deriva severa pode invalidar garantias obtidas em amostras anteriores.

É recomendável adotar níveis graduais: observação de baixa prioridade, investigação assistida e alerta crítico baseado em

múltiplos sinais. Essa hierarquia permite que pequenas divergências permaneçam registradas sem interromper fluxos de trabalho. Também aumenta a capacidade de avaliar impacto das escolhas de limiar, pois registra decisões intermediárias e suas justificativas. Uma redução aparente de falsos positivos não deve resultar apenas do descarte silencioso de alertas antes auditados.

12. DESBALANCEAMENTO E MÉTRICAS DE DESEMPENHO

Eventos de interesse costumam ser raros em comparação aos registros rotineiros. Nesse cenário, a acurácia global pode parecer elevada mesmo quando o sistema ignora quase todos os incidentes. Precisão, revocação, especificidade, taxa de falsos positivos e área sob a curva precisão–revocação descrevem aspectos diferentes do desempenho. Saito e Rehmsmeier (2015) demonstram a importância da curva precisão–revocação quando a classe positiva é minoritária, pois ela torna explícita a proporção de alertas efetivamente relevantes.

A avaliação também deve contabilizar o volume absoluto de trabalho. Uma taxa aparentemente pequena de falsos positivos pode produzir milhares de ocorrências quando o sistema processa milhões de eventos. É útil registrar falsos alarmes por serviço ou por dia, tempo médio de triagem e proporção de alertas agrupados como episódio único. Uma mesma ocorrência distribuída por dezenas de componentes não deve ser contabilizada como dezenas de incidentes independentes sem justificativa.

Comparações de algoritmos exigem uma mesma definição de evento positivo e intervalos de teste equivalentes. Mudanças na prevalência dos incidentes alteram a precisão mesmo que

sensibilidade e especificidade permaneçam estáveis. Por isso, resultados de benchmarks sintéticos ou laboratórios controlados não podem ser transferidos diretamente para produção. Deve-se apresentar a matriz de confusão, distribuição das classes e incerteza estatística, além dos indicadores agregados.

13. AVALIAÇÃO TEMPORAL E PREVENÇÃO DE VAZAMENTO DE DADOS

Dividir aleatoriamente registros de uma mesma sequência entre treinamento e teste pode produzir estimativas artificiais de desempenho. Logs contíguos compartilham configurações, identificadores e padrões que tornam o teste excessivamente semelhante ao treino. Em aplicações temporais, convém separar períodos cronologicamente, utilizar intervalos de afastamento quando necessário e registrar a data em que cada rótulo se tornou disponível. A avaliação deve simular a situação real em que o modelo é treinado sobre o passado e opera sobre dados futuros.

Em federações, há uma segunda fonte de vazamento: o mesmo incidente pode atingir diversos participantes e aparecer simultaneamente nos conjuntos de treinamento e teste. A divisão por organização, campanha operacional ou episódio reduz dependências ocultas. Para medir generalização a clientes não vistos, é necessário reservar participantes inteiros para avaliação externa, e não apenas amostras aleatórias de cada cliente.

Alterações de atributos feitas com conhecimento do conjunto de teste também comprometem validade. Escalonamento, seleção de variáveis e calibração devem ser ajustados somente com dados autorizados do treino e validação. Documentar essas fronteiras é tão

importante quanto a arquitetura algorítmica, pois modelos complexos podem aparentar excelente desempenho por falhas básicas de desenho experimental.

14. PRESERVAÇÃO DE PRIVACIDADE: ALCANCE E LIMITES

Manter dados brutos nos participantes reduz um caminho relevante de exposição, mas não assegura anonimato das atualizações. Gradientes e parâmetros podem conter informação sobre exemplos de treinamento, especialmente quando a quantidade de registros é pequena ou o modelo possui alta capacidade. A privacidade federada precisa ser definida mediante um modelo de ameaça: quais entidades são confiáveis, que informações o coordenador observa e que riscos surgem de participantes desonestos ou de comprometimento de dispositivos.

A primeira camada consiste em minimização de atributos, autenticação dos participantes e criptografia das comunicações. A segunda inclui restrições de acesso a atualizações, políticas de retenção e eventual uso de agregação segura ou privacidade diferencial. Esses mecanismos respondem a riscos diferentes e podem ter custos distintos de comunicação e desempenho. A eliminação do registro bruto no coordenador não torna o sistema imune a inferência, nem dispensa análise jurídica do tratamento.

A transparência com as organizações participantes deve descrever finalidades, campos coletados, versões de modelos e possíveis repercussões de uma atualização comprometida. O desenho de privacidade não deve ser reduzido a uma frase promocional sobre descentralização. Exige governança demonstrável e avaliação

contínua das superfícies de exposição criadas pelo próprio processo de aprendizagem.

15. AGREGAÇÃO SEGURA E CONTROLE DE ACESSO

A agregação segura procura permitir ao coordenador calcular uma combinação das contribuições locais sem inspecionar cada atualização individual em texto claro. Bonawitz et al. (2017) propuseram um protocolo tolerante a desistências de participantes, demonstrando a viabilidade de agregação protegida em determinados cenários. O mecanismo limita a observação individual, mas não garante que todos os participantes enviem atualizações corretas ou que o resultado agregado seja imune a manipulações.

O uso operacional exige definir tamanho mínimo de grupos, tratamento de clientes desconectados, rotação de chaves e proteção das credenciais de autenticação. Federações com poucos clientes podem ter risco residual diferente de sistemas com milhares de dispositivos. Também é necessário coordenar a agregação com mecanismos de detecção de atualizações anômalas, pois algumas formas de inspeção individual tornam-se mais difíceis quando o servidor não observa contribuições separadas.

