Barreiras Discursivas: Uma Análise da Toxicidade em Interações de Comunidades de TI no Reddit

Maicon Silva[1], José Iago Lima[1], Douglas Araújo[1]
Francisco Victor da S. Pinheiro[1], Marcelo Martins da Silva[1]

[1]Universidade Federal do Ceará (UFC)
Av. José de Freitas Queiroz, 5003 – Cedro – Quixadá – Ceará 63902-580 – Brazil

{maicon.silvadonascimento, joseiagolima, dougaraujo}@alu.ufc.br

{victorpinheiro, mmartins}@ufc.br

Resumo. *A troca de conhecimento em comunidades de Tecnologia da Informação (TI) no Reddit é frequentemente prejudicada por comportamentos hostis que desestimulam a colaboração. Este artigo investiga a presença de "barreiras discursivas"nessas interações. Utilizando técnicas de Mineração de Dados e Processamento de Linguagem Natural (PLN), o estudo aplica um modelo pré-treinado de classificação de toxicidade para detectar e categorizar automaticamente comentários tóxicos. A análise busca compreender como esses padrões linguísticos impactam a dinâmica social e a difusão de conhecimento, oferecendo uma perspectiva de Sistemas Colaborativos sobre a sustentabilidade de comunidades técnicas online.*
Palavras-chave: *Reddit; Toxicidade Discursiva; Comunidades de Tecnologia da Informação; Processamento de Linguagem Natural; Detecção de Toxicidade; CSCW.*

Abstract. *Knowledge exchange in Information Technology (IT) communities on Reddit is often hindered by hostile behaviors that discourage collaboration. This paper investigates the presence of "discursive barriers"within these interactions. Using Data Mining and Natural Language Processing (NLP) techniques, the study applies a pre-trained toxicity classification model to automatically detect and categorize toxic comments. The analysis aims to understand how these linguistic patterns impact social dynamics and knowledge diffusion, offering a Computer-Supported Cooperative Work (CSCW) perspective on the sustainability of online technical communities.*
Keywords: *Reddit; Discursive Toxicity; Information Technology Communities; Natural Language Processing; Toxicity Detection; CSCW.*

1. Introdução

As plataformas de redes sociais consolidaram-se como espaços vitais para a construção e validação coletiva de conhecimento. Especialmente na área de Tecnologia da Informação (TI), o desenvolvimento profissional não ocorre de forma isolada; ele depende intrinsecamente de "comunidades de prática"onde o aprendizado é contínuo e colaborativo. Nesses ambientes, desenvolvedores e profissionais de tecnologia buscam mentoria, compartilham soluções para erros complexos, discutem tendências de mercado e oferecem suporte emocional e técnico a seus pares.

Nesse contexto, o Reddit destaca-se como uma plataforma singular e possui características que o diferenciam de sites de Q&A estritos, como o Stack Overflow, ou redes sociais generalistas. A estrutura do Reddit, baseada em comunidades temáticas (subreddits) e discussões em formato de *threads* aninhadas, favorece debates aprofundados e narrativas longas, indo além da simples troca de pergunta e resposta. Além disso, o pseudonimato e o sistema de votação (*upvotes/downvotes*) criam uma dinâmica única de visibilidade e reputação, permitindo que os usuários expressem opiniões sinceras sobre a cultura de trabalho e tecnologias, embora isso também possa facilitar comportamentos hostis.

Contudo, a qualidade dessa troca colaborativa depende diretamente de um ambiente social saudável. A problemática surge quando esse espaço de cooperação é afetado por barreiras discursivas, como comportamentos tóxicos, hostilidade e elitismo. No âmbito da TI, tais barreiras manifestam-se frequentemente por meio de respostas desdenhosas a iniciantes (o chamado *gatekeeping*), linguagem agressiva em debates técnicos e posturas que desencorajam a participação [Campos 2009]. Esse cenário revela um paradoxo: plataformas projetadas para promover a inteligência coletiva podem, simultaneamente, reforçar dinâmicas de exclusão, minando o próprio propósito colaborativo da comunidade. Compreender esses fenômenos exige mais do que observações pontuais; requer a análise sistemática de grandes volumes de dados para identificar padrões recorrentes de interação. Nesse contexto, técnicas de mineração de dados e análise automatizada tornam-se essenciais para investigar como a cultura técnica e os mecanismos de incentivo da plataforma — como pontuações e visibilidade — influenciam o comportamento discursivo [Brugnera et al. 2024, Franco 2018].

Este trabalho propõe uma análise automatizada da toxicidade em comunidades de TI no Reddit, utilizando técnicas avançadas de Processamento de Linguagem Natural (PLN). Diferentemente de abordagens tradicionais baseadas em palavras-chave, são empregados Grandes Modelos de Linguagem (LLMs) com classificação *zero-shot*, capazes de detectar nuances semânticas relacionadas a hostilidade, sarcasmo e comportamentos excludentes. O objetivo central é identificar e categorizar barreiras discursivas que comprometem a colaboração e a difusão do conhecimento técnico, evidenciando como tecnologias modernas podem contribuir para o entendimento das dinâmicas sociais em ambientes digitais.

