チャットツールにおける動的表示遅延を用いたネットいじめ削減手法の提案

川口尊琉[†1]　神場知成[†2]

概要：ソーシャルメディア上のネットいじめは，フィルタリングや警告メッセージといった既存の技術的対策にもかかわらず依然として深刻な課題である．本研究では，攻撃的なメッセージの送信に対して動的な表示遅延を適用することで，衝動的な有害発言を抑制しエスカレーションを遅らせるインタラクション手法を提案する．提案システムは，会話コンテクストリスクの変化を大規模言語モデル（ChatGPT API）で計算するとともに，高リスク語彙リストを用いた評価結果に応じて 5〜25 秒の表示遅延を加える．動作シミュレーションを目的とし，大規模言語モデルを用いて軽度・中度・重度の攻撃性を持つ模擬チャットログを生成し，可視化ツールを開発し，それらに「遅延なし / 発話者のみの表示遅延 / 全参加者の表示遅延」 の 3 手法を適用したときの影響を観察した．参加者全員の表示を遅延させるグローバル遅延 では会話の流動性を低下させて冷静になる時間を与えることができ，リスクのある発話表示だけを遅延させる個人遅延 では発話順序の崩れや共感発話の消失が生じることが期待できる．表示遅延の挿入は単なる技術的試みではなく，会話の社会的ダイナミクスの再構成とみなすことができる．

1. はじめに

ソーシャルメディアの急速な普及は，特に 10 代から 20 代の若年層におけるコミュニケーション方法を大きく変容させ，彼らの交流の多くがオンライン空間で行われるようになった．この変化は，即時性やアクセスの容易さといった多くの利点をもたらす一方で，同時に新たなリスクも生み出した．その一つは，衝動的かつ攻撃的な発言の増加であり，ネットいじめを助長している．デジタル環境における匿名性，低摩擦性，即時的な送信機能が相まって，社会的抑制が失われ，対面では決して発しないような攻撃的メッセージを送信する傾向が強まっている．これまでにブロックフィルター，自動警告，投稿削除機能といった教育的・技術的介入が提案されてきたが，いずれも包括的な解決策とはなっていない．

そこで本研究では，高リスクメッセージが発見された際に，表示を動的に遅延させるシステムの検討を行った．遅延により再考の時間を発言者もしくは会話の場に与えることで，衝動性を低下させ，攻撃性のエスカレーションを防止できることが期待できる．

2. 関連研究

2.1 ネットいじめの定義づけ

インターネットの発達により発生率が上昇しているネットいじめに関する研究は，心理学，教育学，情報科学など幅広い学術分野にまたがり，多様な研究が行われている．ネットいじめの定義づけについては概念の複雑さや，従来のいじめの定義をデジタル環境に適用する際の課題から，研究者の間でも統一された見解を得ることが困難だとされているが[1]-[4]，一例として Zhang 等は「ネットいじめとは，電子的な接触手段を用いて，集団または個人によって，繰り返し，時間をかけて，容易に自己を守ることができない被害者に対して行われる，攻撃的かつ意図的な行為または行動である」と定義している[5]．また，ネットいじめには，単一の行為であってもオンライン上に無期限に残存してしまうコンテンツの永続性や，無数の人に閲覧または拡散されてしまう反復性も特徴の一つである．これらに加えて，意図の立証の困難さも指摘されており，単なる「意地の悪い」オンラインでのやり取り（不適切な冗談など）と，深刻な被害を引き起こす「意図的で反復的な危害」をいかに区別するかという課題も存在している[6].

2.2 介入方法

Chen 等は，攻撃的言語の高精度な検出を目指すために Lexical Syntactic Feature（LSF）フレームワークを提案している．同手法ではメッセージレベルとユーザレベルの二段階の検出をしている．メッセージレベルでは，語彙的特徴と統語的特徴を統合して文脈依存の攻撃性を推定し，ユーザレベルでは投稿スタイルやコンテンツ特性を含む分析により，従来手法を上回る高精度な検出性能を示している[7].

技術的介入策としては，有害言語を識別する機械学習アルゴリズムの活用，攻撃性コンテンツの自動削除，および不適切な投稿の可能性をユーザに警告するプロンプトの提示などが挙げられるが[8]-[17]，先行研究では習慣化の発生が問題とされている．これはユーザが繰り返し提示される警告に慣れてしまうことで，意図された抑止効果が減衰する現象である[18][19].

