KETRIN: Elicitação Adaptativa de Conhecimento para Diagnóstico de Maturidade em Proteção de Dados

Modalidade: Código Aberto

Ketrin Diovana Alves Rodrigues Vargas[1], Tuigg R. Barcelos[1], Camilla B. Quincozes[1,2], Silvio E. Quincozes[1,2], Paulo Silas Severo de Souza[1]

[1]AI Horizon Labs, PPGES, Universidade Federal do Pampa (UNIPAMPA)
[2]AI Horizon Labs, PPGCO, Universidade Federal de Uberlândia (UFU)
{ketrinvargas, tuiggbarcelos, camillaborchhardt}.aluno@unipampa.edu.br

{silvioquincozes, paulosilas}@unipampa.edu.br

Abstract. *KETRIN is a web platform with publicly available source code for response-adaptive elicitation of organizational knowledge about personal-data processing. Instead of presenting a fixed checklist, it generates questions in four stages, uses previous answers and detected weaknesses to select follow-up probes, and produces a contextualized maturity report and action plan. Its architecture separates a React interface, a Node.js orchestration layer with automatic failover among four LLM providers, and Firebase persistence. A pilot recorded 18 completed diagnoses, an 86% completion rate, and a mean user rating of 3.21/4. A reproducible campaign with 55 synthetic personas generated 550 question-answer pairs and exposed strengths and limitations of the lexical detection layer. Source code, scripts, personas, oracles, and logs are publicly available.*

Resumo. *A* KETRIN *é uma plataforma web com código-fonte publicamente disponível para elicitação adaptativa de conhecimento organizacional sobre o tratamento de dados pessoais. Em vez de aplicar uma lista fixa, ela gera perguntas em quatro estágios, usa respostas anteriores e fragilidades detectadas para aprofundar a investigação e produz diagnóstico contextualizado de maturidade e plano de ação. A arquitetura separa interface React, orquestração Node.js com contingência entre quatro provedores de LLM e persistência no Firebase. Um piloto registrou 18 diagnósticos concluídos, taxa de conclusão de 86% e avaliação média de 3,21/4. Uma campanha reprodutível com 55 personas sintéticas gerou 550 pares de pergunta e resposta e revelou capacidades e limites da camada lexical de detecção. Código, scripts, personas, oráculos e registros estão publicamente disponíveis.*

Artefato e demonstração online: `https://github.com/KetrinDiovanaVargas/PlatformLGPDCompliance`; `https://platformlgpdcompliance.com.br`.
Vídeo técnico de uso: `https://drive.google.com/file/d/1qc-2eQvFO7-oeuLly41Yc9WmcMsbNpmM/view?usp=sharing`.

1. Introdução

A Lei Geral de Proteção de Dados Pessoais (LGPD) exige que organizações conheçam e consigam justificar como coletam, acessam, compartilham, retêm e protegem dados

pessoais [Brasil, 2018]. Esse conhecimento, porém, costuma estar distribuído entre pessoas, sistemas e práticas informais. Uma política pode determinar o uso de repositório corporativo enquanto a rotina real recorre a planilhas locais ou aplicativos de mensagens. Assim, avaliar maturidade em proteção de dados não é apenas verificar documentos: é elicitar como o trabalho de fato ocorre e relacionar as evidências encontradas a riscos e controles.

Questionários estáticos e auditorias manuais têm limitações complementares. O primeiro mantém comparabilidade, mas não esclarece respostas vagas nem investiga uma pista inesperada. O segundo permite aprofundamento, porém exige especialistas, tempo e esforço difíceis de sustentar em avaliações recorrentes. Modelos de linguagem de grande escala (LLMs) abrem espaço para instrumentos intermediários, capazes de conduzir perguntas abertas e adaptar a sondagem ao contexto [Hassani et al., 2024, Adhikari et al., 2025]. O risco é trocar rigidez por uma conversa sem controle: perguntas repetidas, mudanças bruscas de tema, diagnósticos sem evidência ou dependência de um único provedor.