A adoção de criptografia ou agregação segura deve vir acompanhada de análise de custo. Restrições de memória, largura de banda e disponibilidade de participantes podem aumentar a latência de cada rodada. A proteção efetiva depende do conjunto de controles: autenticação, integridade do software cliente, política de atualização e auditoria de resultados. Um protocolo criptograficamente correto não substitui gestão de identidade e supervisão do ambiente.

16. PRIVACIDADE DIFERENCIAL E UTILIDADE ANALÍTICA

A privacidade diferencial oferece um arcabouço matemático para limitar quanto a participação de um registro ou usuário pode modificar uma saída observável. Em aprendizado de máquina, a aplicação pode envolver recorte de contribuições e introdução controlada de ruído. A garantia, quando formalmente implementada, depende da unidade protegida, dos parâmetros de privacidade, do mecanismo empregado e da composição de múltiplas rodadas. Não basta declarar que foi adicionado ruído para afirmar existência de proteção adequada.

Em detecção de anomalias, o compromisso entre utilidade e privacidade merece atenção especial porque os eventos críticos são raros. Ruído excessivo pode apagar justamente os padrões minoritários relevantes, prejudicando a capacidade de identificação. Por outro lado, ruído insuficiente pode manter informação sensível nas atualizações. A análise deve comparar resultados por classe e participante, documentar orçamento acumulado de privacidade e explicitar se a proteção se aplica a registros, dispositivos ou usuários.

A decisão organizacional envolve reconhecer riscos residuais, finalidade legítima de tratamento e necessidade proporcional de informação. Privacidade diferencial não torna qualquer conjunto de dados automaticamente adequado para toda finalidade. O mecanismo atua em dimensões específicas da exposição, enquanto retenção, controle de acesso e resposta a incidentes continuam necessários.

17. ROBUSTEZ CONTRA ATUALIZAÇÕES INCONSISTENTES

Uma federação pode receber atualizações de clientes com falhas de configuração, dados corrompidos ou comportamento adversarial. Mesmo sem intenção maliciosa, diferenças extremas de escala e distribuição podem deslocar o modelo global para uma região pouco útil aos demais. São necessários critérios de integridade, limitação de magnitude e monitoramento da influência de cada rodada, compatíveis com o modelo de ameaça e com a proteção de privacidade adotada.

Agregações robustas por estatísticas alternativas e ponderações de confiança podem reduzir a influência de contribuições extremas em alguns cenários, mas também descartar atualizações legítimas de participantes minoritários. A escolha não deve presumir que diferença equivale a ataque. Um pequeno cliente com padrão operacional raro pode ser precisamente aquele que o sistema precisa proteger. Validação por cliente e avaliação de justiça de desempenho ajudam a reconhecer esses efeitos.

Um processo prudente mantém versões estáveis e critérios de rollback quando a versão recém-agregada provoca deterioração observável de precisão ou sensibilidade. As alterações devem ser registradas com trilha de proveniência, sem expor informações além do necessário. Antes de combinar medidas de robustez com agregação segura, é preciso verificar compatibilidade técnica: invisibilidade das atualizações individuais pode limitar testes de plausibilidade que dependem de inspeção granular.

18. REDUÇÃO DE FALSOS POSITIVOS POR CONTEXTO

Os falsos positivos não decorrem somente de limitações estatísticas. Eles também resultam de interpretações sem contexto: variações

previstas em horários de fechamento, testes de carga, atualizações automatizadas e rotinas de backup podem disparar detectores que ignoram calendário e função do ativo. Integrar metadados de mudança autorizada, dependências de serviços e criticidade operacional ajuda a diferenciar atividade inesperada de comportamento apenas incomum.

Uma estratégia promissora utiliza múltiplos estágios. Um modelo identifica desvios candidatos, um componente contextual avalia coerência com a operação e uma camada final considera impacto potencial e prioridade de investigação. Essa sequência pode reduzir a quantidade de alertas apresentados ao analista, mas seus filtros precisam ser avaliados quanto a omissões. Uma operação marcada como manutenção não deve excluir automaticamente todo sinal, pois mudanças autorizadas podem coincidir com incidentes ou erros de configuração.

A deduplicação por episódio permite relacionar alertas que compartilham ativo, janela temporal e dependência de serviço. A redução no volume de notificações deve ser acompanhada da preservação de rastreabilidade dos sinais originais. Caso contrário, o sistema pode parecer mais preciso por ocultar evidências, dificultando investigações retrospectivas. A qualidade da triagem deve ser medida pela utilidade do agrupamento, não apenas por um contador menor de alarmes.

19. APRENDIZAGEM COM RETORNO DOS ANALISTAS

Analistas produzem conhecimento operacional ao investigar alertas e classificar eventos como pertinentes, benignos ou inconclusivos. Esse retorno pode alimentar processos de recalibração, mas precisa

ser tratado como evidência sujeita a erro e desacordo. A capacidade do profissional, o volume de trabalho e a disponibilidade de registros variam; uma decisão rápida diante de pressão operacional pode não representar uma confirmação técnica definitiva.

Uma proposta de aprendizagem assistida deve registrar justificativa da decisão, grau de confiança, evidências consultadas e eventual revisão posterior. Alertas inconclusivos não devem ser automaticamente rotulados como negativos. Também é útil monitorar divergências entre equipes e mudanças nos critérios de classificação, evitando que o modelo incorpore preferências transitórias como se fossem fatos permanentes.