心理学的観点からは，認知的停止や遅延が自制心を直接回復させるものではないものの，冷却要因，つまり攻撃的な行動に対する負の結末について考える時間となり攻撃性

†1 東洋大学 ライフデザイン学部
†2 東洋大学 情報連携学部

の抑制に寄与する可能性が示されている[20]-[22]．Warner らはこの概念を「時間的摩擦」として提唱し，プロアクティブ（事前対応型）なモデレーションにおける有効性を示している．彼らは，設計要因（タイミング，摩擦，提示方法）がユーザ体験に与える影響を体系的に検証しており，入力中または送信後に毒性が検出された際に機能を 30 秒無効化する手法を提案した[19]．

これまでの研究は主に「入力停止」や「警告表示」に焦点を当てており，ユーザの文脈やリスクレベルに応じて動的に遅延を調整するインタラクション設計については十分に検討されていない．本研究はこの課題に着目し，ユーザの攻撃性を低下させるための動的遅延システムを検討する．

3. 提案手法

3.1 システム概要

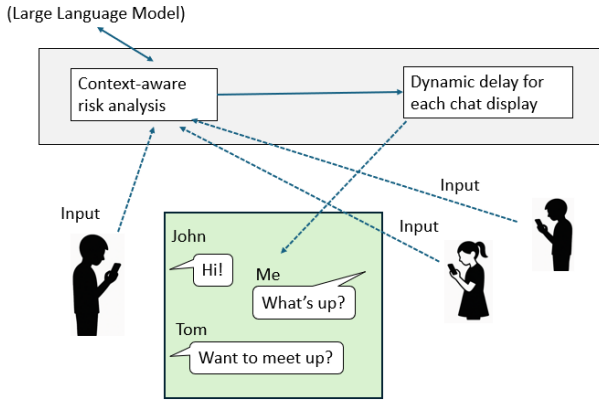

図 1　システム概要

本研究では，心理学的理論を基盤とし，リアルタイム自然言語処理を統合したネットいじめ軽減システムを提案する（図 1）．本システムは，発話者の攻撃的発言にのレベルに応じて動的な遅延を付与するインタラクション設計を中核的特徴とするが，各発話に含まれる攻撃的単語だけでなく，会話全体の流れやトーンを踏まえて，時点ごとの「コンテクストリスク」を推定している．

3.2 コンテクストリスクの推定

本システムの処理は以下の通りである．

① 解析対象となるコンテクスト

解析の単位は，ある時点までに蓄積された会話ログである．各時点 t において，そこまでに登場した全ての発話を [発話者 ID：発話の生テキスト] のペア列としてまとめ，時点 t におけるリスクを判定する．

② 大規模言語モデル（Large Language Model：以下 LLM）による敵対性の評価

LLM に対して次のようなプロンプトを与える：

・会話全体の敵対性を 0〜100 のスコアで評価する

・明示的な暴言や脅迫に加えて，皮肉・嘲笑・軽視・一方的な非難の継続，仲間外れ・排除の示唆，それに対する被害者側の萎縮や謝罪の繰り返しといった文脈的な攻撃性も考慮する

また，プロンプトでは「説明文は出力せず，0〜100 の数値のみ返すこと」と指定し，返ってきた数値をコンテクストリスクの一次的な推定値として用いる．

③ 事前定義したキーワードベースによる補完的評価

LLM が利用できない場合や時間内に応答が取得できない場合を想定し，②の代替として，事前に定義した攻撃的単語リスト（レベル 0〜5）を発話のコンテクストリスク（0〜100）に変換し，それまでの会話の平均値により時刻 t に対して 0〜100 のスコアを得る

④ 時系列としてのリスク変動

上記に加え，リスクの滑らかな変動を考慮して以下の処理を行う：

・新しいリスクスコアが得られるたびに，直前のスコアと新しいスコアから指数移動平均を得る．

・時間の経過に応じて，リスクが徐々に下がるような自然減衰をかけ，新しい攻撃が出てこない状態では，リスクが少しずつ低下させる．

これにより，一瞬だけ攻撃性の高い単語が出たから最大のリスクになるという挙動ではなく，会話全体の雰囲気が徐々に高まるまたは落ちついていくとみなしている．

なお，本システムのユーザは高リスク語彙リストを手動で追加でき，リスト内の各語彙はリスクレベルの再設定が可能である．この機能により，特定のチャット環境に適応したリスク検出精度の向上が期待できる．