Este artigo apresenta a KETRIN (*Knowledge Elicitation Through Response-Adaptive Inquiries*), uma ferramenta que operacionaliza a elicitação em quatro estágios. Cada estágio possui um objetivo, mas suas perguntas são geradas durante a sessão. O histórico, o perfil da avaliação e sinais de dez categorias de fragilidade orientam perguntas de aprofundamento; ao final, a ferramenta calcula indicadores, organiza os achados e gera recomendações priorizadas. Administradores criam avaliações e acompanham respostas; participantes acessam um link sem cadastro e sem fornecimento obrigatório de identidade.

As contribuições da ferramenta são: (i) fluxo multiestágio que combina estrutura global e perguntas adaptativas; (ii) arquitetura de contingência entre *Claude, Groq, DeepSeek e Gemini*, com fila, limitação de requisições, *cache* e respostas seguras de contingência; (iii) diagnóstico visual com evidências e plano de ação; e (iv) um conjunto reprodutível de 55 personas, oráculos separados, *scripts* e logs para testar a capacidade de elicitação. A avaliação combina um piloto com respondentes reais e uma campanha controlada. O restante do artigo discute trabalhos relacionados, arquitetura, implementação, resultados e a demonstração planejada para o Salão de Ferramentas.

2. Contexto e Trabalhos Relacionados

Estruturas como o NIST (National Institute of Standards and Technology) *Privacy Framework* organizam a gestão de privacidade como processo contínuo e baseado em risco [National Institute of Standards and Technology, 2020]. Ferramentas de maturidade convertem essa orientação em dimensões e níveis, mas dependem da qualidade das informações fornecidas. Em um formulário fixo, respostas curtas ou excessivamente positivas permanecem sem contraditório; uma pergunta adaptativa pode solicitar exemplo, responsável, evidência, canal usado ou exceção observada.

A LGPD difere do GDPR em ênfase: enquanto GDPR foca em direitos individuais, LGPD prioriza responsabilização por processos de tratamento. Maturidade em proteção de dados é progressão contínua, não certificação binária. A KETRIN mapeia práxis organizacional para diagnóstico, não garante conformidade legal, auditoria e parecer jurídico permanecem essenciais.

Agentes já foram empregados na adaptação de questionários [Paduraru et al., 2024], e LLMs vêm sendo investigados tanto como entrevistadores [Korn et al., 2025] quanto

como respondentes sintéticos para testar instrumentos [Bachmann et al., 2025]. *Elicitron* usa agentes para simular atores durante elicitação de requisitos [Ataei et al., 2025]. A literatura, contudo, alerta que personas geradas por LLMs não substituem populações humanas e podem reproduzir vieses ou homogeneizar respostas [Argyle et al., 2023, Bisbee et al., 2024]. Por isso, na KETRIN, personas são usadas como testes controlados do instrumento, com fragilidades definidas previamente e oráculos ocultos, enquanto clareza, utilidade e intenção de reuso são aferidas com pessoas.

A diferença central da KETRIN é integrar, em uma ferramenta executável, quatro elementos que usualmente aparecem isolados: questionário adaptativo, taxonomia explícita de fragilidades, relatório acionável e bancada de validação versionada. O sistema não pretende certificar conformidade jurídica; ele apoia a descoberta e a priorização de práticas que devem ser verificadas por responsáveis de privacidade e segurança.