O feedback pode alimentar tanto o treinamento local quanto ajustes de limiar, mas a natureza da alteração deve ser explicitada. Se a equipe identifica uma rotina legítima recorrente, uma regra contextual pode resolver o problema com menor custo do que retreinar toda a rede. Se a alteração envolve um padrão de comportamento realmente novo, uma atualização de representação pode ser necessária. A seleção do mecanismo adequado depende da causa do erro.

20. EXPLICABILIDADE DOS ALERTAS E CONFIANÇA OPERACIONAL

A confiança de uma equipe depende da possibilidade de compreender por que determinado evento foi priorizado. Alguns detectores oferecem atributos diretamente interpretáveis; outros, sobretudo modelos profundos, produzem escores cuja origem não é transparente. Técnicas de atribuição podem indicar variáveis influentes, mas não devem ser tratadas como provas causais. Uma

explicação útil precisa conservar relação com os registros, as versões e o contexto disponíveis na investigação.

Em sistemas distribuídos, uma justificativa pode combinar desvio temporal, serviço afetado, comparação com histórico local e relação com eventos dependentes. É recomendável apresentar o grau de incerteza e os limites do que o modelo inferiu. A categoria “anomalia em autenticação”, por exemplo, é informativa como hipótese de triagem, mas não determina identidade de agente nem natureza de eventual infração.

Auditorias devem examinar se explicações permanecem consistentes após atualização federada. Alterações profundas de representação podem modificar os atributos destacados, mesmo quando a pontuação final é semelhante. Essa instabilidade precisa ser documentada para evitar excesso de confiança. Uma boa interface auxilia decisão humana e preserva possibilidade de contestação e revisão, especialmente quando alertas podem provocar suspensão de serviços ou investigação de pessoas.

21. ARQUITETURA CONCEITUAL EM CAMADAS

Como síntese da literatura, propõe-se uma arquitetura conceitual de cinco camadas, sem alegação de implementação experimental. A primeira estabelece observabilidade local com coleta mínima, sincronização temporal e documentação dos atributos. A segunda executa pré-processamento e estimação de desvios no cliente, preservando padrões operacionais específicos. A terceira coordena treinamento federado com autenticação, verificação de versões e proteções apropriadas das atualizações. A quarta calibra alertas

conforme contexto e risco; a quinta registra avaliação humana, métricas, respostas e aprendizado institucional.

A separação das camadas permite atualizar componentes sem transformar toda alteração em reconfiguração global. Um novo coletor pode exigir revisão de esquema antes de alimentar a aprendizagem, enquanto um ajuste de prioridade de alertas pode ser realizado sem alterar o modelo de representação. Essa modularidade facilita testes de regressão e facilita identificar a origem de uma piora: coleta, distribuição, algoritmo, limiar ou fluxo de resposta.

A proposta assume que organizações participantes dispõem de governança mínima de identidades, infraestrutura e comunicação. Não constitui protocolo universal para redes industriais, sistemas clínicos ou ambientes militares, cujos riscos e tolerâncias podem ser distintos. Sua utilidade reside em tornar explicitáveis decisões arquiteturais, requisitos de teste e pontos de responsabilidade, e não em oferecer garantias automáticas de desempenho.

22. PLANO DE VALIDAÇÃO EXPERIMENTAL

Uma avaliação futura deve começar por hipóteses falsificáveis. Por exemplo, pode-se testar se calibração local reduz alertas improcedentes por unidade de tempo sem deterioração relevante da revocação; se treinamento federado melhora generalização entre clientes com dados escassos; ou se agregação segura altera significativamente custo de comunicação. Cada hipótese deve possuir desfecho primário definido antes da execução e método de comparação adequado.

O desenho deve contemplar pelo menos uma referência centralizada quando juridicamente e tecnicamente possível, uma linha de base local sem federação, um modelo federado simples e variantes com personalização ou calibração. Conjuntos de teste devem representar períodos futuros, participantes diferentes e ocorrências de diferentes intensidades. Os resultados devem ser apresentados por cliente e agregados com pesos transparentes, evitando que desempenho de um participante volumoso mas fácil esconda falhas em ambientes críticos.

Intervalos de confiança e análise de sensibilidade a prevalência são necessários para interpretar ganhos. Quando a classe positiva é rara, uma diferença de poucos incidentes pode alterar fortemente a precisão. É indispensável identificar os eventos utilizados, o processo de rotulagem, a disponibilidade de dados e os critérios de exclusão. Sem esse detalhamento, um resultado numérico não poderá ser reproduzido nem transferido de maneira responsável.

23. VALIDAÇÃO SOB DERIVA E EVENTOS RAROS

Um detector que funciona em laboratório pode deteriorar-se após mudança de infraestrutura ou campanha operacional atípica. A validação deve incluir períodos com implantação de novas versões, escalonamento automático, falhas de rede e crescimento sazonal. É importante distinguir adaptação desejável a mudanças benignas de perda de sensibilidade a eventos críticos. O acompanhamento longitudinal permite observar se versões sucessivas mantêm memória suficiente de padrões raros.

Cenários raros não devem ser avaliados somente pela média global. É preferível relatar cobertura por categoria, tempo até o primeiro

alerta e capacidade de identificar episódios de baixa frequência. O uso de registros sintéticos pode ampliar casos de teste, mas não substitui observações reais: geradores podem produzir artefatos que o classificador aprende sem representar o fenômeno operacional. Quando se utiliza simulação, suas regras e limitações precisam ser explicitadas.