3.3 発話への遅延時間設定

上記で算出したコンテクストリスクにもとづいて，各発言に表示までの遅延時間を設定する．遅延時間は判定された攻撃レベルに応じて 5〜25 秒の範囲で動的に設定される．現時点では LLM 応答速度の限界から，主としてキーワードによる遅延設定が優先される場合が多い．

4. 動作例

4.1 会話ログの生成

動的遅延が会話の時間構造および攻撃的発言のエスカレーションに与える影響を検討することを目的に実験を行った．まず，模擬的なオンライン対話を生成するため，LLM を用いて軽度・中度・重度の攻撃性を反映したチャットログを作成した．これは，実際に被験者を集めて意図的にいじめを含む対話をするように指示することは倫理的にできず，そのような既存の対話ログとして公開されているものも入手困難と判断したからである．

各ログは仮想参加者 3 人による約 20 回のやり取りで構成されるものとし，以下の条件で作成した．

・軽度: 軽い皮肉や否定的コメントを含む

・中度: 直接的な侮辱語，批判的発言を含む

・重度: 露骨な暴言，脅迫的表現，強い個人攻撃を含む

生成されたリスク中度の会話ログの一部抜粋を以下に示す．

[00:00] A: 今日のまとめ資料，これで大丈夫かな？
[00:01] B: うーん，正直ちょっと微妙じゃない？
[00:02] C: てかA って，毎回クオリティ低くない？
[00:03] A: そ，そうかな…直すよ．
[00:04] B: いや普通に分かりづらいよ．読む気失せるし．
[00:05] C: ほんとセンスないよな，こういうの．

4.2 可視化ツール

システムの動作確認を目的とする可視化ツールを開発した．本ツールは以下の 5 種類を出力する．

① 会話コンテクストのリスク度

図 2 は，上記に示した会話例の進行に伴って ChatGPT が推定したコンテクストリスクの変動を可視化したものである．攻撃的発言の発生に応じてリスク値が上昇し，平穏な状態が続く場合には減衰するなど，会話全体の雰囲気や緊張度の変化を理解できる．このようなグラフにより，エスカレーションの兆候や沈静化のタイミングを含む，会話ダイナミクスの特徴を把握することが可能となる．

図 2. 会話コンテクストのリスク度

② 遅延を適用した際の会話ログ

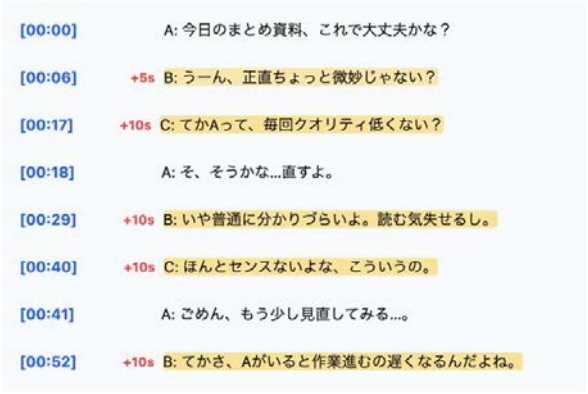

図 3. 遅延を適用した会話ログ

図 3 は，動的遅延を適用した際の会話ログの表示例である．青字の時刻 [00:06]で各発言が画面に表示される時刻を表す．赤字で示す「+5s」「+10s」は，当該発言に対してシステムが付与した遅延時間であり，リスク判定に基づきメッセージが送信から表示まで何秒間保留されたかを示す．背景が黄色でハイライトされた発言は，LLM および高リスク語彙リストにより，攻撃的または潜在的に有害と判定されたメッセージである．

③ 遅延あり/なしの会話タイムライン比較

図 4 は，動的遅延を適用した場合と適用しなかった場合の，会話が終了するまでに要した総時間を表示したものである．点線は遅延を挿入しなかった場合で，各発話が 1 秒ごとに発生したと想定している．実線は遅延を挿入した場合を表す．これにより，遅延介入が会話全体の時間構造をどの程度伸長させたかを定量的に確認できる．