O uso de redes sociais cresceu significativamente nos últimos anos. Segundo estimativas recentes Priorino et al. (2025) , existem 5,5 bilhões de usuários ativos na Internet, dos quais 5,24 bilhões utilizam alguma forma de mídia social, representando aproximadamente 64% da população mundial. Nesse cenário, o Reddit [Medvedev et al. 2019], inserido nesse contexto, tornou-se um *hub* central para o compartilhamento de experiências. Entretanto, essas interações envolvem tanto aspectos positivos quanto negativos [Piorino et al. 2025]. Um estudo realizado pelo *Pew Research Center* em 2020 indica que 41% dos adultos norte-americanos já sofreram algum tipo de assédio online, sinalizando a amplitude do problema

A toxicidade, entendida como linguagem desrespeitosa, hostil ou prejudicial, tem sido amplamente discutida na literatura sobre comunidades online. Compreender padrões de comportamento nesses ambientes é crucial para explicar como o pensamento coletivo emerge em grupos sociais [Guimarães et al. 2020]. Em *subreddits* voltados à tecnologia,

observa-se que a intensa troca de conhecimento técnico convive com interações marcadas por agressividade, comprometendo a qualidade das discussões e o senso de pertencimento dos participantes.

Estudos específicos sobre comunidades de engenharia de software reforçam essa problemática. Pesquisas recentes de Paiva (2024) apontam a recorrência de comentários hostis em debates sobre práticas profissionais e cultura organizacional em empresas de TI. Trabalhos brasileiros sobre rotulação e análise de conteúdo tóxico também ressaltam desafios associados à moderação, à propagação de discursos nocivos e à subjetividade da classificação humana. Assim, ao investigar *subreddits* de tecnologia, este estudo busca ampliar discussões sobre toxicidade em comunidades digitais, destacando como a cultura técnica influencia — e é influenciada por — a forma como indivíduos interagem nesses espaços.

2. Referencial Teórico

2.1. Mineração de Dados em Redes Sociais e Sistemas Colaborativos

A mineração de dados em redes sociais consolidou-se como uma área fundamental para a extração de conhecimento em larga escala, permitindo a identificação de padrões comportamentais que não seriam visíveis em análises manuais [Aggarwal 2011]. No contexto de sistemas colaborativos, onde o valor é gerado pela interação entre usuários, a mineração permite mapear a estrutura dessas interações e a qualidade do conteúdo gerado.

O Reddit exemplifica esse cenário como um ecossistema de *"sabedoria das multidões"* [Medvedev et al. 2019]. Estruturado em comunidades temáticas (*subreddits*), que, diferente de redes sociais genéricas, tem foco na construção coletiva de conhecimento sobre tópicos específicos. Portanto, aplicar técnicas de mineração nesse ambiente é primordial não apenas para coletar dados, mas para entender como a colaboração emerge ou é inibida por fatores sociais dentro dessas comunidades.

2.2. O Reddit como Ferramenta de Gestão de Conhecimento Técnico

O Reddit transcende a função de entretenimento, atuando como um repositório dinâmico de conhecimento técnico. Sua estrutura organizacional, baseada na folksonomia (criação de tags e comunidades pelos próprios usuários) e em mecanismos de *Crowdsourcing*[1] de qualidade (o sistema de *upvotes* e *downvotes* e o *score*, que é a diferença numérica entre esses dois), favorece a validação coletiva da informação [Medvedev et al. 2019]. Em comunidades de TI, esse mecanismo de ranqueamento, ou como é conhecido, *score*, funciona como uma revisão por pares informal: respostas tecnicamente precisas tendem a ganhar visibilidade, enquanto informações incorretas ou ruídos são filtrados. Para este trabalho, o Reddit é analisado sob a ótica de um sistema sócio-técnico, onde a eficiência na resolução de problemas depende diretamente da saúde das interações entre os membros [Proferes et al. 2021].

2.3. Processamento de Linguagem Natural e Detecção de Toxicidade

O Processamento de Linguagem Natural papel central na análise da qualidade do discurso em ambientes online. Técnicas de classificação textual possibilitam a detecção

[1] https://ideascale.com/pt-br/blogue/o-que-e-crowdsourcing/

de discursos tóxicos, agressivos ou potencialmente prejudiciais, contribuindo para a avaliação da qualidade da interação entre usuários. Técnicas clássicas de classificação textual evoluíram para detectar não apenas palavras ofensivas, mas nuances de discursos tóxicos, agressivos ou excludentes que minam a coesão do grupo [Wulczyn et al. 2017]. Em comunidades técnicas e outras comunidades em geral, tal análise torna-se especialmente relevante, uma vez que comportamentos hostis podem impactar negativamente a participação de novos membros e a difusão de conhecimento nessas comunidades.