A comparação também deve examinar o custo de adaptação: número de rodadas federadas, quantidade de bytes transmitidos, duração das atualizações e disponibilidade dos clientes. Um método que melhora uma métrica marginalmente, mas amplia muito a carga de computação, pode ter baixa viabilidade. A decisão de implantação exige considerar benefício operacional líquido e riscos de degradação, não apenas uma curva de desempenho produzida em condições favoráveis.

24. GOVERNANÇA, RESPONSABILIDADES E AUDITORIA

Um sistema federado atravessa fronteiras administrativas e precisa estabelecer responsabilidades antes do compartilhamento de qualquer atualização. A governança deve determinar finalidades, papéis do coordenador e participantes, autorização de modelos, níveis de acesso, prazos de retenção e procedimentos para suspensão de clientes. Um contrato técnico ou acordo institucional precisa esclarecer se atualizações constituem informação confidencial e quem pode realizar auditorias de desempenho ou privacidade.

O NIST recomenda integrar detecção e resposta aos processos de gestão de risco cibernético. A SP 800-61 revisão 3, publicada em 2025, relaciona resposta a incidentes às funções do Cybersecurity

Framework 2.0 (Nelson et al., 2025). No contexto aqui analisado, esse princípio significa que um detector não deve operar isolado de responsáveis por investigação, contenção, recuperação e aprendizagem. Indicadores de desempenho precisam dialogar com capacidade de resposta efetiva.

A trilha de auditoria deve conservar versões de parâmetros, políticas de limiar, justificativas de atualização, resultados de validação e incidentes de desempenho, respeitando minimização de dados. Também deve existir procedimento para retirar versões defeituosas e corrigir alertas que tenham produzido consequências indevidas. A responsabilidade institucional não pode ser transferida integralmente a um modelo, por mais sofisticado que seja.

25. LEI GERAL DE PROTEÇÃO DE DADOS E MINIMIZAÇÃO

No contexto brasileiro, a Lei nº 13.709/2018 estabelece princípios de finalidade, adequação, necessidade, transparência, segurança e responsabilização no tratamento de dados pessoais. Registros técnicos podem conter dados pessoais mesmo quando produzidos para fins operacionais. Endereços associados a indivíduos, identificadores persistentes e sequências de acesso podem permitir identificação direta ou indireta. Por isso, um sistema de monitoramento não está dispensado de análise de base legal e de necessidade simplesmente por ter finalidade de segurança.

O artigo 46 da LGPD estabelece o dever de adoção de medidas técnicas e administrativas para proteção dos dados. No desenho federado, medidas pertinentes podem incluir redução de identificadores, segregação de acesso, autenticação forte, proteção do tráfego e restrições à retenção de atualizações. A escolha precisa

ser proporcional aos riscos concretos e ao tipo de informação. Aprendizagem descentralizada constitui um instrumento possível de minimização, mas não substitui conformidade jurídica nem avaliação do ciclo de vida dos dados.

Em ambientes que envolvem múltiplas organizações, a governança deve descrever os fluxos informacionais em linguagem compreensível aos responsáveis. Devem ser examinadas transferências internacionais, prestação de serviços por terceiros e finalidade secundária das métricas. Um mecanismo destinado a identificar falhas de infraestrutura não deve ser redirecionado automaticamente para vigilância individual sem fundamentação específica e controles adequados.

26. SEGURANÇA DE IMPLANTAÇÃO E CONTINUIDADE

O processo de implantação constitui fase de risco porque introduz código de coleta e inferência em sistemas que já possuem exigências próprias de disponibilidade. A execução local deve ter limites de CPU e memória, monitoramento de falhas e capacidade de desativação segura. Um agente que comprometa o desempenho de um serviço crítico pode criar risco superior ao evento que procura identificar. Procedimentos de atualização, assinatura de pacotes e rollback fazem parte da arquitetura de segurança.

A coordenação federada depende de canais de comunicação confiáveis e de controle de versões. Se um cliente operar com esquema de atributos incompatível, a atualização pode deteriorar o modelo mesmo sem intenção adversarial. Testes de compatibilidade e protocolos explícitos de mudança são essenciais, sobretudo quando há múltiplos fornecedores e ambientes legados.

Observabilidade do próprio mecanismo de detecção deve permitir identificar falhas de coleta, atraso de sincronização e inconsistências de estado.

Durante indisponibilidades da federação, o cliente deve preservar capacidade local básica de alerta, com regras documentadas para operar em modo degradado. O mecanismo de recuperação não pode depender de disponibilidade permanente do coordenador, especialmente em situações que coincidam com incidentes de infraestrutura. A continuidade exige planejar armazenamento de eventos essenciais, sincronização posterior e prevenção de duplicação de alarmes no retorno da comunicação.

27. CUSTO OPERACIONAL E SUSTENTABILIDADE

A avaliação econômica de uma plataforma de detecção não deve concentrar-se apenas em licenças ou custos computacionais de treinamento. Alarmes improcedentes consomem tempo de profissionais, podem gerar interrupções desnecessárias e desgastar confiança no monitoramento. Em contrapartida, redução agressiva de alertas pode aumentar custos de incidentes não detectados. É necessário construir uma função de custo que contemple trabalho de triagem, risco residual, manutenção, infraestrutura e impacto sobre continuidade.