図 4. 遅延を適用した会話タイムライン

④ 各発言の遅延時間

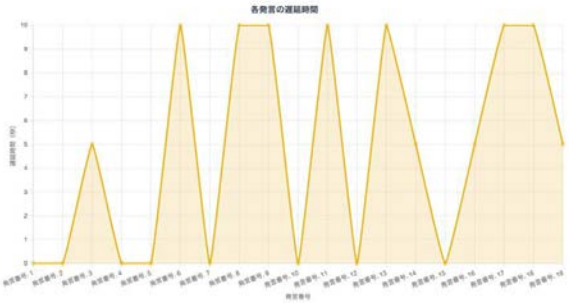

図 5. 各発言に設定された遅延時間

図 5 は，各発言がシステムによってどの程度の遅延を引き起こしたのかを示すものである．発言ごとの遅延時間を視覚的に比較することで，どの発言が高い攻撃性として評価され，より長い遅延を引き起こしたかを即座に判別できる．これにより，会話内における攻撃的発話の局所的分布や突出した発言を分析することが可能となる．

⑤ ユーザごとの累積遅延時間

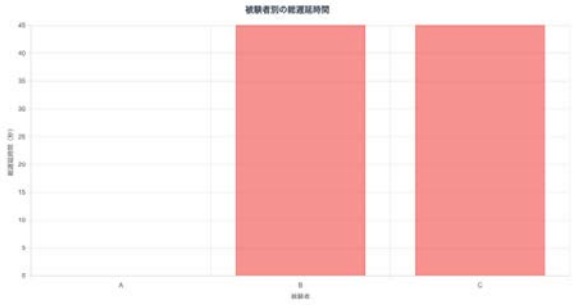

図 6. 被験者別の総遅延時間

図 6 は，個々のユーザが会話全体を通じて引き起こした遅延時間を累積したものである．ユーザごとの累積値を比較することで，攻撃的発言を最も多く行ったユーザ，つまり主導的加害者を発見したり，逆に遅延を一切発生させていないユーザ，つまり潜在的被害者・非加害者の存在を把握したりすることが可能となる．累積遅延は各ユーザの発話傾向を特徴づける重要な指標となる．

5. 考察

5.1 遅延挿入による抑止効果

チャットツールにおける動的遅延が，会話の時間構造を変化させることによって攻撃的相互作用の抑制に寄与し得ると考えている．中度の攻撃性を含むシナリオでは，遅延を適用しない条件では 20 秒前後で終了していた対話が，遅延を適用した条件では 7 倍にまで伸長した．遅延が導入されることにより，加害者は次の発言を行うまでの待機時間を経験し，その間に自らの発言内容や相手への影響を再考する機会を得る．このような遅延を設ける介入は，従来の心理学研究で議論されてきた認知的停止やクールダウンの効果と整合する[20–22]．すなわち，遅延そのものが自制心を直接回復させるわけではないものの，攻撃的行動の直前に行動の結果を想起させる時間的余白を挿入することで，衝動的反応が緩和される可能性がある．本システムでは，攻撃性が高まるほど遅延時間が長くなるよう設計されているため，エスカレーションに伴って遅延も増大し，結果として高リスク状態への遷移を遅らせる効果が期待できる．

一方で，遅延は発話のテンポや会話リズムを変化させる副作用も伴う．攻撃的メッセージが頻発する場面では，連続した遅延により対話全体のペースが著しく低下し，ユーザにフラストレーションや違和感を生じさせる可能性がある．このような負の体験は，一部のユーザにとっては抑止力として機能する一方で，システムそのものへの反発や回避行動（プラットフォーム移動や別チャネルでの会話継続）を誘発するリスクも内包している．そのため，遅延時間の上限値や発動頻度の設計は，抑止効果と利用継続性のバランスを踏まえて慎重に調整する必要がある．

動的表示遅延は「内容を検閲して削除する」直接的な介入ではなく，「時間軸を再編成する」ことによって攻撃的相互作用の流れを変える介入として位置づけられる．遅延が導入する時間的摩擦は，攻撃的発言の勢いを弱めるための冷却要因として機能し得るが，同時に会話の自然さや快適さを損なう可能性も持つ．本研究の知見は，ネットいじめ抑止システムの設計において，遅延を単なる技術的なペナルティとしてではなく，ユーザの思考プロセスや感情の変化を支える行動変容デザインの一要素として位置づける必要性を示している．