3. Arquitetura e Fluxo Adaptativo

3.1. Camadas e componentes

"Na Figura 1 esquematiza-se a arquitetura que separa apresentação, orquestração e persistência. A aplicação *React/TypeScript* atende participantes e gestores. O servidor *Express* recebe metadados e respostas, valida a entrada, produz o *prompt* do estágio, seleciona o provedor e normaliza a saída estruturada. O Firestore armazena avaliações, sessões, estágios, relatórios e *feedbacks*. Serviços de \llm permanecem externos e somente a orquestração os aciona

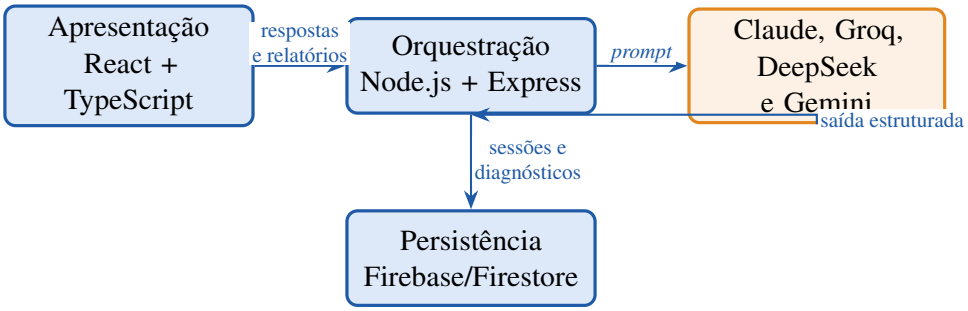


Figure 1. Arquitetura lógica da KETRIN.

O acesso possui três perfis. O participante responde pelo link público da avaliação e visualiza seu diagnóstico. O ADMIN configura objetivo, contexto, público e provedor preferencial, distribui o link e acompanha suas avaliações. O MASTER gerencia administradores e possui visão consolidada. A separação permite a aplicação sem cadastro de respondentes e mantém a gestão em área autenticada.

3.2. Quatro estágios com sondagem orientada a evidências

O fluxo preserva comparabilidade por meio de quatro estágios, mas não fixa as perguntas. O primeiro estabelece contexto e rotina; o segundo percorre controles e processos; o terceiro investiga riscos, responsabilidades e terceiros; e o quarto procura maturidade e evidências. A Tabela 1 resume a divisão. A implementação gera quatro questões no primeiro estágio e duas em cada estágio subsequente, totalizando dez por sessão. Antes de aceitar o conjunto, remove duplicatas e perguntas semanticamente próximas das anteriores.

Table 1. Estágios e decisões adaptativas do questionário.

Etapa	Foco	Exemplos de aprofundamento
1	Contexto organizacional	Papel, rotina, dados tratados, ferramentas e frequência de contato.
2	Controles e processos	Canais reais, armazenamento, acesso, coleta, retenção e exceções.
3	Riscos e governança	Responsáveis, terceiros, dependências, incidentes e escalonamento.
4	Maturidade e evidências	Políticas, registros, revisão, monitoramento e comprovação de controles.

Após cada resposta, uma camada lexical procura sinais em dez eixos: compartilhamento informal (F1), armazenamento inadequado (F2), retenção indefinida (F3), coleta excessiva (F4), acesso além da função (F5), falta de transparência (F6), uso secundário (F7), terceiros sem controle (F8), dados sensíveis sem salvaguardas (F9) e incidente mal tratado (F10). Os sinais não constituem o diagnóstico por si sós; eles geram instruções de sondagem para o LLM. Por exemplo, a menção a envio por aplicativo pessoal leva a perguntar quem recebe, se o arquivo permanece no dispositivo e se existe registro do compartilhamento. Categorias abstratas, como transparência, dependem da interpretação contextual do modelo.

Concluído o quarto estágio, o histórico integral alimenta o relatório final. A saída contém escore e nível de maturidade, distribuição de risco, achados vinculados a evidências textuais, pontos fortes, lacunas, recomendações priorizadas e um plano de ação. Essa rastreabilidade reduz o risco de uma recomendação genérica desconectada das respostas.