2.4. Modelos de Linguagem e Classificação *Zero-Shot*

A evolução dos LLMs permitiu a realização de tarefas complexas de compreensão textual sem a necessidade de extensos conjuntos de dados anotados. A abordagem *zero-shot*, utilizada nesta pesquisa, permite classificar a toxicidade baseando-se em instruções semânticas e raciocínio lógico do modelo, em vez de apenas padrões estatísticos(fixos) de palavras. Isso é especialmente aplicável a este trabalho, pois a toxicidade em TI muitas vezes é sutil (ex: condescendência técnica ou elitismo) e difícil de capturar com léxicos tradicionais. Ferramentas que viabilizam a execução local desses modelos, como aquelas baseadas em quantização, ampliam o acesso à inferência de alto desempenho e favorecem a implementação de soluções independentes de serviços externos.

2.5. Dinâmica Social em Comunidades de Software

As comunidades técnicas online, como as de Engenharia de Software, apresentam padrões específicos de interação, influenciados por normas meritocráticas e, por vezes, competitivas como a experiência dos participantes, normas coletivas e objetivos colaborativos, aponta que a colaboração nesses ambientes é sensível a fatores humanos: a forma como perguntas são respondidas impacta diretamente a preservação de membros e o engajamento. Analisar essas dinâmicas é essencial para entender o fenômeno da toxicidade não como casos isolados, mas como um padrão sistêmico que pode degradar o ambiente colaborativo. A análise dessas dinâmicas contribui para compreender como surgem conflitos, como determinados tipos de discurso se propagam e de que forma o ambiente social pode interferir no engajamento, na aprendizagem e na retenção de membros em fóruns orientados ao compartilhamento de conhecimento.

3. Trabalhos Relacionados

Esta seção apresenta e discute os principais estudos relacionados à análise de toxicidade em interações online, com ênfase em comunidades digitais, redes sociais e ambientes colaborativos. O objetivo é situar o presente trabalho no contexto da literatura, destacando abordagens, métodos e resultados de pesquisas anteriores que investigam comportamentos hostis, discurso tóxico e seus impactos sobre o engajamento e a dinâmica social dos participantes. A partir dessa revisão, busca-se evidenciar a lacuna que este estudo pretende explorar, especialmente no que diz respeito à análise da toxicidade como barreira discursiva em comunidades de TI no Reddit.

3.1. Rotulação e Caracterização de Conteúdo Tóxico de Comunidades do Reddit no Brasil.

O artigo descreve um experimento em que foi construído um conjunto de dados anotados manualmente de comentários das 10 maiores comunidades brasileiras do Reddit. O estudo

apresenta uma análise linguística das manifestações de toxicidade, identificando padrões discursivos característicos e avaliando modelos automáticos de classificação, como a API Perspective e LLMs.

Assim como o trabalho mencionado, o presente estudo investiga fenômenos relacionados à toxicidade em interações online no Reddit, empregando técnicas de PLN para analisar comportamentos discursivos. Contudo, enquanto o artigo enviado concentra-se exclusivamente no ecossistema brasileiro do Reddit e em textos produzidos em língua portuguesa, este trabalho amplia o escopo de análise para comunidades globais de TI. Essas comunidades são predominantemente em língua inglesa e inseridas em *subreddits* de alcance internacional. Além disso, o estudo relacionado prioriza a construção de um conjunto de dados rotulado e a descrição sociolinguística da toxicidade no Brasil; por sua vez, o presente artigo enfatiza a investigação de como a toxicidade opera como barreira discursiva em comunidades técnicas, afetando processos de colaboração, engajamento e circulação de conhecimento entre os mais diversos níveis de *senioridade* dos membros da comunidade.

A principal distinção deste trabalho em relação ao estudo analisado reside no foco em comunidades de TI e na exploração da toxicidade como mecanismo de exclusão discursiva em ambientes altamente especializados. Enquanto o artigo relacionado aborda comunidades generalistas e discute a toxicidade sob uma perspectiva sociolinguística nacional, o presente estudo concentra-se nas dinâmicas comunicativas específicas do domínio técnico, caracterizadas por jargões, debates especializados e interações orientadas à resolução de problemas. Além disso, este trabalho aplica técnicas *Natural Language Processing* (NLP), voltadas à análise de dados multilíngues e internacionalizados, permitindo identificar padrões de toxicidade que impactam a participação de usuários em comunidades globais de TI. Dessa forma, a contribuição aqui apresentada complementa e expande a literatura ao examinar efeitos práticos da toxicidade na formação de barreiras discursivas em contextos técnicos internacionais.

3.2. Assédio e Comportamento Hostil em Plataformas Digitais

O relatório do *Pew Research Center* (2021) apresenta estatísticas amplas sobre experiências de assédio online entre adultos nos Estados Unidos. Os resultados indicam que 41% dos entrevistados já sofreram algum tipo de assédio e que 75% dos casos mais recentes ocorreram em redes sociais. O relatório diferencia agressões leves — como xingamentos — de agressões mais severas, como perseguição e ameaças físicas, além de discutir motivações como política, gênero e raça.