5.2 いじめの深刻度と攻撃対象に応じた遅延介入設計

遅延介入は，攻撃的発言が検出されたタイミングで会話の時間構造を変化させるという点で一貫しているが，その望ましい設計は，いじめの内容の深刻度や攻撃対象との関係性によって大きく左右される．

まず，軽い「からかい」や冗談の延長線上に位置づけられる発言の場合，参加者自身もそれを深刻ないじめとして認識していないことが多い．このような状況に対して強い遅延介入を一律に適用すると，むしろ些細なやり取りに対する過剰反応として受け止められ，システムへの不信感や回避行動を生むおそれがある．軽度の攻撃に対しては，短時間の遅延やソフトな注意喚起といった弱めの介入にとどめ，場の雰囲気を大きく損なわずに「言い過ぎ」を抑制するバランスが求められる．

一方で，継続的な人格攻撃や脅迫的表現，特定の属性を標的とした中傷など，明らかに深刻度の高いいじめに対しては，軽度のケースと同一の閾値や遅延時間を適用するだけでは不十分である．

また，たとえば被害者が会話グループの内部に存在する場合，攻撃的メッセージが可視化されるたびに心理的負荷が蓄積し，安全感が損なわれていく．そのため，度を超えた深刻ないじめが検出された場合には，会話の強制的シャットダウンも考慮する必要がある．ここでは会話の流れを維持することよりも，被害者の保護とエスカレーションの即時停止が優先されるべきであり，遅延パラメータの設計はこうした優先順位を反映して調整される必要がある．

一方，攻撃対象が会話グループ外の人物である場合には，介入の意味づけは異なってくる．外部の第三者に対する悪口や中傷は，その場に被害者が存在しないため，当事者の心理的負荷という観点では軽視されがちである．しかし，そのような発言が常態化すると，グループ内部で攻撃的言語が許容される規範が形成され，将来的に内部メンバーが標的になるリスクを高める．こうした外部への攻撃は，コミュニティ全体の言語文化や倫理規範に影響を及ぼすため，遅延介入は被害者保護だけでなく，参加者に対してどのような発言が許容され，どのような発言が境界線を越えるのかを可視化する役割も担うことになる．

以上に示すように，遅延介入の強度や発動条件は，いじめの深刻度と攻撃対象が誰であるかという文脈的要因と密接に結びついている．

5.3 遅延の設定手法

前章では，リスクのある発言があった際に，その発言により会話全体に遅延が引き起こされ，会話全体の進行が遅くなると想定したが，別の方法として，リスクのある発言だけを遅延して表示し，そうでない発言は通常通り表示するという実装も想定される．このような遅延方式を以下では個人遅延と呼び，これと対比して前章で述べた方式をグローバル遅延と呼ぶものとする．前述の会話例に対し，個人遅延を適用した際の会話ログを図 7 に示す．

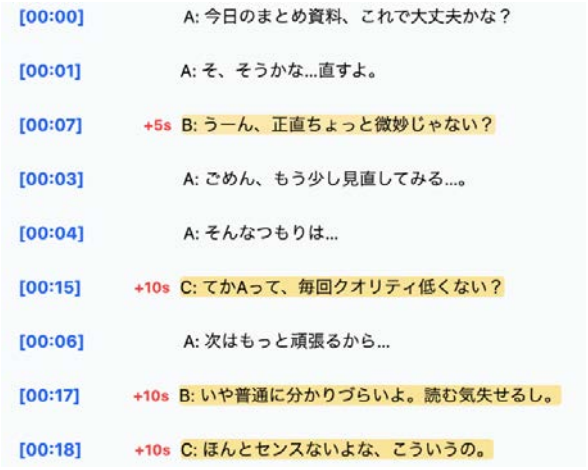

図 7. 個人遅延を適用した会話ログ

図に示すように，この条件下では，加害者の発言が遅延されている間にも他のユーザは通常どおり発言を行うため，送信時系列では加害者の発言が先であったにもかかわらず，受信側の表示順序では後続の非攻撃的なメッセージが先に表示されることになる．