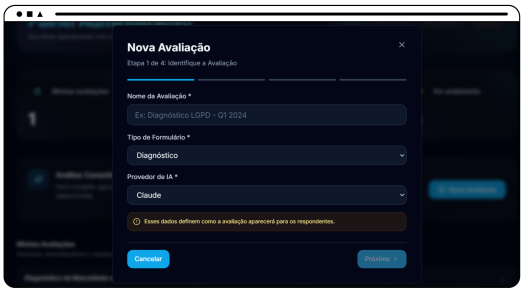
4. Implementação e Funcionalidades

A versão avaliada utiliza React 18, TypeScript e Vite no *frontend*; Tailwind CSS e shadcn/ui na interface; Recharts para visualização; Node.js/Express no servidor; e Firebase Authentication/Firestore. O artefato contém scripts de desenvolvimento, produção, testes e simulação. A instalação local requer Node.js 18 ou superior, `npm install`, configuração das variáveis do Firebase e de ao menos um provedor de LLM, seguida de `npm run dev`. O sistema também possui implantação web.

A camada de LLM aceita um provedor preferencial, mas percorre uma ordem de contingência se ocorrer ausência de chave, indisponibilidade ou limite de requisições. Uma fila configura concorrência e intervalo por provedor; um *cache* evita repetir gerações equivalentes; limites por IP reduzem abuso; e perguntas genéricas distintas por estágio funcionam como contingência segura caso a geração estruturada falhe.

A validação de entrada funciona em três camadas. Primeiro, tamanho: respostas de usuários são limitadas a 2.000 caracteres, enquanto personas ficam em 5.000, evitando overflow. Segundo, padrões: regex busca frases típicas de prompt injection como "ignore", "override" e "disregard", neutralizando-as; URLs externas e blocos de código também são detectados e marcados. Terceiro, estrutura: respostas são parametrizadas de forma segura antes de ir ao prompt, reduzindo risco de injeção. Personas também usam as mesmas proteções (Seção 5.2).

Na Figura 2 apresentam-se dois momentos do uso. O assistente de criação recebe nome, tipo e provedor; no formulário, a próxima pergunta é produzida a partir do contexto acumulado. Na Figura 3 mostra-se a síntese visual e os achados detalhados. O painel administrativo (Figura 4) agrega quantidade de avaliações, respostas e diagnósticos, e permite exportar *feedbacks* em planilha.

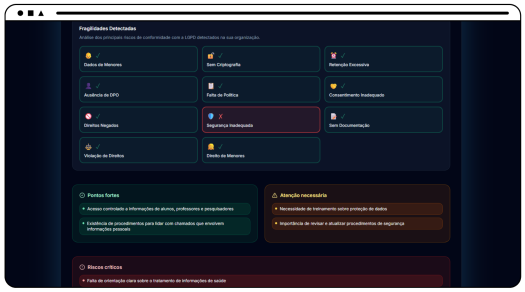


(a) Configuração de uma avaliação.

(b) Pergunta aberta no primeiro estágio.

Figure 2. Criação e aplicação do questionário adaptativo.

(a) Escore e distribuição de risco.



(b) Achados, pontos fortes e lacunas.

Figure 3. Relatório contextualizado entregue ao respondente.

5. Avaliação e Resultados

A avaliação seguiu duas frentes complementares. O piloto com pessoas mede aceitação e utilidade percebida. A campanha com personas testa, sob condições conhecidas, se o instrumento faz emergir fragilidades predefinidas. Os números foram extraídos do painel, dos logs e dos scripts versionados. Não se atribui às personas validade ecológica; elas funcionam como casos de teste com gabarito, enquanto a interpretação dos resultados observa ameaças usuais à validade [Wohlin et al., 2012].