Esse panorama reforça a relevância social do presente estudo ao demonstrar que comportamentos tóxicos são comuns em ambientes digitais. As formas de agressão mapeadas pelo relatório se alinham às categorias analisadas nas comunidades de TI investigadas neste trabalho. Contudo, há distinções importantes: o relatório baseia-se em dados perceptivos e macrosseriais de uma população ampla, enquanto esta pesquisa adota uma análise técnica e detalhada de dados reais do Reddit, focando no modo como a toxicidade se manifesta discursivamente.

3.3. Ética e Comportamento Tóxico nas Comunidades Virtuais

Diversos estudos têm explorado a toxicidade em ambientes virtuais e os impactos que eles podem causar nas interações sociais. Paiva (2024) analisou interações em fóruns de

engenharia de software no Reddit. O estudo buscou investigar como comentários ofensivos emergem em comunidades técnicas e discutiu as implicações que eles podem causar no engajamento dessas comunidades de Engenharia de Software. O autor concluiu que, embora a toxicidade esteja presente, o engajamento das publicações não foi significativamente afetado, sugerindo que fatores culturais e contextuais podem influenciar a uma certa tolerância de interações hostis.

De forma complementar, Guimarães et al. (2020), no estudo *Characterizing Toxicity on Facebook Comments in Brazil*, realizaram uma analise em larga escala de comentários associados a notícias durante alguns eventos políticos relevantes na sociedade. Utilizando a API Perspective, os autores identificaram padrões de toxicidade associados a essas notícias, e, destacaram como figuras públicas desencadeiam tais comportamentos. Esse estudo evidencia que a toxicidade não apenas permeia discussões políticas, mas também molda a percepção pública em ambientes digitais.

Apesar das contribuições desses trabalhos, observa-se uma lacuna na literatura quanto à comparação entre diferentes tipos de comunidades virtuais e como a toxicidade afeta o engajamento e a ética das interações. O presente artigo visa avançar nessa direção propondo uma análise integrada que considere aspectos culturais e contextuais dessas comunidades.

A Tabela 1 sintetiza os principais estudos relacionados, destacando diferenças de escopo, técnicas e objetivos.

Tabela 1. Comparação entre estudos relacionados

Aspectos	**Toxicidade e Gatilhos (2020/2025)**	**Relatório *Pew Research* (2021)**	**Paiva (2024)**	**Guimarães et al. (2020)**
Objeto de Estudo	Comunidades do Reddit Brasil (análise de toxicidade e gatilhos)	Assédio online nos EUA (adultos americanos)	Comunidades de Engenharia de Software no Reddit	Comentários em notícias políticas no Facebook
Técnicas de Avaliação	Análise linguística, API Perspective e LLMs	*Survey* (*American Trends Panel*)	Análise de conteúdo qualitativa	Análise em larga escala com API Perspective
Participantes	Usuários de comunidades brasileiras do Reddit	Amostra aleatória estratificada	Usuários e autor do estudo	Usuários comentadores
Objetivo	Analisar padrões discursivos de toxicidade e avaliar modelos automáticos	Investigar experiências e padrões de assédio online	Analisar emergência de comentários ofensivos e impacto no engajamento	Investigar toxicidade em eventos políticos e influência de figuras públicas

Observa-se que a maioria das pesquisas foca em contextos amplos de interação social, como assédio online e discussões políticas, enquanto estudos em comunidades técnicas ainda exploram de forma limitada os impactos da toxicidade na colaboração. Nesse contexto, este trabalho diferencia-se ao analisar a toxicidade como uma barreira discursiva em comunidades de TI, integrando técnicas de Processamento de Linguagem Natural com uma perspectiva de Sistemas Colaborativos.

4. Metodologia

Esta Seção descreve os procedimentos metodológicos adotados neste estudo, incluindo a coleta e o pré-processamento dos dados, a definição do modelo de classificação de toxi-

cidade e os critérios de análise empregados. O objetivo é garantir a reprodutibilidade da abordagem, detalhando as etapas desde a preparação dos dados até a inferência automatizada e a interpretação dos resultados no contexto das comunidades analisadas.

4.1. Conjunto de dados e Pré-processamento

O conjunto de dados foi composto por comentários extraídos de nove comunidades de TI da plataforma *Reddit*, selecionadas por sua relevância tecnológica a nível global que são: *r/webdev, r/programming, r/programminghorror, r/computers, r/Python, r/computerscience, r/developersIndia, r/technology e r/AskProgramming*.

Para garantir a integridade da análise e a eficiência computacional, os dados brutos foram submetidos a um rigoroso protocolo de limpeza e normalização:

1. **Tratamento de Vacuidade:** Valores nulos foram preenchidos com strings vazias e submetidos a uma conversão explícita de tipo para garantir a compatibilidade com o modelo de linguagem.
2. **Filtragem de Metadados Irrelevantes:** Foram removidos registros contendo os marcadores automáticos [*deleted*] e [*removed*], que indicam intervenção prévia da moderação ou do usuário; portanto, não tinham conteúdo textual passível de análise.
3. **Truncamento e Normalização:** Os comentários foram limitados aos primeiros 512 caracteres (removendo os espaços em branco excedentes). Essa etapa é crítica, pois modelos baseados em *Transformers* possuem limites intrínsecos de *tokens* para processamento.
4. **Critério de Comprimento Mínimo:** Comentários com menos de três caracteres foram descartados para evitar ruídos estatísticos e falsos positivos oriundos de onomatopeias ou caracteres isolados.

