Severity-aware Decision Support System for Women's Online and Offline Safety Using Large Language Models

Christlin Jefrina R [1], Mythily M [1,*], Vaisnave I M [1], and Iwin Thanakumar Joseph Swamidason [2]

[1] Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India
[2] Computer Science and Engineering, School of Advanced Computing, Alliance University, Bangalore, India
Email: christlinjefrina2k4@gmail.com (C.J.R.); mythily.m@gmail.com (M.M.); vaisnave01@gmail.com (V.I.M.); iwineee2006@gmail.com (I.T.J.S.)
*Corresponding author

***Abstract*—Safety issues faced by women today occur in both online and offline settings, as they develop gradually without clear warning signs. Most existing safety technologies focus on emergency responses and incident reporting, but they do not provide support for ongoing situations that require context-specific guidance. This research presents a decision-support framework that identifies threats using structured threat modeling and operates with a constrained large language model. The multi-label threat taxonomy system recognizes all simultaneous online and offline security breaches. The heuristic severity index incorporates experiential indicators, such as fear, loss of control, escalation, physical boundary violations, and misuse of authority. Experiential data were gathered through a trauma-informed survey with 240 responses, of which 204 remained after consent and quality checks. Compared to standard instruction-following large language models, the proposed framework consistently provides guidance more aligned with perceived threat severity (89.2% vs. 54.3%), better acknowledges personal boundaries (93.6% vs. 61.5%), maintains a non-judgmental tone (96.1% vs. 58.1%), and gives proportionate escalation advice based on context (90.4% vs. 49.7%). Additionally, violations of predefined safety constraints are significantly reduced (3.9% compared to 27.8%). Feedback from user-centered evaluations indicates that responses are clearer, emotionally appropriate, and more trustworthy. These findings suggest that large language models can be effectively used as supportive tools in women's safety applications rather than as autonomous or unconstrained systems.**

***Keywords*—large language model, decision-support framework, context-specific guidance, human-computer interaction**

I. Introduction

Digital spaces increasingly overlap with physical environments, creating a growing need for technological solutions that protect women from violence. Digital platforms, which include social media and messaging applications, serve as the main routes through which people experience harassment, coercion, intimidation, and boundary violations because these online interactions can lead to actual physical violence. The need to maintain a constant online presence introduces additional challenges and perpetual contact with others creates other obstacles that users must overcome. The incidents that occur typically do not show clear evidence of criminal conduct or present immediate threats to life, yet they produce long-lasting psychological problems that require standard treatment methods instead of emergency-only solutions.

The existing technologies that aim to protect women's safety focus on reactive systems that include panic buttons, emergency alerts, location sharing, and post-incident reporting. The emergency systems demonstrate effectiveness during critical situations, yet they do not provide solutions for early threat assessment or for users who need to establish control over moderate levels of danger while they receive assistance through established systems. The lack of decision-making frameworks that take into account particular contexts leads women to use informal methods for coping with their situations.

Recent improvements in Large Language Models (LLMs) have created conversational systems that can analyze narratives and produce context-specific responses. Organizations can use these models to deliver personalized safety support that continues throughout their operations. The use of general-purpose LLMs in safety-critical environments remains restricted because these systems produce unfiltered content, which includes inappropriate responses that create victim-blaming situations and provide guidance that does not match the user's emotional state or risk level. The user loses trust because of this situation, which leads to additional psychological damage.

The AI safety frameworks now face a critical problem because they cannot properly assess the seriousness of safety incidents. The current methods that exist to handle safety situations treat them as two distinct categories,

Manuscript received February 4, 2026; revised March 25, 2026; accepted May 8, 2026; published August 19, 2026.

which represent safe and unsafe conditions, without recognizing how different safety levels, emotional effects, and escalation methods create unique hazards. Women experience safety problems through a continuum that authorities evaluate from direct behavioral observations, their fear, loss of control, and authority avoidance patterns. Systems that do not capture this continuum will either miss essential early indicators or they will react excessively to situations which require minimal response.

Security applications today use two primary methods, which include rule-based systems and end-to-end systems. The operation of rule-based systems provides transparent visibility to users, while they fail to offer needed adaptability for multiple situational contexts. The end-to-end models create black-box systems because they keep their decision-making processes hidden, which makes their decisions impossible to track, whereas these systems do not meet the security requirements needed for sensitive environments. The need exists for systems that deliver transparent decision-making functions, together with systems that handle natural language generation through adaptable solutions.

The study introduces a decision-making framework that assesses risks to ensure women's security through digital and physical environments. The framework contains three parts, which include threat identification and severity assessment, and guidance creation, which uses LLMs for decision assistance instead of independent operation. A structured threat taxonomy defines typical online and offline safety violations to establish security controls. The heuristic severity index uses experiential indicators, which include fear, escalation, physical boundary violations, and authority misuse to create its design. Structured signals function as control variables that determine the guidance's tone, urgency, and content.