5.1. Piloto com respondentes reais

Foram iniciadas 21 sessões; 18 percorreram os quatro estágios, taxa de conclusão de 86%. A amostra piloto era constituída predominantemente por estudantes de pós-graduação , profissionais de tecnologia em organizações de médio porte e consultores especialistas em conformidade. Quanto à experiência profissional, 67% relataram entre dois e cinco anos de atuação em compliance ou governança de dados; nenhum respondente possuía certificação formal em LGPD. Esse perfil, ainda que compatível com o contexto acadêmico de coleta de dados, compromete a generalização dos achados para populações de auditores sênior e executivos de segurança uma ameaça à validade externa detalhada na Seção 7.

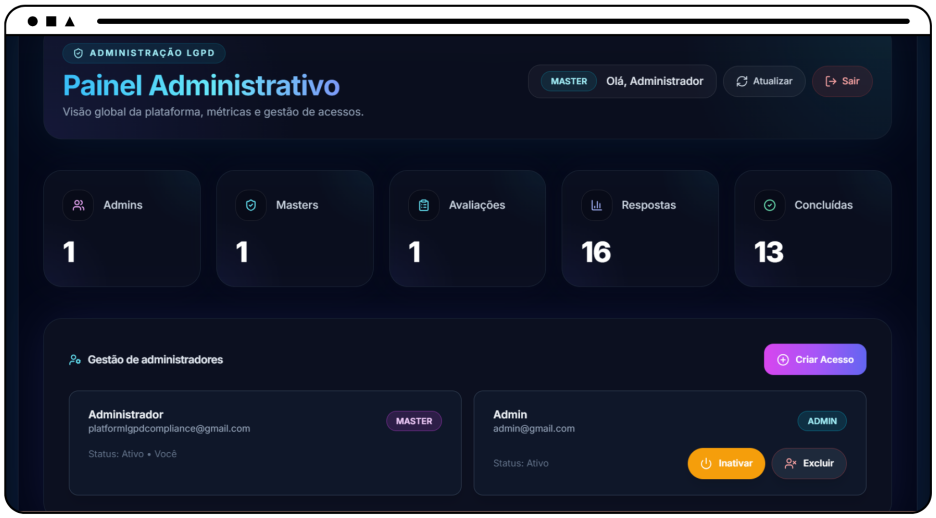


Figure 4. Painel do ADMIN para gestão e acompanhamento das avaliações.

Doze concluintes responderam a dez itens em escala Likert de 1 a 4. A média geral foi 3,21; utilidade percebida e experiência geral alcançaram 3,29, e clareza/facilidade, 3,08. O item "pode contribuir para melhorar processos de LGPD" teve média 3,50 e 92% de concordância. Todos concordaram parcial ou plenamente que utilizariam a ferramenta novamente. O principal ponto de melhoria foi a clareza e coerência da sequência adaptativa: média 2,75 e 67% de concordância, sugerindo explicitar por que uma pergunta deriva da anterior.

Nos 18 diagnósticos concluídos, a conformidade média calculada foi 54% e a plataforma sinalizou 31 fragilidades críticas. Compartilhamento informal concentrou 23 ocorrências e acesso excessivo, seis; armazenamento e retenção tiveram uma ocorrência cada. Esses valores descrevem o piloto e não estimam prevalência populacional.

A análise mostrou desempenho variável por categoria de fragilidade. F1, F2 e F9 atingiram alta precisão (0,95, 0,93, 0,89), enquanto F8 ficou em apenas 0,36. Isso ocorre porque respondentes usam termos genéricos ao descrever fornecedores e parceiros, gerando muitas detecções desnecessárias mesmo quando contratos já existem. F10 (incidentes) teve baixa revocação (0,13), emergindo apenas em estágios posteriores. Esses resultados indicam que a análise léxica captura fragilidades explícitas, mas lacunas estruturais precisam de diálogo mais profundo.

5.2. Campanha com personas e oráculos

O conjunto possui 50 personas operacionais e cinco adversariais, distribuídas por setores, cargos e dez categorias. Cada persona tem um arquivo de comportamento fornecido ao modelo e um oráculo YAML separado, inacessível durante a sessão. Em 2 de julho de 2026, todas as 55 sessões foram concluídas pelo fluxo adaptativo de produção. Claude Haiku 4.5, com temperatura 0,2, representou os respondentes. O corpus contém 550 pares de pergunta e resposta, dez por persona, e média aproximada de 735 palavras respondidas por sessão.