A aprendizagem federada acrescenta despesas de coordenação, comunicação e gestão de identidades, mas pode reduzir custos associados à centralização de grandes volumes de registros. A direção do balanço depende do tamanho dos modelos, frequência de atualização, infraestrutura de rede e exigências regulatórias. Para alguns participantes, uma arquitetura híbrida com modelos locais e

atualização eventual pode ser mais adequada do que treinamento frequente de alta complexidade.

A sustentabilidade inclui capacidade de manutenção por equipes técnicas. Um sistema só é reprodutível quando versões, dependências, critérios de avaliação e procedimentos de calibração estão documentados. A dependência de especialistas muito específicos pode limitar adoção, mesmo quando o método mostra potencial científico. O planejamento deve prever capacitação, auditorias recorrentes e evolução controlada da infraestrutura.

28. APLICAÇÃO EM INSTITUIÇÕES DE ENSINO E PESQUISA

Universidades reúnem laboratórios, serviços administrativos, redes acadêmicas, ambientes de ensino remoto e equipamentos de pesquisa, frequentemente administrados por equipes com capacidades distintas. O comportamento legítimo varia entre horários de aula, períodos de matrícula, avaliações e atividades científicas de longa duração. Modelos centralizados podem confundir essas diferenças com anomalias; uma federação entre unidades preservaria parte do contexto local e permitiria compartilhar representações de padrões recorrentes.

A adoção, porém, deve respeitar autonomia administrativa e proteção de dados de estudantes, docentes e técnicos. O monitoramento precisa definir quais informações são necessárias à segurança de sistemas e quais práticas poderiam gerar vigilância desproporcional. Perfis acadêmicos atípicos, como transferências volumosas em pesquisa científica, não podem ser automaticamente classificados como comportamento suspeito. Políticas de exceção e canais de contestação favorecem uso legítimo da tecnologia.

Em pesquisa, conjuntos de logs reais frequentemente contêm informações confidenciais, dificultando abertura integral. Protocolos de anonimização, dados sintéticos documentados e ambiente de avaliação com acesso controlado são opções, cada qual com limitações. A cooperação interinstitucional pode favorecer estudos mais representativos desde que existam critérios compartilhados de documentação, proteção e responsabilização. O ganho científico vem da comparabilidade, não apenas da quantidade de participantes.

29. APLICAÇÃO EM SAÚDE E INFRAESTRUTURAS CRÍTICAS

Em organizações de saúde, dispositivos clínicos, sistemas de prontuário, serviços de imagem e infraestrutura administrativa possuem requisitos distintos de disponibilidade e segurança. Uma detecção sensível demais pode desencadear investigações que interrompem fluxos assistenciais; uma detecção insensível pode retardar resposta a falhas importantes. A avaliação de alertas precisa considerar criticidade operacional e protocolos de segurança próprios da instituição. Não é adequado transferir diretamente limiares treinados em comércio eletrônico ou redes acadêmicas.

A arquitetura federada apresenta interesse quando instituições desejam compartilhar aprendizado sem consolidar registros sensíveis em um repositório comum. Entretanto, diferenças entre fornecedores, dispositivos e processos podem tornar o treinamento conjunto especialmente heterogêneo. É necessário validar se características compartilhadas possuem o mesmo significado em todos os participantes. As proteções de privacidade devem contemplar possíveis informações pessoais em telemetria e

restrições técnicas à execução de agentes em equipamentos regulados.

Qualquer aplicação em infraestrutura crítica exige decisões proporcionais ao risco e testes de operação degradada. A proposta conceitual deste artigo não é um sistema clínico certificado nem especifica protocolos para controle autônomo de equipamentos. Seu valor está em esclarecer requisitos de governança, avaliação e integração ao trabalho profissional, evitando transformar inferências algorítmicas em ações automáticas sem validação.

30. RELEVÂNCIA DA RESPOSTA A INCIDENTES

Detectar uma anomalia é apenas uma etapa do ciclo de gestão de incidentes. A utilidade do alerta depende de informação suficiente para triagem, disponibilidade de responsáveis e mecanismos de investigação proporcionais. As recomendações do NIST para sistemas de detecção e prevenção descrevem a integração de diferentes fontes de monitoramento (Scarfone; Mell, 2007). A atualização de 2025 sobre resposta a incidentes enfatiza a inserção dessas atividades no gerenciamento mais amplo de riscos (Nelson et al., 2025).

A redução de falsos positivos deve ser avaliada à luz desse fluxo. Um modelo que produz alertas menos numerosos, mas menos explicáveis, pode aumentar tempo médio de investigação. Já a agregação contextual que apresenta um episódio com evidências relacionadas pode tornar o processo mais eficiente mesmo sem modificar drasticamente a sensibilidade estatística. É útil aferir tempo até triagem, confirmação, encaminhamento e encerramento, sem confundir esses indicadores com tempos de detecção isolados.

Os resultados das investigações devem retornar à governança do modelo por um processo rastreável. Incidentes confirmados podem exigir revisão de atributos e casos de teste; alarmes indevidos podem motivar calibração; falhas de coleta demandam correção de infraestrutura. Essa classificação por causa raiz impede que todo problema seja tratado como insuficiência algorítmica e favorece melhorias em diferentes partes do sistema.

31. DISCUSSÃO INTEGRADA

A literatura revisada sustenta que algoritmos de anomalia, federação e proteção de privacidade respondem a problemas relacionados, mas distintos. Métodos de detecção procuram reconhecer desvios; a federação coordena aprendizagem sem exigir transferência integral dos dados; e mecanismos de privacidade limitam formas específicas de exposição. Nenhum desses componentes garante, isoladamente, menor taxa de falsos positivos. O desempenho resulta da combinação entre definição de eventos, qualidade da telemetria, distribuição dos participantes, calibração e ação humana.