The framework uses real-world experiential data that researchers gathered through a trauma-informed survey that protects participant confidentiality to evaluate safety experiences, coping methods, and essential guidance. The dataset enables the framework to create responses that show real human experiences instead of abstract theoretical rules. The LLM uses parameter-efficient fine-tuning together with constrained prompting to create non-judgmental responses that display contextual suitability while matching severity levels and ethical requirements.

The primary contributions of this work are threefold. First, the study establishes an interpretable modeling framework that assesses severity levels to connect online and offline environments used by women. Second, it develops a controlled LLM-based decision-support framework that provides contextually accurate guidance tailored to specific situations, rather than offering limitless conversational support. Third, through qualitative analysis, user-centered evaluation, and statistical assessment, the research demonstrates that severity-aware control improves response alignment, emotional appropriateness, and user trust compared to a baseline instruction-following LLM.

This work proposes a new framework that provides a different research methodology from existing predictive risk assessment methods. The system enables decision makers to make better choices during uncertain safety situations, which require them to respond in proportionate ways. The proposed method shows that structured control signals enable LLMs to operate as responsible elements in safety-critical decision support systems while they maintain their connection to natural language generation.

II. Literature Review

The studies about violence against women in public health and social sciences have resulted in multiple research results. The World Health Organization found that a large number of women around the world go through physical or sexual violence at some point in their lives, showing how widespread and ongoing the problem is [1]. Devries *et al.* [2] gathered global data on intimate partner violence because their work showed that this type of violence happens in many different cultures and places around the world. These studies demonstrate two important aspects that show how frequently the problem happens and its effects, which continue over time. The study shows which security methods users should use to protect themselves from threats.

Women now take steps to protect themselves online because digital communication has become more common. The Pew Research Center [3] found that women are more likely to face online harassment, like stalking, sexual abuse, and threats, especially among younger women. Henry and Powell [4] called these behaviors "technology-enabled sexual violence" because they show that online attacks worsen existing power imbalances instead of creating new ones. The study demonstrates that online threats that exist today can produce real-world consequences. The research lacks an explanation that helps users to protect themselves from immediate threats.

The study shows that legal research and feminist research together create a complex understanding of women's safety across different environments. Citron [5] introduced the idea of "cyber hate crimes" as a new type of crime, which demonstrates that online harassment produces equivalent damage to physical violence because it generates fear and breaks social ties and harms a person's reputation. The organization Plan International [6] found that online harassment acts as a barrier that prevents girls and young women from accessing digital platforms since it causes them to either self-censor or withdraw from online spaces entirely.

Kelly's [7] framework explains how women experience violence along a continuum, which starts with mild discomfort before reaching extreme physical danger through the course of their life, which shows that safety exists beyond simple binary definitions. The European Union Agency for Fundamental Rights [8] discovered that numerous women choose not to report their experiences because they lack certainty about the severity of the situation or whether it meets the standard for reportable incidents. The results demonstrate the requirement for models that establish safety measures during uncertain

circumstances, which can develop into dangerous situations.

Research has also examined the impact of online and offline abuse on mental health. Cyberbullying creates a direct link to higher anxiety levels and lower self-esteem, which results in lasting emotional damage, according to research conducted by Hinduja and Patchin [9]. UN Women [10] recognized technology-facilitated violence as a primary obstacle to achieving gender equality, while they called for technological solutions that can assist survivors. The existing initiatives concentrate their efforts mainly on developing policies and guidelines, which results in insufficient attention towards creating actual safety tools that users can use for protection.

Social science research has advanced through natural language processing, which enables computer systems to produce text that resembles human speech patterns. Brown *et al.* [11] showed that LLMs enable large language models to learn new skills through short training sessions. The current LLMs were developed from earlier systems, which included Bidirectional Encoder Representations from Transformers (BERT) [12] and the attention mechanism that Vaswani *et al.* [13] created. Radford *et al.* [14] demonstrated that unsupervised models possess the ability to handle multiple tasks without needing dedicated training for each specific task.

Foundation models enable users to execute various tasks, yet they pose dangers when used in domains that handle confidential information. Bommasani *et al.* [15] studied the advantages and disadvantages of these models, which showed problems about bias, dangerous results, and missing responsibility. Zhao *et al.* [16] demonstrated that LLMs need proper configuration and alignment processes to achieve dependable functioning for practical usage scenarios. The massive models PaLM [17] and GPT-4 [18] exhibit greater capabilities with increasing size, yet they need effective safety protocols and restricted operational methods. Users tend to take safer actions based on their subjective risk assessment because chain-of-thought prompting techniques enhance reasoning abilities [19].