As 55 personas tiveram distribuição desbalanceada por tamanho organizacional:

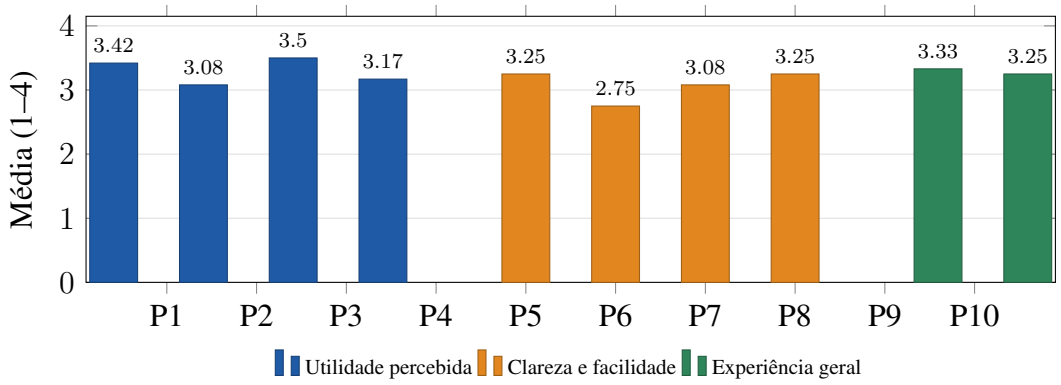


Figure 5. Média dos itens de *feedback* do piloto ($n = 12$).

aproximadamente 60% médias, 20% pequenas e 20% grandes. Sem estratificação, métricas agregadas podem mascarar desempenho em cada faixa de tamanho. Trabalhos futuros devem incluir pelo menos 15 personas por faixa e reportar F1-score estratificado para validação mais robusta.

A análise reproduzível, implementada pelo script de avaliação lexical incluído no repositório, isola essa camada: concatena as respostas, aplica os mesmos padrões usados na ferramenta e compara detecções ao oráculo. Portanto, os números seguintes não medem o diagnóstico híbrido completo nem comparam um questionário fixo ao adaptativo. As cinco personas adversariais foram excluídas da matriz por não possuírem vetor categórico completo, restando 50.

Nas oito categorias com expressões regulares, precisão agregada foi 0,625, revocação 0,509 e F1 0,561. A fragilidade central foi encontrada lexicalmente em 31 das 47 personas que a possuíam (66%). F1, F2 e F9 tiveram alta precisão (0,952; 0,929; e 0,895), mostrando que sinais específicos geram poucos alarmes falsos. F8 disparou em todas as personas porque palavras como “fornecedor” e “parceiro” são amplas, reduzindo sua precisão a 0,36. F10 teve revocação 0,133, pois incidentes frequentemente só apareceram após perguntas de aprofundamento. F6 e F7 não possuem detector lexical e dependem da interpretação contextual.