実際にはこのような会話ログは長時間に渡っては成立しえず，一時的に発生したとすれば攻撃的発言の発話者は自分の発言が明らかに遅れて表示されていることに気づき，自主的に発言を控えることが期待できる．また，攻撃的発言に対する他者による共感的発言も発生しなくなると想定できる．このような場合に実際にどのような対話になるかは，実際に人間の参加者による対話実験が行われないとわからないため，今後の検討課題となる．

いずれにせよ遅延挿入はチャットシステム上における社会的・認知的ダイナミクスを再構成するデザイン要素であり，破壊的，建設的，保護的のいずれにも作用し得る．さらに，本稿では考慮しなかったが会話が遅延していることを示すユーザインタフェースデザインとしても，カウントダウン付きのグレー表示バブル，メッセージ凍結アニメーション，全体停止の視覚的表現などさまざまな手法が考えられる．

以上のように本研究の知見は，遅延を用いたネットいじめ抑止システムのデザインにおいて，技術的要因（検出精度・遅延制御）と体験的要因（認知負荷・対話の自然さ・社会的関係の変容）を統合的に検討する必要性を示している．

6. 今後の課題

本研究では，動的遅延により攻撃的相互作用の抑制可能性を示したが，実社会での運用に向けていくつかの課題が残されている．まず初めに，倫理的に管理された環境でユーザ実験を行い，動的遅延が加害者・被害者・傍観者に与える心理的影響や行動変容を検証する必要がある．また，遅延メッセージの視覚的表現や通知方法はユーザの受容性に大きく影響するため，負担の少ない提示方法を検討するとともに，言い換え提案や再考を促すメッセージなど他手法との統合による複合的な介入設計が求められる．さらに，遅延介入が短期的効果にとどまらず長期的な行動変容に結びつくかどうかを明らかにするため，縦断的な評価によりエスカレーション抑制効果や遅延への慣れを検討する必要がある．最後に，教育現場やオンラインコミュニティにおけるモデレーションツールとしての実装に向けて，遅延による会話阻害とユーザの安全確保のバランスをどのように取るかが課題となる．プライバシー保護や遅延理由の提示を含む説明責任を考慮した包括的なシステム設計を通じて，実運用に耐えうる枠組みを構築することが今後の目標である．

謝辞 本研究は東洋大学重点研究推進プログラム 2025T2 の助成を受けたものです．

参考文献

[1] Patton, D. U., Hong, J. S., Ranney, M., Patel, S., Kelley, C., Eschmann, R., & Washington, T. (2014). Social media as a vector for youth violence: A review of the literature. Computers in human behavior, 35, 548-553. https://doi.org/10.1016/j.chb.2014.02.043

[2] Englander, E., Donnerstein, E., Kowalski, R., Lin, C. A., & Parti, K. (2017). Defining cyber-bullying. Pediatrics, 140(Suppl.2), S148–S151. https://doi.org/10.1542/peds.2016-1758U

[3] Kumar, V. L., & Goldstein, M. A. (2020). Cyberbullying and adolescents. Current pediatrics reports, 8(3), 86–92. https://doi.org/10.1007/s40124-020-00217-6

[4] Langos, C. (2012). Cyberbullying: The challenge to define. Cyberpsychology, Behavior, and Social Networking, 15(6), 285–289. https://doi.org/10.1089/cyber.2011.0588

[5] Zhang, W., Huang, S., Lam, L., Evans, R., & Zhu, C. (2022). Cyberbullying definitions and measurements in children and adolescents: Summarizing 20 years of global efforts. Frontiers in public health, 10, 1000504. https://doi.org/10.3389/fpubh.2022.1000504

[6] Patchin, J. W., & Hinduja, S. (2015). Measuring cyberbullying: Implications for research. Aggression and Violent Behavior, 23, 69–74. https://doi.org/10.1016/j.avb.2015.05.013

[7] Chen, Y., Zhou, Y., Zhu, S., & Xu, H. (2012). Detecting offensive language in social media to protect adolescent online safety. In 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing (pp. 71–80). IEEE. https://doi.org/10.1109/SocialCom-PASSAT.2012.55

[8] Gao, Z., Jing, S., & Zhang, L. (2025). A study on the application of large language models based on LoRA fine tuning and difficult sample adaptation for online violence recognition. Symmetry, 17(8), 1310. https://doi.org/10.3390/sym17081310