The LLM assessment requires evaluation through performance metrics and measurement of its effects on society and practical results according to present assessment methods. The evaluation frameworks that Liang *et al.* [20] proposed combine ethical assessment with performance assessment. The existing concerns connect to international initiatives, which include the United Nations 2030 Agenda for Sustainable Development [21] and the WHO World Report on Violence and Health [22] that emphasize preventive methods as essential. The existing frameworks fail to present operational methods that can deliver immediate security assistance to users.

Women's safety has received multiple technological enhancements through mobile applications, wearable devices, and alert systems, which function as safety solutions. The researchers Sarkar and Das [23] discovered that early safety tools mainly functioned through their alarm systems and emergency contact functions. The researchers Gurjar and Bhardwaj [24] demonstrated that wearable devices face two main problems, which include device reliability problems and false alert notifications. The emergency solutions provide effective assistance during critical situations, but they do not handle situations with undefined conditions or changing circumstances.

Legal and policy frameworks determine the direction of women's safety initiatives. The Criminal Law (Amendment) Act of 2013 established a clear definition of sexual offenses, which India enacted as a law [25]. Kaur and Garg [26] evaluated the success of legal tools that handle domestic violence cases. The OECD [27] said that gender-based violence needs different agencies to work together in approved partnerships. Bhattacharyya [28] used spatial analysis to demonstrate that public spaces now function as dangerous areas, which people perceive as places to expect threats. The National Crime Records Bureau [29] reports rising crime rates against women, which shows the need for safety solutions that are scalable and accessible to everyone.

Multi-Criteria Decision-Making (MCDM) serves as a method for handling complex decision processes that involve multiple factors and their associated uncertainties. The methods combine various indicators to assess risk because they prove useful in assessing situations that involve rare but significant events. Recent studies have looked at combining criteria to improve decision-making in uncertain environments [30, 31]. The proposed heuristic severity index follows this idea by using behavioral and experiential indicators to create a structured measure of severity.

Despite these advancements, existing research on women's safety still does not adequately address real-time decision support. New large language models now enable conversational intelligence, so there is a need to find better ways to handle safety problems. Right now, safety solutions for women mainly include emergency response models and ways to report incidents after they happen. But general language models don't help in checking how dangerous a situation is, how someone feels emotionally, or how the danger might escalate over time.

AI models mostly use two main ways to make decisions, but many of these models use hidden processes. This renders them "black-box" frameworks, which are hard to understand and limit how safely they can be used in important areas. The data that is used for training these models doesn't include important feelings like fear or losing control, or personal preferences that help make real-time decisions. To improve safety for women, a structured framework integrating severity-aware control with constrained language generation is required that understands danger levels. This model would combine threat analysis with careful language generation to protect women both online and in real life.

Existing AI-based solutions for women's safety, such as wearable technology, alerting systems, and mobile apps, have mostly been created as reactive solutions intended to be used in emergencies [23, 24]. In these situations, trigger mechanisms like panic buttons and GPS-based notifications offer an effective emergency response, but

they are unable to identify any narrative of the scenario, assess varying degrees of danger intensity, or offer suitable guidance in an ambiguous circumstance.

Nevertheless, these systems can already identify such circumstances because of recent advancements in AI and language modeling. However, they continue to produce unpredictable and potentially dangerous data without enough management, and lack ethics and transparency [15, 16]. Additionally, they do not incorporate organized decision-making from experience indications; instead, they rely on black-box reasoning.

The novelty introduced by this paper is the use of a severity-aware approach to support decisions that will help overcome these limitations.

III. Materials and Methods

This section presents a modular decision-support framework that works with awareness of how serious a situation is, to give safety advice that fits the specific situation of women who face harassment both online and offline. The method used is a type of control framework that follows clear steps. The framework establishes three main objectives, which include creating an understandable system for users, achieving consistent results through testing, and establishing safe operational protocols for critical security environments. The system uses separate modules for threat detection and impact assessment, together with system guidance and language enhancement functions, which differ from systems that operate through a single processing path.

A. System Overview and Architecture

The proposed method functions as a decision-support tool that uses severity assessment to deliver safety recommendations. The system operates through a combination of deterministic control logic and restrictions that limit Large Language Model (LLM) response generation. The framework provides safety recommendations that address the distinct safety requirements of women who experience online and offline and mixed environment harassment. The system maintains predictable safety conditions through its design, which allows for straightforward understanding and auditing.

The system uses a modular layered framework which enables its two main functions of decision-making and language generation to operate independently, as illustrated in Fig. 1. It uses structured analysis parts to process user input through threat detection and severity evaluation. These parts create control signals that guide the framework's next steps. The LLM functions as a controlled natural language module and can only be used after the framework has determined both the severity level and the required guidance.