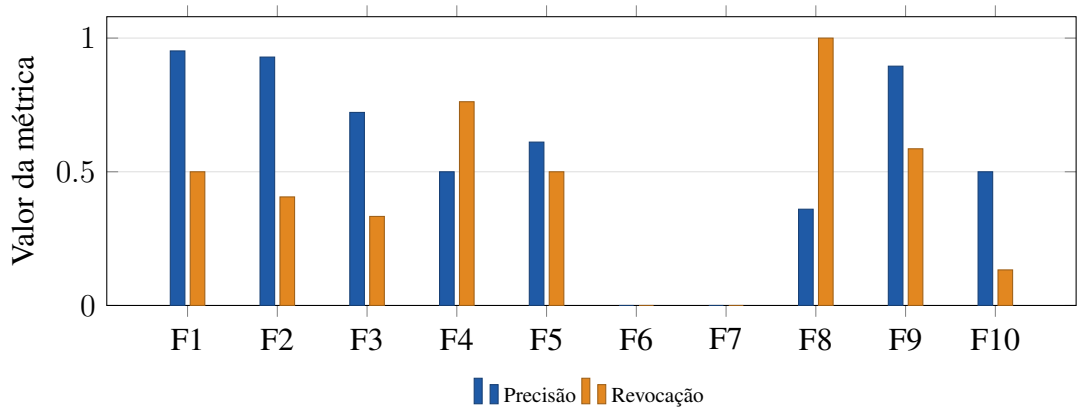


Figure 6. Precisão e revocação da camada lexical por categoria ($n = 50$).

A revocação cresceu com a severidade atribuída pelo oráculo: 0,348 para fragili-

dades leves, 0,522 para moderadas e 0,733 para severas. Além disso, 123 das 176 primeiras detecções (70%) ocorreram no estágio inicial. Em conjunto, os dados mostram que a camada lexical reconhece práticas enunciadas diretamente, mas perde sinais sutis ou contextuais. O resultado orienta duas melhorias: substituir padrões amplos por classificadores semânticos calibrados e tornar a sondagem adaptativa mais explícita ao participante.

6. Demonstração Planejada e Disponibilidade

A demonstração no Salão de Ferramentas reproduzirá um ciclo completo, com três papéis. Primeiro, no perfil ADMIN, será criada uma avaliação de diagnóstico, com objetivo, contexto, público e provedor preferencial. Segundo, um participante abrirá o link anônimo, responderá a uma prática de compartilhamento e o público observará uma pergunta de aprofundamento gerada a partir da resposta. Terceiro, serão exibidos o relatório individual e o painel consolidado. Como extensão reprodutível, será executada uma persona e mostrada a separação entre especificação, oráculo e log.

Não há requisito de *hardware* especializado. A demonstração necessita de notebook com navegador atual, projetor e acesso à Internet para a implantação e os provedores de LLM. Como contingência, serão mantidos uma compilação local, uma sessão previamente registrada e o vídeo técnico. O repositório inclui código, `package-lock.json`, guia de instalação, testes, 55 personas, oráculos, logs e scripts de consolidação. Em validação independente para esta submissão, `npm test` executou 54 testes com sucesso e `npm run build` produziu a versão de produção.

Para reprodução da avaliação lexical, após instalar as dependências, basta executar:

```
node scripts/avaliar_regex_vs_oraculo.mjs
```

O comando lê os logs e oráculos versionados e regenera as matrizes e métricas em CSV/JSON. A disponibilidade de uma implantação pública facilita a inspeção funcional, enquanto o pacote local permite auditar o código e reproduzir a análise sem depender de dados privados.

7. Limitações

A KETRIN apoia autoavaliação e não substitui auditoria, parecer jurídico ou evidência independente. Respostas podem ser incompletas, incorretas ou estratégicas; saídas de LLMs permanecem probabilísticas; e a campanha sintética não representa a diversidade integral de pessoas e organizações. O piloto é pequeno e majoritariamente acadêmico. O piloto oferece insights úteis, mas com ressalvas importantes. Os 12 respondentes eram majoritariamente acadêmicos e iniciantes em compliance, o que pode explicar tanto as avaliações altas de usabilidade (3,21/4) quanto a valorização de clareza sobre outras dimensões. É possível que auditores experientes ou especialistas legais, familiarizados com metodologias de avaliação mais rigorosas, tivessem reações diferentes mais críticas ou, alternativamente, mais apreciativas de especificidades. A estratificação da amostra por experiência e contexto profissional teria revelado essas possíveis diferenças, permitindo conclusões mais robustas.