O principal risco conceitual é tratar a federação como substituta de validação. Um modelo compartilhado pode incorporar variações legítimas de vários ambientes e tornar-se mais tolerante, mas também perder sensibilidade para eventos raros em clientes particulares. De maneira semelhante, um limiar local pode reduzir alarmes, mas esconder ocorrências importantes se for ajustado apenas para satisfazer um orçamento de notificações. As métricas precisam ser examinadas conjuntamente e por contexto, como salientam Saito e Rehmsmeier (2015).

A proposta de arquitetura em camadas oferece um caminho para tornar decisões explícitas: coleta proporcional, aprendizagem compartilhada quando pertinente, proteção das atualizações, calibração contextual, rastreabilidade e avaliação humana. Essa organização não representa resultado de ensaio e deve ser submetida a provas independentes. Seu mérito potencial consiste em orientar hipóteses, facilitar auditorias e evitar a falsa oposição entre privacidade e eficácia operacional, reconhecendo compromissos mensuráveis.

32. LIMITAÇÕES DA REVISÃO

Por se tratar de revisão narrativa, este estudo não apresenta busca exaustiva nem síntese quantitativa. Foram selecionadas referências fundamentais e documentos técnicos pertinentes, mas trabalhos adicionais podem abordar arquiteturas específicas não contempladas. Também não se realizou comparação experimental entre algoritmos, conjuntos de dados ou cenários de ameaça. As afirmações sobre possibilidades de redução de falsos positivos são, portanto, recomendações fundamentadas em mecanismos e evidências gerais, não medidas novas de superioridade.

As práticas de aprendizado federado evoluem rapidamente e podem apresentar limitações distintas conforme número de clientes, arquitetura de rede e natureza dos registros. Garantias de privacidade dependem de parametrização e modelos formais de ameaça; não é possível extrapolá-las de uma referência geral a todos os ambientes. O contexto regulatório também exige revisão diante de mudanças normativas e diferentes finalidades institucionais.

Por fim, desempenho técnico e impacto organizacional são dimensões diferentes. Reduzir uma métrica de falso positivo não demonstra automaticamente maior segurança, menor custo ou maior confiança dos analistas. Estudos prospectivos deveriam medir eventos efetivamente investigados, carga de trabalho, omissões e consequências das decisões. Essas limitações delimitam o alcance deste manuscrito e reforçam a necessidade de transparência metodológica.

33. AGENDA DE PESQUISA FUTURA

A agenda futura deve priorizar experimentos com divisão temporal estrita, participantes heterogêneos e prevalência realista de eventos relevantes. Testes com clientes totalmente não vistos seriam particularmente úteis para avaliar generalização. Também é necessário comparar FedAvg, métodos tolerantes à heterogeneidade e mecanismos de personalização sob o mesmo orçamento de comunicação. Os resultados devem ser relatados por unidade, com intervalos de confiança e decomposição de falsos positivos por causa operacional.

Uma segunda linha envolve a relação entre privacidade e utilidade. Experimentos controlados podem investigar como diferentes políticas de agregação segura, minimização de atributos e ruído alteram a sensibilidade a eventos raros e o custo de treinamento. Os modelos de ameaça precisam ser explícitos, evitando afirmações genéricas de confidencialidade. Estudos de governança poderiam analisar confiança entre organizações e condições de compartilhamento de atualizações em setores regulados.

Finalmente, recomenda-se avaliar a interação humano–algoritmo por medidas de trabalho real. A comparação entre alertas brutos e episódios consolidados, com análise de tempo de triagem e qualidade da explicação, pode produzir resultados mais úteis à prática do que uma variação isolada de acurácia. A agenda deve incluir testes de deriva, revisão de versões, documentação de incidentes e impactos não intencionais de decisões automatizadas.

33.1. Padronização de Dados e Interoperabilidade

A compatibilidade entre participantes não se esgota na escolha de um formato de arquivo. Duas organizações podem utilizar o mesmo campo chamado erro, mas empregar regras de contagem distintas, produzir agregações em janelas incompatíveis ou registrar eventos com granularidades divergentes. Antes do treinamento conjunto, é necessário construir um dicionário de dados que especifique significado, unidade, periodicidade, limites aceitáveis e transformações de cada atributo. Testes automatizados podem identificar alterações abruptas de esquema sem revelar os registros de origem. Esse procedimento permite que problemas de interoperabilidade sejam detectados antes de produzir atualizações de modelo potencialmente enganosas.

Os participantes também precisam concordar sobre regras de qualidade e ausência de dados. Substituir valores ausentes por zero pode fazer sentido para um contador, mas induzir erro quando a ausência resulta de falha de instrumentação. Normalizações calculadas individualmente podem proteger peculiaridades locais; por outro lado, sua incompatibilidade pode dificultar interpretação de parâmetros agregados. A decisão deve ser registrada junto da versão do esquema. Em estudos empíricos, apresentar

documentação de atributos e testes de compatibilidade é tão relevante quanto divulgar a arquitetura neural empregada.

33.2. Equidade de Desempenho e Clientes Minoritários

Um sistema compartilhado pode alcançar boa média global e, ao mesmo tempo, falhar em participantes com baixa capacidade computacional ou perfis pouco frequentes. A expressão equidade de desempenho refere-se aqui à necessidade operacional de avaliar se clientes pequenos ou distintos sofrem desvantagem sistemática. Não implica que todos devam produzir taxas idênticas, pois prevalência de incidentes, criticidade e disponibilidade de rótulos variam. Requer, porém, transparência sobre diferenças e justificativas de decisões de ponderação que afetem segurança de unidades específicas.