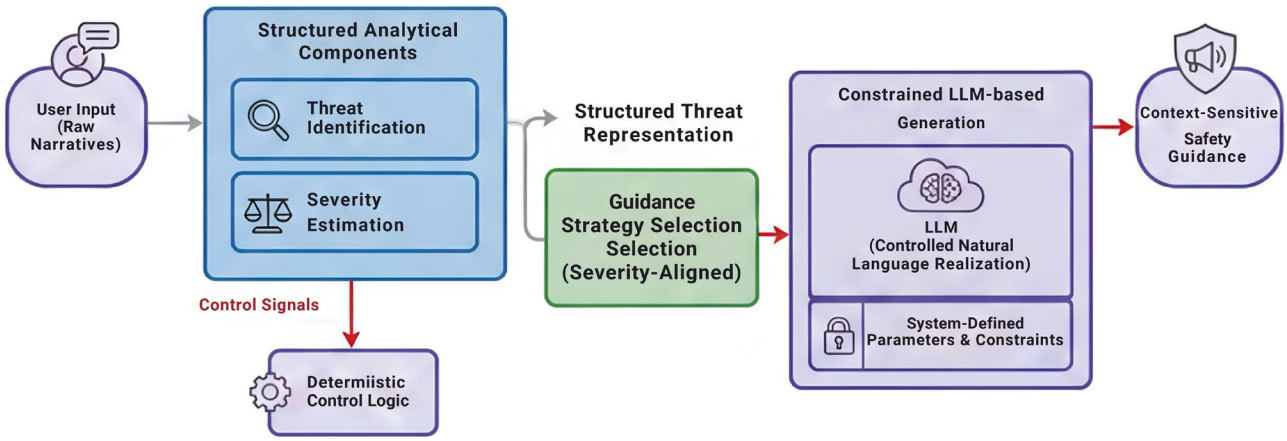

Fig. 1. System architecture of severity-aware decision support.

The process begins with raw user narratives, which are first converted into structured threat descriptions. The data is utilized to determine situation severity, which helps in selecting suitable response strategies. The LLM produces its response during the last phase, but the response must follow the established framework restrictions. The design protects against three main dangers that chat-based systems face because it prevents LLMs from operating beyond their limits, stops users from wrongly interpreting their intentions, and eliminates the chance of producing different or opposing answers. The framework keeps its safety standards intact because it creates boundaries between its decision-making systems and its language generation system. The system enables users to assess threat levels consistently throughout their interaction, which supports them in making educated choices.

The system uses modular components, enabling independent testing, and maintenance of each component while providing support for future system extensions that will not affect its fundamental safety features. The framework operates as a decision-support system because it provides a well-organized system that enables users to make decisions while presenting all the information they need to understand.

B. Dataset Description and Data Collection

The dataset used to develop and align the proposed framework was collected through a structured survey designed to capture women's safety experiences in both online and offline environments. The survey was distributed through Google Forms, while the data was maintained in ethical standards by ensuring participants

could choose to participate or not, and their answers would remain private, and they would receive full information about the research before the start of data collection.

A total of 240 responses were received. After removing incomplete responses, those that lacked proper consent, or had conflicting information, 204 valid responses were used to align the framework.

Dataset Structure:

Each response is composed of structured indicators and narrative contexts that can be used for both rule-based modeling and qualitative grounding

- Demographic context (age group, occupation, environment).
- Incident context (online, offline, or online-to-offline).
- Online threat indicators (O1–O7).
- Offline threat indicators (F1–F6).
- Experiential severity indicators (fear, escalation, avoidance, loss of control).
- Short narrative description.
- User actions and guidance preferences.

This dataset will not be used for prediction but to engineer severity-aware decision-making protocols and response behaviors. The framework was built based on the analysis of data obtained from 204 surveys conducted with young women in a particular socio-cultural environment. Even though this data reflects a diversity of experience, we recognize that the understanding of indicators of fear, abuse of authority, and boundary violations varies across other socio-cultural, legal, and cultural settings.

This means that our framework cannot be generalized to any setting without modification. Instead, it functions as a decision-support tool that is contingent upon the patterns revealed by the database. Though there exist universal human responses, such as discomfort, threat perception, and loss of control, their meaning and level of severity may vary from region to region.

To resolve this problem, we plan to expand the database to include data samples collected from a wider audience with diverse demographic characteristics. Besides, we could modify the decision-making structure of the framework to enable customizable severity assessments.

C. Problem Formulation and Framework Inputs

The proposed framework is formulated as a controlled decision-support mapping problem, in which unstructured user input is transformed into structured safety guidance through explicit intermediate representations. Let each user interaction be represented as a composite input vector:

$$X = \{N, O, F, E\} \tag{1}$$

In Eq. (1), the free-flowing commentary provided by the user regarding an incident is denoted by the letter *N*. This narrative is considered contextual grounding rather than a decision variable. The set $O = \{o_1, o_2, \dots, o_m\}$ represents binary online threat indicators related to predefined digital safety violations, while $F = \{f_1, f_2, \dots, f_n\}$ represents binary offline threat indicators that capture physical-world behaviors. The set $E = \{e_1, e_2, \dots, e_k\}$ encodes experiential severity signals reflecting the user's perceived emotional and behavioral impact, such as fear, escalation, avoidance, or loss of control.