4.2. Modelo de Inferência: *Detoxify (unbiased)*

A classificação de toxicidade foi realizada por meio da biblioteca *Detoxify*, que utiliza a arquitetura *Bidirectional Encoder Representations from Transformers* (BERT). Pelo fato de os dados serem em língua inglesa, usamos o modelo *"unbiased"*, que tem uma eficácia maior na língua inglesa e foi treinado especificamente para mitigar vieses algorítmicos relacionados a termos de identidade (raça, religião, orientação sexual), garantindo que a classificação foque na agressividade do conteúdo e não apenas em palavras-chave sensíveis. O processo de inferência foi otimizado por meio de processamento em lote (*batch processing*), permitindo que o modelo predissesse a toxicidade de múltiplos comentários simultaneamente, otimizando o uso de memória e tempo de processamento

4.3. Critério de Classificação e Exportação

O modelo retorna um score de probabilidade contínuo no intervalo $[0, 1]$. Para a categorização binária dos dados, estabeleceu-se um limiar crítico (*threshold*) de 0,5, no qual comentários com $score \geq 0,5$ foram classificados como tóxicos, enquanto comentários com $score < 0,5$ foram classificados como seguros. Os dados classificados foram consolidados em novos arquivos estruturados em formato CSV, codificados em utf-8-sig para garantir a preservação de caracteres especiais e a portabilidade para ferramentas de análise estatística.

4.4. Análise Contextual e Relacionamento com CSCW

Para estabelecer a relação entre os índices de toxicidade detectados e as características de colaboração das comunidades, foi feita uma abordagem analítica comparativa baseada no propósito do *subreddit*. As comunidades selecionadas foram categorizadas qualitativamente em três perfis de colaboração distintos:

1. **Comunidades de Suporte/Aprendizado (ex: *r/AskProgramming*):** Onde a colaboração é definida pela troca direta de dúvidas e soluções (Q&A).
2. **Comunidades de Discussão Geral (ex: *r/technology*):** Focadas em notícias e opiniões, onde a colaboração ocorre via debate de ideias.
3. **Comunidades de Crítica/Humor (ex: *r/programminghorror*):** Focadas no compartilhamento de experiências negativas ou falhas, onde a interação serve para reforço de identidade de grupo.

As comunidades foram categorizadas qualitativamente com base em seus objetivos e dinâmicas de interação, seguindo princípios de estudos em CSCW [Kraut and Resnick 2012] e comunidades de prática [Wenger 1999], que indicam que diferentes propósitos comunitários influenciam diretamente os padrões de colaboração e comportamento.

A relação com a colaboração foi inferida sob a ótica de Sistemas Colaborativos da seguinte forma: a presença de toxicidade em comunidades do tipo (1) é interpretada como uma barreira de entrada (*gatekeeping*), pois viola a expectativa de acolhimento necessária para o aprendizado. Já em comunidades do tipo (3), a toxicidade foi analisada como potencial manifestação de elitismo técnico, onde a hostilidade contra o código alheio atua como um mecanismo de validação de senioridade. Dessa forma, a toxicidade não é tratada apenas como ”ofensa”, mas como um ruído comunicacional que altera a dinâmica de transferência de conhecimento dependendo do contexto onde ocorre.

5. Resultados e Discussão

Nesta seção, apresentam-se os dados obtidos após o processamento de 862 comentários distribuídos entre os nove *subreddits* selecionados. A análise foca na prevalência de interações hostis, na relação entre engajamento e toxicidade, e nas características estruturais das mensagens coletadas.

Após a aplicação do modelo *Detoxify*, observou-se que a grande maioria das interações nas comunidades de tecnologia é classificada como “Segura”. Do total de amostras processadas, apenas 2,90% (25 comentários) foram rotuladas como tóxicas, enquanto 97,10% mantiveram-se abaixo do limiar crítico de 0,5.

O volume de dados por comunidade é ilustrado na Figura 1. Embora a coleta tenha sido equilibrada entre os *subreddits* (variando entre 77 e 100 comentários por nicho), a proporção de conteúdo tóxico (representada em vermelho) é visualmente marginal. Isso sugere que o tom predominante nas discussões técnicas é majoritariamente informativo ou neutro, com o conflito aparecendo de forma pontual.