A análise deve comparar sensibilidade, quantidade de alarmes e latência de inferência por classe de cliente, não apenas por unidade individual. Se um laboratório de pesquisa apresentar fluxo de dados muito distinto de serviços administrativos, pode precisar de camada de personalização ou de um modelo de referência próprio. A exclusão de clientes por indisponibilidade frequente também merece monitoramento: o modelo global não deve ser tratado como representativo se alguns grupos nunca participam das rodadas. Uma federação responsável registra cobertura de participação e avalia se as melhorias alegadas se distribuem entre segmentos.

33.3. Auditoria da Redução de Alarmes

Quando a meta organizacional é reduzir falsos positivos, há risco de criar um incentivo para descartar alertas em vez de melhorar decisões. Uma auditoria deve preservar a distinção entre total de

candidatos gerados, alertas consolidados, casos encaminhados e episódios efetivamente analisados. A queda de notificações por agrupamento não equivale necessariamente à elevação de precisão do classificador, assim como uma alteração do limiar não demonstra maior qualidade da representação aprendida. Cada estágio deve ter métricas próprias e histórico de versões para possibilitar comparação.

Uma amostra independente de episódios não alertados pode ajudar a estimar omissões, desde que selecionada por critérios explícitos e submetida a revisão qualificada. A análise pode incluir períodos de maior carga e clientes de menor representatividade para revelar falhas encobertas na média. Painéis operacionais devem apresentar tendências de revocação, precisão, volume de trabalho e casos revisados, sem reduzir segurança a um único indicador. A prestação de contas é particularmente relevante quando a diminuição de alertas altera escala de equipes ou procedimentos de resposta.

33.4. Registro de Modelos e Reprodutibilidade

A reprodutibilidade de uma arquitetura federada requer mais do que disponibilizar um conjunto de hiperparâmetros. Devem ser identificados código utilizado, dependências, versão do esquema de telemetria, conjunto de clientes elegíveis, política de amostragem, número de rodadas, tratamento de interrupções, critérios de parada e versão dos limiares. A divulgação de sementes aleatórias pode apoiar experimentos, mas não elimina a variabilidade gerada por diferenças de infraestrutura e alterações temporais no conjunto local. Um registro de modelos permite reconstituir decisões e distinguir execução de teste e implantação efetiva.

A governança de versões deve determinar quais evidências autorizam promoção de um modelo para produção. Recomenda-se conservar relatórios de validação em períodos futuros, resultados por participante, falhas conhecidas e condições que motivam rollback. Quando existem requisitos de privacidade, os registros devem evitar armazenamento desnecessário de informações que permitam reconstrução de exemplos locais. O objetivo é produzir auditabilidade técnica proporcional ao risco, garantindo que mudanças sejam explicáveis mesmo após substituição de equipamentos ou responsáveis pelo sistema.

33.5. Participação Organizacional e Ciclo de Vida

Incluir e retirar participantes de uma federação altera composição estatística, relações de confiança e representatividade. Uma unidade recém-integrada pode ter poucos dados para calibrar sua distribuição e não deveria influenciar imediatamente todo o modelo sem avaliações iniciais. Da mesma forma, a saída de uma organização exige procedimentos para revogar credenciais, interromper participação e examinar retenção de atualizações anteriores. Esses eventos devem fazer parte do planejamento do ciclo de vida, e não ser tratados apenas como operações administrativas.

Mudanças de finalidade também precisam de avaliação. Um detector inicialmente desenhado para disponibilidade de serviços pode ser reaproveitado para investigar comportamento individual, alterando significativamente implicações de privacidade e expectativas dos titulares. A governança deve restringir usos secundários não autorizados, preservar a separação entre funções de operação e investigação e prever revisão periódica das hipóteses de

risco. A aprendizagem federada oferece mecanismo técnico de cooperação, mas a legitimidade de sua continuidade depende de acordos institucionais, supervisão e demonstração de necessidade.

34. CONSIDERAÇÕES FINAIS

A detecção de anomalias em ambientes computacionais distribuídos exige conciliar a diversidade de operações, a raridade de eventos relevantes, as restrições de compartilhamento de registros e a capacidade limitada de investigação humana. A aprendizagem federada oferece um mecanismo de coordenação entre participantes sem obrigar centralização dos dados brutos, mas seu uso exige enfrentar heterogeneidade, participação parcial, atualização segura e possível exposição de informação por parâmetros. A proteção de privacidade deve combinar decisões arquiteturais, mecanismos técnicos e governança institucional.

A redução de falsos positivos depende sobretudo de definição consistente de eventos, validação temporal, calibração por contexto, agrupamento de episódios e incorporação criteriosa do retorno de analistas. Nenhuma técnica pode ser considerada superior sem teste com dados representativos e avaliação simultânea da sensibilidade. A proposta conceitual em camadas sintetiza requisitos para uma plataforma auditável, mantendo separadas a geração de pontuações, a decisão sobre alertas e a resposta institucional.

Conclui-se que a integração entre algoritmos adaptativos, federação e privacidade pode orientar soluções promissoras, desde que sustentada por avaliação reprodutível, controles de segurança e acompanhamento humano. O avanço da área depende menos da adoção indiscriminada de novos modelos e mais da demonstração

transparente de quais condições permitem reconhecer incidentes relevantes com menor custo operacional e proteção adequada dos dados.