The system supports multiple threat detection through its online and offline threat indicators, which were created to handle simultaneous threats during a single event. The framework requires two different indicator types because their combination would make it difficult to evaluate actual behavior and the perceived level of harm. The process of separation functions as an essential component for constructing safety models which demonstrate how various experiences that lack physical danger affect mental health outcomes. Given the structured input X, the framework calculates two intermediate control variables: a dominant threat category T and an inferred severity level S. These variables are deterministically derived using rule-based mappings described in later sections and function as non-generative control signals. The final framework output is defined as:

$$Y = G(N, T, S) \tag{2}$$

In Eq. (2), the function $G(\cdot)$ is a framework that gives limited guidance. It creates safety rules by looking at the story context *N*, inferred threat category *T*, and threat severity level *S*. Importantly, this function doesn't try to guess if a threat will happen or what the result will be. The established rules which T and S determine already provide guidance that maintains fairness and ethical standards without making judgments about people.

The framework development process requires three distinct tasks, which include understanding context, evaluating risks, and forming language separately. The system operates according to predictable patterns, which prevent the language model from making unexpected assessments about the urgency of situations.

D. Threat Taxonomy and Signal Encoding

Threat behaviors are encoded using an explicit multi-label threat taxonomy that differentiates between online and offline safety violations while allowing their simultaneous occurrence within a single interaction. Each threat indicator is represented as a binary variable:

$$O_i, f_j \in 0,1 \tag{3}$$

In Eq. (3), o_i denotes the presence of the i^{th} online threat behavior and f_j denotes the presence of the j^{th} offline threat behavior. Online indicators capture digital violations such as persistent messaging, emotional manipulation, sexual pressure, blackmail, privacy intrusion, and explicit threats, while offline indicators represent physical-world behaviors including unwanted attention, stalking, non-consensual contact, coercive pressure to meet, verbal harassment, and misuse of authority.

The taxonomy incorporates multiple labels because safety incidents in real life result from various concurrent unsafe actions. An interaction presents different online and offline threat indicators that must be maintained together as they represent complete threat scenarios. The system

control threat assessment begins with physical boundary violations, which act as the primary threat category, before moving through coercive threats, authority misuse, and persistent harassment. The framework uses priority rules to control system operation by determining which paths to take and which guidance to use while maintaining its original multi-label functionality. The system implements multiple threat encoding methods through its dual representation system, which connects to one control category and maintains accurate analytics and predictable system performance while providing ongoing severity-based guidance for safety-critical operations that need both interpretability and auditability.

E. Severity Estimation Engine

The framework uses a heuristic severity index to assess severity through experiential and behavioral signs, while it needs to operate without using legal definitions. The framework functions as a decision-support tool, as its design must align response intensity with urgent situations perceived by individuals who believe they have experienced harm. The formula for calculating the heuristic severity index in each interaction *i* is as follows:

$$HSI_i = f_i + l_i + e_i + 2p_i + a_i + t_i \quad (4)$$

In Eq. (4), f_i denotes self-reported fear, l_i denotes perceived loss of control, e_i represents escalation, p_i indicates physical boundary violation, a_i captures authority misuse, and t_i represents explicit threats or blackmail. All the indicators have been converted into binary values. Physical boundary crossing is weighted more heavily than other safety violations because it creates immediate safety threats that cannot be undone according to trauma-informed safety guidelines.

The sensitivity analysis conducted for the severity indicators suggests that an increased weight for physical violations systematically results in cases being categorized as higher severity, which makes it a safer approach. Although this simple weighting framework has improved the safety consciousness approach, further efforts are planned to find optimal weights based on the data analysis.

Models based on soft computing technologies, such as fuzzy logic, have traditionally modeled subjective notions like fear and loss of control with increased resolution. In contrast, the proposed framework employs binary representation to guarantee that the resulting solutions can be interpreted, reproduced, and audited, thus fulfilling the requirements set by critical frameworks. The use of deterministic representation shows the clear and explicit translation from inputs into outputs. The system has the capability to expand its functions through fuzzy or probability-based methods, which serve as temporary solutions before the system proceeds to severity assessment.

The framework establishes four heuristic severity index scores, which include S1 (Low), S2 (Medium), S3 (High), and S4 (Critical). The framework uses these levels to determine its response through simple and understandable methods, which direct users and manage situations that become more serious. The severity engine acts as a built-in framework that turns user experiences into different steps of guidance. It focuses on what users feel is dangerous, how much control they lose, and if any boundaries are broken, following guidelines that are based on understanding and supporting users who have experienced trauma.