A análise por expressões regulares avalia somente uma parte do mecanismo híbrido e revelou alarmes falsos em categorias de vocabulário amplo. Além disso, a implantação de produção deve aplicar regras de acesso e retenção compatíveis com o contexto de cada organização antes de receber dados reais.

7.1. Uso de Inteligência Artificial

Ferramentas de IA Generativa foram utilizadas em duas frentes de apoio neste trabalho. No desenvolvimento da ferramenta, auxiliaram na implementação de código, geração de trechos repetitivos, depuração e versionamento, sempre sob revisão e teste dos autores; a concepção da arquitetura, o fluxo adaptativo de quatro estágios, a taxonomia de fragilidades e a bancada de validação são contribuições intelectuais dos autores. Na redação deste artigo, o apoio limitou-se à revisão linguística e à organização textual. Em nenhuma das frentes essas ferramentas geraram resultados, analisaram dados ou definiram conclusões. Todas as decisões de projeto e metodológicas, bem como a responsabilidade pelo conteúdo e pelo artefato, permanecem integralmente com os autores.

8. Conclusão

A KETRIN transforma uma lista de verificação em uma entrevista estruturada que reage às respostas, solicita evidências e devolve um diagnóstico acionável. A ferramenta combina quatro estágios, múltiplos provedores, gestão de avaliações, relatórios e uma bancada reprodutível de validação. O piloto indica boa conclusão e intenção de reuso; a campanha com personas revela, de forma auditável, onde a detecção lexical é precisa e onde a interpretação contextual deve evoluir. Como trabalhos futuros, serão priorizados detector semântico calibrado, explicações de transição entre perguntas e avaliação comparativa com especialistas e organizações de diferentes portes.

Referências

Adhikari, D. M., Hartland, A., Weber, I., and Cannanure, V. K. (2025). Exploring LLMs for automated generation and adaptation of questionnaires. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces*.

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., and Wingate, D. (2023). Out of one, many: Using language models to simulate human samples. *Political Analysis*, 31(3):337–351.

Ataei, M., Cheong, H., Grandi, D., Wang, Y., Morris, N., and Tessier, A. (2025). Elicitron: A large language model agent-based simulation framework for design requirements elicitation. *Journal of Computing and Information Science in Engineering*, 25(2):021012.

Bachmann, F., van der Weijden, D., Heitz, L., Sarasua, C., and Bernstein, A. (2025). Adaptive political surveys and GPT-4: Tackling the cold start problem with simulated user interactions. *PLOS ONE*, 20(5):e0322690.

Bisbee, J., Clinton, J. D., Dorff, C., Kenkel, B., and Larson, J. M. (2024). Synthetic replacements for human survey data? the perils of large language models. *Political Analysis*, 32(4):401–416.

Brasil (2018). Lei n. 13.709, de 14 de agosto de 2018: Lei geral de proteção de dados pessoais. Diário Oficial da União.

Hassani, S., Sabetzadeh, M., Amyot, D., and Liao, J. (2024). Rethinking legal compliance automation: Opportunities with large language models. In *Proceedings of the 32nd IEEE International Requirements Engineering Conference*, pages 432–440. IEEE.

Korn, A., Gorsch, S., and Vogelsang, A. (2025). LLMREI: Automating requirements elicitation interviews with LLMs. In *Proceedings of the 33rd IEEE International Requirements Engineering Conference*, pages 19–30.

National Institute of Standards and Technology (2020). NIST privacy framework: A tool for improving privacy through enterprise risk management, version 1.0. Technical report, NIST, Gaithersburg, MD.

Paduraru, C., Cristea, R., and Stefanescu, A. (2024). Adaptive questionnaire design using AI agents for people profiling. In *Proceedings of the 16th International Conference on Agents and Artificial Intelligence*, volume 3, pages 633–640.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., and Wesslén, A. (2012). *Experimentation in Software Engineering*. Springer, Berlin, Heidelberg.