REFERÊNCIAS BIBLIOGRÁFICAS

BONAWITZ, Keith et al. Practical secure aggregation for privacy-preserving machine learning. In: ACM SIGSAC CONFERENCE ON COMPUTER AND COMMUNICATIONS SECURITY, 2017. Proceedings [...]. New York: ACM, 2017. p. 1175–1191. https://doi.org/10.1145/3133956.3133982.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018. Disponível em: https://www2.camara.leg.br/legin/fed/lei/2018/lei-13709-14-agosto-2018-787077-normaatualizada-pl.html. Acesso em: 10 nov. 2026.

CHANDOLA, Varun; BANERJEE, Arindam; KUMAR, Vipin. Anomaly detection: a survey. ACM Computing Surveys, v. 41, n. 3, art. 15, 2009. https://doi.org/10.1145/1541880.1541882.

KAIROUZ, Peter et al. Advances and open problems in federated learning. Foundations and Trends in Machine Learning, v. 14, n. 1–2, p. 1–210, 2021. https://doi.org/10.1561/2200000083.

LI, Tian et al. Federated optimization in heterogeneous networks. Proceedings of Machine Learning and Systems, v. 2, 2020. Disponível em: https://proceedings.mlsys.org/paper_files/paper/2020/hash/1f5fe83998a09396ebe6477d9475ba0c-Abstract.html. Acesso em: 10 nov. 2026.

MCMAHAN, H. Brendan et al. Communication-efficient learning of deep networks from decentralized data. Proceedings of Machine Learning Research, v. 54, p. 1273–1282, 2017. Disponível em: https://proceedings.mlr.press/v54/mcmahan17a.html. Acesso em: 10 nov. 2026.

NELSON, Alexander; REKHI, Sanjay; SOUPPAYA, Murugiah; SCARFONE, Karen. Incident response recommendations and considerations for cybersecurity risk management: a CSF 2.0 community profile. NIST Special Publication 800-61, rev. 3. Gaithersburg: NIST, 2025. https://doi.org/10.6028/NIST.SP.800-61r3.

RUFF, Lukas et al. A unifying review of deep and shallow anomaly detection. Proceedings of the IEEE, v. 109, n. 5, p. 756–795, 2021. https://doi.org/10.1109/JPROC.2021.3052449.

SAITO, Takaya; REHMSMEIER, Marc. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS ONE, v. 10, n. 3, e0118432, 2015. https://doi.org/10.1371/journal.pone.0118432.

SCARFONE, Karen; MELL, Peter. Guide to intrusion detection and prevention systems (IDPS). NIST Special Publication 800-94. Gaithersburg: NIST, 2007. https://doi.org/10.6028/NIST.SP.800-94.

---

[1] Especializações em Cybersecurity: Detecção de Ameaças e Proteção de Sistemas de Informática e Comunicações – FANS; Redes de Computadores – FANESE; Inteligência Artificial – FANS. Graduações em Sistemas de Informação – UNIT e Administração – ETEP. Lattes: http://lattes.cnpq.br/8752905296064861. E-mail: acesse o artigo original para visualizar o e-mail

[2] Especializações em Linguagens, suas Tecnologias e o Mundo do Trabalho – UFPI; Metodologia do Ensino da Língua Inglesa – FAVENI; MIT em Engenharia de Software (.NET/C#) – INFNET. Segunda Licenciatura em Informática – UNIASSELVI. Graduações em Letras–Inglês – ESTÁCIO e Análise e Desenvolvimento de Sistemas – UNICARIOCA. ORCID: https://orcid.org/0009-0005-3263-7879. Lattes: https://lattes.cnpq.br/9898138931949263. E-mail: acesse o artigo original para visualizar o e-mail

[3] Mestrado em Ciências da Educação – INTERAMERICANA; especializações em Redes de Computadores – FANESE e Metodologia e Didática de Nível Superior – SERIGY; graduação em Gestão de Tecnologia da Informação – FANESE. Lattes: http://lattes.cnpq.br/0606923050760749. E-mail: acesse o artigo original para visualizar o e-mail

[4] Especializações em Engenharia de Software – FOCUS; Gestão de Riscos e Cibersegurança – FOCUS; Tecnologia da Informação – UNIFAVENI; graduação em Pedagogia – FSLF. Lattes: http://lattes.cnpq.br/8302147764638511. E-mail: acesse o artigo original para visualizar o e-mail

[5] Mestrado em Tecnologia Nuclear – Aplicações – IPEN; graduação em Física Médica – UFS/SE. ORCID: https://orcid.org/0009-0003-1003-1994. Lattes: https://lattes.cnpq.br/9601959592189180. E-mail: acesse o artigo original para visualizar o e-mail

[6] Especializações em Administração Hospitalar – Centro Universitário São Camilo e Sistema de Informação para Web – FANESE; graduação em Administração com Ênfase em Análise de Sistemas – UNIT; Formação Pedagógica para Formadores de Educação Profissional –

UNISUL. ORCID: https://orcid.org/0000-0001-6676-575X. Lattes: https://lattes.cnpq.br/6247498924148402. E-mail: acesse o artigo original para visualizar o e-mail

[7] Especialização em Segurança de Redes de Computadores – FACUMINAS; tecnólogo em Gestão da Tecnologia da Informação – FANESE; Formação Pedagógica em Informática – UNIASSELVI. ORCID: https://orcid.org/0009-0002-8135-3248. Lattes: https://lattes.cnpq.br/8383104135339670. E-mail: acesse o artigo original para visualizar o e-mail