Moreover, the proposed framework works in an interactive dialogue fashion as opposed to a one-shot input approach. If user data is incomplete or unclear, the framework actively seeks more context using clarification questions. This allows for minimizing uncertainty in severity assessment by making sure that all critical behavioral and experiential features have been fully identified before determining the severity value. Additionally, there is a temporal severity averaging scheme utilizing the average of the last few severity values to smooth out any short-term severity variations. Thus, uncertainty in severity determination does not arise from probability distributions; instead, it is tackled using context elicitation.

F. Guidance Strategy Mapping and Response Control Logic

The system provides guidance based on direct assessment of threat levels because it does not have autonomous operational capabilities. The framework has a component that assesses threat levels, and this component implements specific strategies to address various threat levels and their corresponding severity. The framework uses ethical guidelines to determine the appropriate assistance level that should be provided to users.

The main threat type T proceeds to lead the threat identification process which uses Section III-D multi-label threat list to identify potential security threats. The framework contains multiple guidance strategies which provide different emotional support tones to users who need help with establishing protective boundaries, making emergency safety decisions, preserving evidence, learning about upcoming actions, and developing awareness skills. These strategies are only used when certain conditions are met, and the main factor in deciding which strategy to use is the level of seriousness. Formally, choosing the right guidance is a clear and specific process:

$$M(T, S) \rightarrow \{G_{k1}, G_{k2}, \dots\} \quad (5)$$

The framework works through M which controls how the guidance strategies are applied depending on the level of seriousness in Eq. (5). The framework uses three different methods to handle different levels of seriousness: for low seriousness, it offers reassurance and awareness; for moderate seriousness, it provides clear boundaries and support; and for high or critical seriousness, it steps in with immediate safety measures, documentation, and options to escalate the situation. The specific content for each threat category T adds more details, helping ensure the response is appropriate and preventing over-reaction in unclear situations.

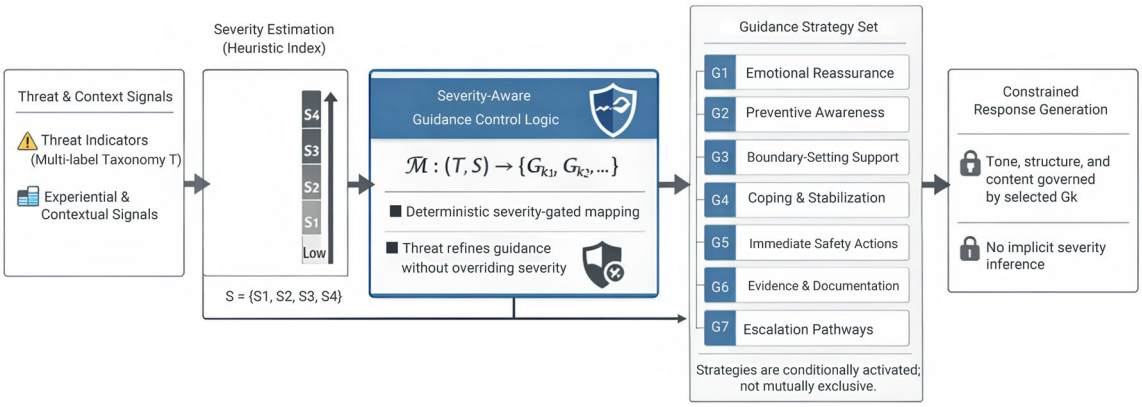

Fig. 2. Guidance strategy mapping and response control logic.

As shown in Fig. 2, the guidance control framework acts as an intermediate layer that links the assessed severity of a situation with the generation of responses. The system establishes guidance techniques that assist users in handling different situations through its combination of threat detection signals and experience-based performance assessment methods. The framework triggers higher levels of intervention only when required, and it applies clear constraints on both content and tone to ensure that the responses remain appropriate and consistent. The system allows tracking of system performance through (T, S) configuration data, which connects to every output in the system.

G. LLM Integration, Constraint Enforcement, and Alignment Strategy

The framework uses the Large Language Model (LLM) to generate language but not to make decisions. The system identifies threats, evaluates their severity, and determines appropriate actions through established rules that operate before safety-related decisions. The system introduces the LLM after completing decision-making because this process ensures its outputs will follow established safety requirements. The model operates through a fixed prompt, which describes its functioning as:

$$P = \{N, T, S, G_S, C\} \tag{6}$$

In Eq. (6), N represents the user's narrative context, T denotes the dominant threat category, S indicates the inferred severity level, and G_S corresponds to the set of guidance strategies selected by the severity–guidance mapping layer. The component C establishes multiple strict behavioral rules that stop the use of victim-blaming language and disallow any guessing about intent. The system guards all user interactions with secure rules that the model cannot bypass.