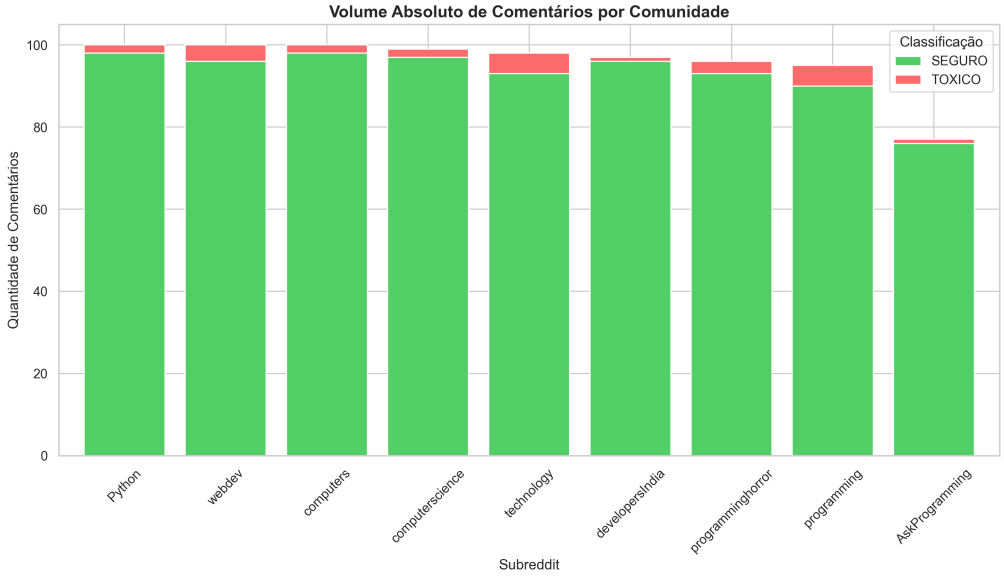

Figura 1. Volume Absoluto de Comentários por Comunidade e sua Classificação.

A análise individualizada revelou disparidades significativas entre as comunidades analisadas, conforme detalhado na Tabela 2. Essas diferenças sugerem que a incidência de toxicidade não é homogênea, variando de acordo com o propósito e a dinâmica de interação de cada *subreddit*. Em particular, comunidades orientadas à discussão geral e opinião tendem a apresentar maiores níveis de hostilidade, enquanto aquelas focadas em suporte técnico e resolução de problemas exibem padrões mais colaborativos e menos propensos à ocorrência de comportamentos tóxicos.

Tabela 2. Distribuição de Toxicidade por Comunidade (*Subreddit*)

Subreddit	**Total de Comentários**	**Qtd. Tóxicos**	**Taxa de Toxicidade (%)**
r/AskProgramming	77	1	1,03%
r/developersIndia	97	1	1,03%
r/computerscience	99	2	2,02%
r/computers	100	2	2,00%
r/Python	100	2	2,00%
r/programminghorror	96	3	3,12%
r/webdev	100	4	4,00%
r/technology	98	5	5,10%
r/programming	95	5	5,26%
MÉDIA GERAL	**862**	**25**	**2,90%**

Fonte: Elaborada pelo autor.

Os dados indicam que comunidades com viés crítico ou de notícias gerais apresentam maior incidência de hostilidade. O *r/programming* (5,26%) e o *r/technology* (5,10%)

lideram o ranking. No caso do *r/technology*, a natureza generalista atrai discussões que transcendem o código, abordando política e ética, temas naturalmente mais polarizados. Em contrapartida, o *r/AskProgramming* (1,03%) demonstrou o menor índice, reforçando que o propósito pedagógico de ajuda mútua inibe comportamentos tóxicos.

Um achado relevante desta pesquisa refere-se à relação entre o tom das mensagens e seu nível de engajamento, mensurado por meio de *upvotes*. Conforme ilustrado no *boxplot* da Figura 2, observa-se uma diferença clara na distribuição das pontuações entre conteúdos classificados como tóxicos e seguros. Esse padrão sugere que a natureza do discurso pode influenciar diretamente a visibilidade das interações, indicando que mensagens com maior carga emocional — incluindo aquelas de teor hostil — tendem a gerar maior variabilidade no engajamento, possivelmente devido à polarização ou ao estímulo a debates mais intensos dentro das comunidades.

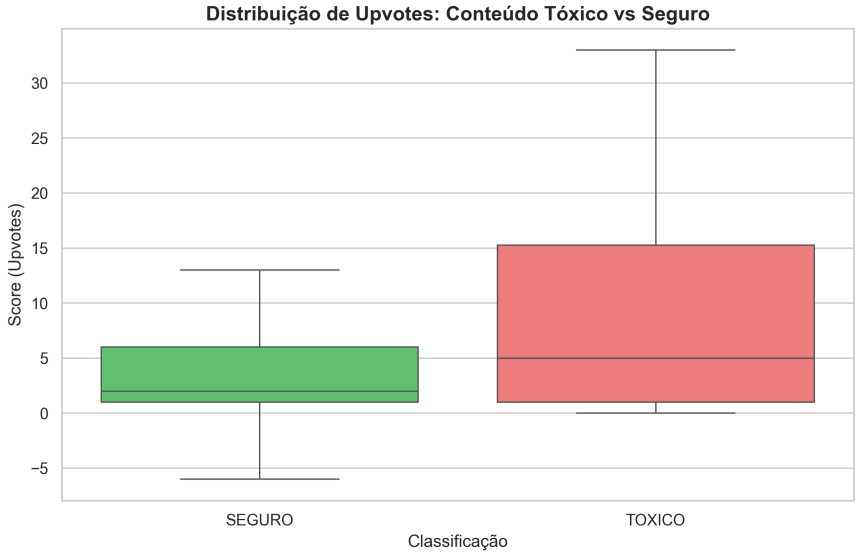

Figura 2. Distribuição de *Upvotes*: Conteúdo Tóxico vs. Seguro.