[9] Aliyeva, A., Kenjayeva, B., Kizdarbekova, M., Kaldarova, B., Mamikov, S., Omarov, B., Omarov, N., Toktarova, A., & Adaly, E. (2025). Toward safer digital communication: A deep hybrid model for detecting abusive language on social networks. Engineering, Technology & Applied Science Research, 15 (5), 27126–27132. https://doi.org/10.48084/etasr.12721

[10] Katsaros, M., Yang, K., & Fratamico, L. (2022). Reconsidering tweets: Intervening during tweet creation decreases offensive content. In Proceedings of the International AAAI Conference on

Web and Social Media (Vol. 16, pp. 477–487). https://doi.org/10.1609/icwsm.v16i1.19308

[11] Kutok, E. R., Dunsiger, S., Patena, J. V., Nugent, N. R., Riese, A., Rosen, R. K., & Ranney, M. L. (2021). A cyberbullying media-based prevention intervention for adolescents on Instagram: Pilot randomized controlled trial. JMIR Mental Health, 8(9), e26029. https://doi.org/10.2196/26029

[12] Mahdi, M. A., Fati, S. M., Ragab, M. G., Hazber, M. A., Ahamad, S., Saad, S. A., & Al Shalabi, M. (2025). A novel hybrid attention-based RoBERTa-BiLSTM model for cyberbullying detection. Mathematical and Computational Applications, 30(4), 91. https://doi.org/10.3390/mca30040091

[13] Modha, S., Majumder, P., Mandl, T., & Mandalia, C. (2020). Detecting and visualizing hate speech in social media: A cyber watchdog for surveillance. Expert Systems with Applications, 161, 113725. https://doi.org/10.1016/j.eswa.2020.113725

[14] Perasso, G. (2020). Cyberbullying detection through machine learning: Can technology help to prevent internet bullying? International Journal of Management and Humanities, 4(11), 57–69. https://doi.org/10.35940/ijmh.K1056.0741120

[15] Royen, K. V., Poels, K., Vandebosch, H., & Zaman, B. (2022). Think twice to be nice? A user experience study on a reflective interface to reduce cyber harassment on social networking sites. International Journal of Bullying Prevention, 4(1), 23–34. https://doi.org/10.1007/s42380-021-00101-x

[16] Salawu, S., He, Y., & Lumsden, J. (2020). Bullstop: a mobile app for cyberbullying prevention. In Proceedings of the 28th International Conference on Computational Linguistics: System Demonstrations (pp. 70–74). https://doi.org/10.18653/v1/2020.coling-demos.13

[17] Amran, A., Zaaba, Z. F., & Mahinderjit Singh, M. K. (2018). Habituation effects in computer security warning. Information Security Journal: A Global Perspective, 27(4), 192–204. https://doi.org/10.1080/19393555.2018.1505008

[18] Anderson, B. B., Vance, A., Jenkins, J. L., Kirwan, C. B., & Bjornn, D. (2016). It all blurs together: How the effects of habituation generalize across system notifications and security warnings. In Information Systems and Neuroscience: Gmunden Retreat on NeuroIS 2016 (pp. 43–49). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-41402-7_6

[19] Warner, M., Strohmayer, A., Higgs, M., Rafiq, H., Yang, L., & Coventry, L. (2024). Key to kindness: Reducing toxicity in online discourse through proactive content moderation in a mobile keyboard, arXiv:2401.10627. https://doi.org/10.48550/arXiv.2401.10627

[20] Osgood, J. M., & Muraven, M. (2016). Does counting to ten increase or decrease aggression? The role of state self‐control (ego‐depletion) and consequences. Journal of Applied Social Psychology, 46(2), 105‐113. https://doi.org/10.1111/jasp.12334

[21] Nesmith, A. (2023). Text-based crisis counseling: an examination of timing, pace, asychronicity and disinhibition. Youth, 3(1), 233–245. https://doi.org/10.3390/youth3010016

[22] McGurry, A. G., May, R. C., & Donaldson, D. I. (2024). Both partners' negative emotion drives aggression during couples' conflict. Communications Psychology, 2, 122. https://doi.org/10.1038/s44271-024-00122-4