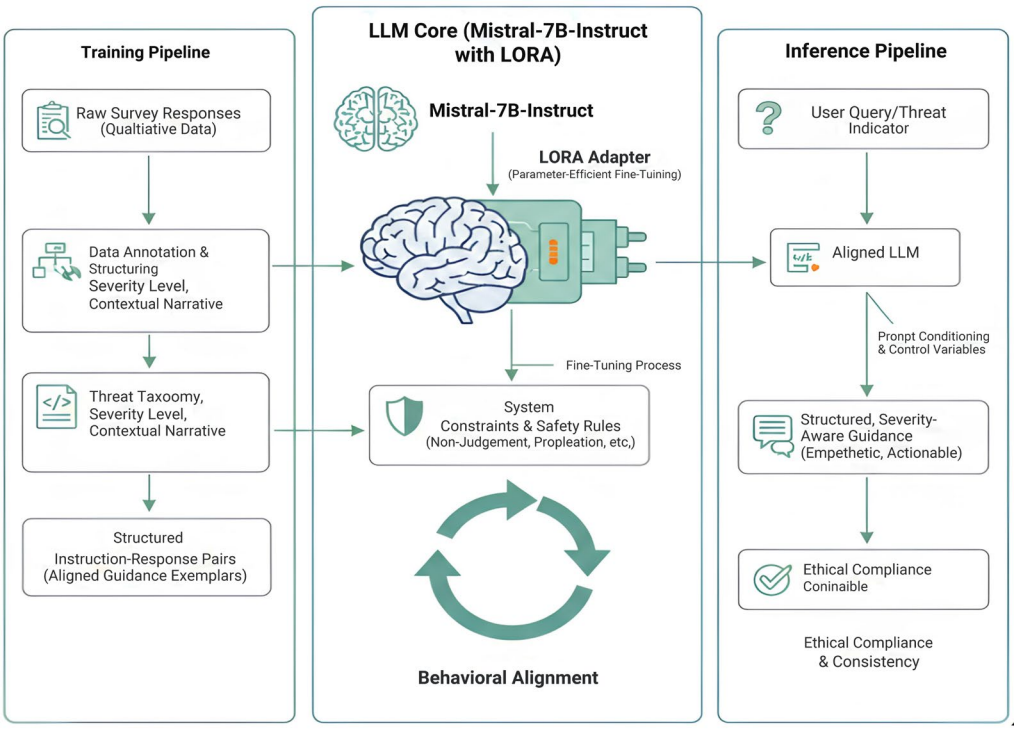

Fig. 3. LLM alignment and fine-tuning pipeline.

TABLE I. LLM ALIGNMENT CONFIGURATION

8 (pts.)	Component	Specification
1	Base language model	Mistral-7B-Instruct
2	Alignment method	Parameter-efficient fine-tuning (LoRA)
3	LoRA rank (r)	8
4	Target modules	Attention projection layers
5	Training samples	204 structured instruction–response pairs
6	Epochs	3
7	Objective	Severity-aligned response realization
8	Full model weight updates	No

The base model, which was instruction-tuned, requires parameter-efficient fine-tuning through Low-Rank Adaptation (LoRA) to achieve model alignment. The training process uses structured instruction-response pairs to show how control variables T, S *and* G_S should control the generated output. The process does not create fresh information; instead, it works to synchronize the model's actions while maintaining its current language skills.

The deployment process uses identical control signals and training constraints through prompt conditioning, which delivers consistent performance across multiple operational situations. The system approach prevents unexpected response variations while stopping the development of unmanageable system behavior. The model produces predefined output patterns that enable it to deliver understandable and consistent answers to users. The LLM operates as a controlled language generation module within a decision-support system, which establishes its operational limits through upstream logic and safety requirements. The process of aligning and deploying the model, which includes adjusting it with LoRA during training and applying prompts during use, is shown in Fig. 3.

To ensure consistent application of severity-aware guidance, parameter-efficient fine-tuning was implemented using LoRA. This alignment process requires only minimal changes, which will not result in new knowledge acquisition or performance increases for the assigned tasks. Rather, it aims to maintain response structure and tone under consistent threat and severity control signals. The language model operates within structured prompting and control variables, while all safety-related decisions depend on deterministic upstream logic. Table I shows the LLM Alignment Configuration.

H. Conversational Context Management and Deployment Workflow

The framework needs clear context management as a key safety rule to prevent three kinds of mistakes that happen during conversations in important safety areas. It uses a limited memory that saves the conversation between the user and the LLM until a set time has passed. The system design directs users to focus on their current requirements while it stops processing past emotional data and previous conversations, and it uses current user activities to test model functionality.