Enquanto o conteúdo seguro possui uma mediana de *upvotes* baixa e estável, o conteúdo tóxico apresenta uma dispersão significativamente maior. Nota-se que o limite superior das mensagens tóxicas atinge mais de 30 *upvotes*, sugerindo que comentários agressivos ou sarcásticos podem gerar um engajamento atípico, seja por identificação dos usuários com a crítica ou pelo desencadeamento de discussões acaloradas.

Para investigar se a toxicidade nas interações analisadas se manifesta de forma mais breve ou mais elaborada, analisou-se a densidade do tamanho das mensagens em número de caracteres, conforme apresentado na Figura 3. Essa análise busca verificar se comentários tóxicos tendem a assumir a forma de respostas curtas e impulsivas ou se, ao contrário, aparecem em mensagens mais extensas, estruturadas e argumentativas. Tal distinção é relevante, pois permite compreender não apenas a presença da toxicidade, mas também a forma como ela se materializa discursivamente nas comunidades de TI.

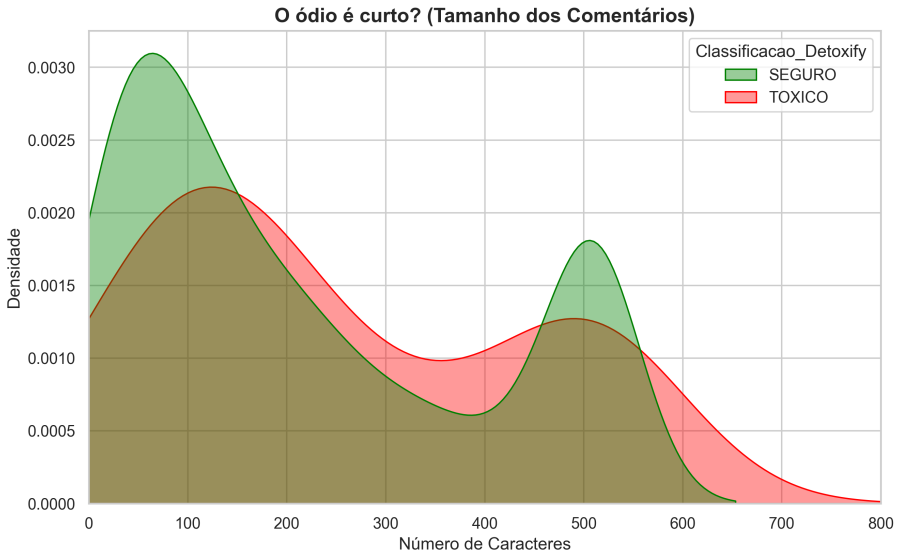

Figura 3. Análise de Densidade: Tamanho dos Comentários.

Diferente do esperado, o conteúdo tóxico (em vermelho) apresenta uma cauda mais longa à direita, estendendo-se até os 800 caracteres. Isso indica que, no contexto de comunidades de TI, a toxicidade não se resume a insultos curtos, mas frequentemente se manifesta em críticas extensas, refutações agressivas ou correções de tom pedante. O conteúdo seguro, por outro lado, apresenta picos bimodais (próximos a 100 e 500 caracteres), refletindo a natureza de respostas técnicas diretas ou tutoriais explicativos.

6. Conclusão

O presente estudo investigou a dinâmica da toxicidade em comunidades de TI no Reddit, com o objetivo de identificar se a cultura técnica promove ”barreiras discursivas”que dificultam a colaboração. Utilizando uma abordagem baseada em Mineração de Dados e PLN, aplicou-se o modelo *Detoxify* em uma arquitetura *Zero-Shot* para classificar a hostilidade em *subreddits* globais. A metodologia provou-se eficaz ao processar interações em larga escala, permitindo mapear não apenas a presença de agressividade explícita, mas também diferenciar nuances de linguagem em um domínio repleto de jargões técnicos.

Do ponto de vista quantitativo, os resultados revelaram um cenário heterogêneo. A taxa média global de toxicidade de apenas 2,9% desafia a percepção de que ambientes online são sistemicamente hostis. Observou-se que comunidades focadas estritamente em suporte e aprendizado (como r/AskProgramming) tendem a manter níveis baixos de agressividade, sugerindo que o foco na resolução objetiva de problemas atua como um regulador social. Por outro lado, a identificação de picos de toxicidade em comunidades de humor depreciativo ou crítica (como r/programminghorror) evidencia nichos onde a hostilidade é normalizada como parte da cultura de entretenimento técnico.

A toxicidade em TI não se manifesta meramente como insulto genérico, mas frequentemente assume a forma de elitismo técnico e condescendência. Mesmo que estatisticamente minoritários, esses comportamentos atuam como barreiras discursivas significativas, funcionando como mecanismos de *gatekeeping* que intimidam iniciantes e validam

a senioridade através da humilhação do erro alheio. A persistência dessas barreiras sugere que ferramentas de moderação baseadas apenas em palavras ofensivas são insuficientes para capturar a sutileza do desestímulo profissional ao ser levado em consideração que a toxicidade não é apenas uma ofensa pessoal, mas um ruído que degrada a dinâmica de transferência de conhecimento entre os membros.

Embora os resultados obtidos ofereçam evidências relevantes sobre a presença de barreiras discursivas em comunidades de TI no Reddit, este estudo apresenta algumas limitações que devem ser consideradas. Primeiramente, o volume de dados analisado é relativamente restrito, o que limita a generalização dos achados para outras comunidades, períodos ou plataformas digitais. Além disso, a classificação de toxicidade foi realizada com base em um modelo pré-treinado, sem validação manual de uma amostra anotada especificamente para o contexto técnico investigado, o que pode afetar a precisão da interpretação em casos de sarcasmo, ironia ou jargão especializado. Por fim, a análise adotada possui caráter exploratório, de modo que os resultados devem ser compreendidos como indícios relevantes para discussão dentro do escopo de *Computer-Supported Cooperative Work*, e não como conclusões definitivas sobre o comportamento discursivo em comunidades técnicas online.

Como ideia para pesquisas posteriores, sugere-se expandir a investigação para além da classificação binária (tóxico/seguro), empregando grandes modelos de linguagem generativos para categorizar tipologias específicas de barreiras, como arrogância intelectual ou sarcasmo técnico. Além disso, estudos longitudinais seriam fundamentais para correlacionar a exposição a essa toxicidade com a taxa de evasão de usuários, permitindo quantificar o impacto real da hostilidade na sustentabilidade e no capital social das comunidades de prática de software.

Declaração do Uso de Inteligência Artificial

Declaro para fins que esse texto usou IA para revisão gramatical, adequação estilística, formatação em LaTeX e auxílio na tradução do *abstract*

Referências

Aggarwal, C. C. (2011). An introduction to social network data analytics. In *Social network data analytics*, pages 1–15. Springer.

Boulic, R. and Renault, O. (1991). 3d hierarchies for animation. In Magnenat-Thalmann, N. and Thalmann, D., editors, *New Trends in Animation and Visualization*. John Wiley & Sons ltd.

Brugnera, E. D., Pedra, R. R., Sousa, D. B., Monge, A. R., de Andrade, T. d. S. C., and Antunes, E. J. (2024). Aprendizado colaborativo: A força da comunidade online. *ARACÊ*, 6(3):5736–5749.

Campos, A. d. (2009). Conflitos na colaboração: um estudo das tensões em processos de escrita coletiva na web 2.0.

Franco, A. H. C. (2018). Inteligência coletiva: manifestações nos ambientes digitais.

Guimarães, S. S., Reis, J. C., Ribeiro, F. N., and Benevenuto, F. (2020). Characterizing toxicity on facebook comments in brazil. In *Proceedings of the Brazilian Symposium on Multimedia and the Web*, pages 253–260.

Knuth, D. E. (1984). *The TEX Book*. Addison-Wesley, 15th edition.

Kraut, R. E. and Resnick, P. (2012). *Building successful online communities: Evidence-based social design*. Mit Press.

Medvedev, A. N., Lambiotte, R., and Delvenne, J. C. (2019). The anatomy of reddit: An overview of academic research. In Ghanbarnejad, F., Saha Roy, R., Karimi, F., Delvenne, J. C., and Mitra, B., editors, *Dynamics On and Of Complex Networks III*, Springer Proceedings in Complexity. Springer, Cham.

Paiva, E. V. O. d. (2024). *Ética e comportamento tóxico em comunidades virtuais de engenharia de software no Reddit*. PhD thesis, Pontifícia Universidade Católica de Minas Gerais.

Piorino, G., Lima, L., Pagano, A., and Silva, A. (2025). Toxicidade e gatilhos: Um estudo de caso em comunidades do reddit no brasil. In *Proceedings of the 31st Brazilian Symposium on Multimedia and the Web*, pages 575–579, Porto Alegre, RS, Brasil. SBC.

Proferes, N., Jones, N., Gilbert, S., Fiesler, C., and Zimmer, M. (2021). Studying reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media+ Society*, 7(2):20563051211019004.

Smith, A. and Jones, B. (1999). On the complexity of computing. In Smith-Jones, A. B., editor, *Advances in Computer Science*, pages 555–566. Publishing Press.

Wenger, E. (1999). *Communities of practice: Learning, meaning, and identity*. Cambridge university press.

Wulczyn, E., Thain, N., and Dixon, L. (2017). Ex machina: Personal attacks seen at scale. In *Proceedings of the 26th international conference on world wide web*, pages 1391–1399.