In Fig. 4, each message from a user is handled in a controlled loop for processing. The message is added to a buffer with a set limit. The basic prompt for the large language model results from combining the model prompt with current control variables, which identify the threat category and its severity level. The framework requires threat assessment and severity evaluation during every conversation because these metrics cannot be retrieved from earlier discussions.

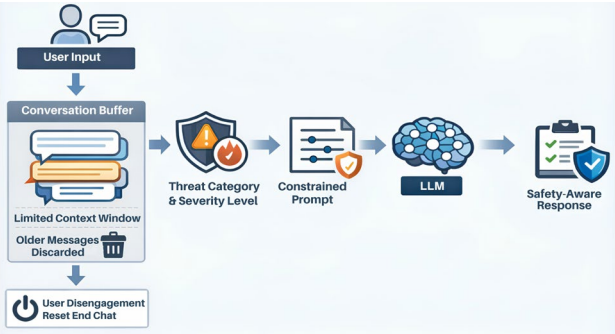

Fig. 4. Conversational context management and deployment workflow.

The framework includes a mechanism to manage conversational context by discarding older messages once the memory limit is reached. The system maintains dialogue length restrictions because it prevents users from understanding extended conversations. It starts a conversation history reset process when user disengagement occurs because this process removes all effects from past interactions on future system responses. The approach helps prevent misunderstandings while keeping the system free from unnecessary contextual information. The framework uses control mechanisms that maintain both the perceived situation severity and emotional tone throughout all interactions. It contains strict ethical guidelines that undergo regular reviews to verify ongoing compliance with those standards. The structured design of the framework enables users to make safer decisions through better situation assessment and controlled informed responses.

$$\bar{S}_t = \frac{1}{k}\sum_{i=t-k+1}^{t} S_i \tag{7}$$

To maintain coherence across multi-turn interactions while preventing error propagation, the framework introduces a short-term severity consistency mechanism. Let S_t denote the severity level computed at the interaction step t, and let the system maintain a bounded history of the last k severity values. A smoothed severity estimate is computed as shown in Eq. (7), which aggregates recent severity values using a mean-based approach. It stabilizes severity by reducing the fluctuations caused by inconsistent inputs.

$$S_t^{\text{final}} = \max(S_t, \bar{S}_t) \quad (8)$$

The framework uses a conservative update method according to Eq. (8) when the existing severity assessment shows differences with previous interaction states. The rule selects the maximum between the current severity and the smoothed estimate, which protects high-risk signals from being filtered out. As a result, the framework maintains essential safety standards while it enables user input to modify its operation.

IV. Result and Discussion

The framework operates as a severity-aware decision-support framework, which does not rely on conventional Natural Language Processing (NLP) metrics to evaluate its performance since it is not designed as a predictive or generative benchmark model. The evaluation process examines four aspects: behavioral alignment, safety compliance, severity proportionality, and user-perceived usefulness, all of which align with the objectives of safety-critical framework design. The evaluation approach employs four methods: distributional analysis, controlled baseline comparison, behavioral auditing, and user-centered statistical validation, to generate evidence of the framework's effectiveness.

In applications of a critical nature and involving users, there cannot be any absolute ground truth for guiding actions since the right thing to do changes according to the person's understanding, context, and risks involved. Thus, the presented framework is designed without any assumption about the existence of ground truth labels. The evaluation process is set as a problem of behavior alignment where the output is compared with the established criteria related to safety and ethics. It includes proportionality concerning the seriousness of the situation, recognition of boundaries, and proper escalation.

Also, user-centered evaluation is conducted in order to determine the perceived usefulness, emotional adequacy, and trust. They can be viewed as practical indicators for estimating the actual performance of the framework.

A. Experimental Setup and Evaluation Protocol

This research tested the study using a controlled dataset that had 204 valid responses from participants who gave their consent to take the trauma-informed survey (Section III-B). The decision-making process went through each response completely, which involved identifying threats using a threat classification framework and estimating the severity using heuristics. Then, it selected a guidance strategy and generated responses from the large language model with specific constraints. All the evaluations were done using the same settings, and for each user interaction, the study recalculated the severity and threat signals.

The framework was evaluated across three complementary dimensions:

- Operational coverage, assessed via threat and severity distributions,
- Behavioral correctness, assessed via comparison with a baseline instruction-following LLM, and
- User-perceived alignment and trust, assessed via structured user feedback and statistical testing. This multi-layered protocol ensures that results are not dependent on a single metric or evaluation perspective.

B. Threat and Severity Distribution Analysis

The system is evaluated by the distribution of threat categories and their corresponding severity levels across the dataset to identify the operational conditions that would allow the framework to function effectively. A rule-based hierarchy from Section III-D assigned each response to its main threat category, while Section III-E established a heuristic severity index to determine severity levels. The framework's decision space relies on these distributions as they do not accurately represent actual population distribution patterns.

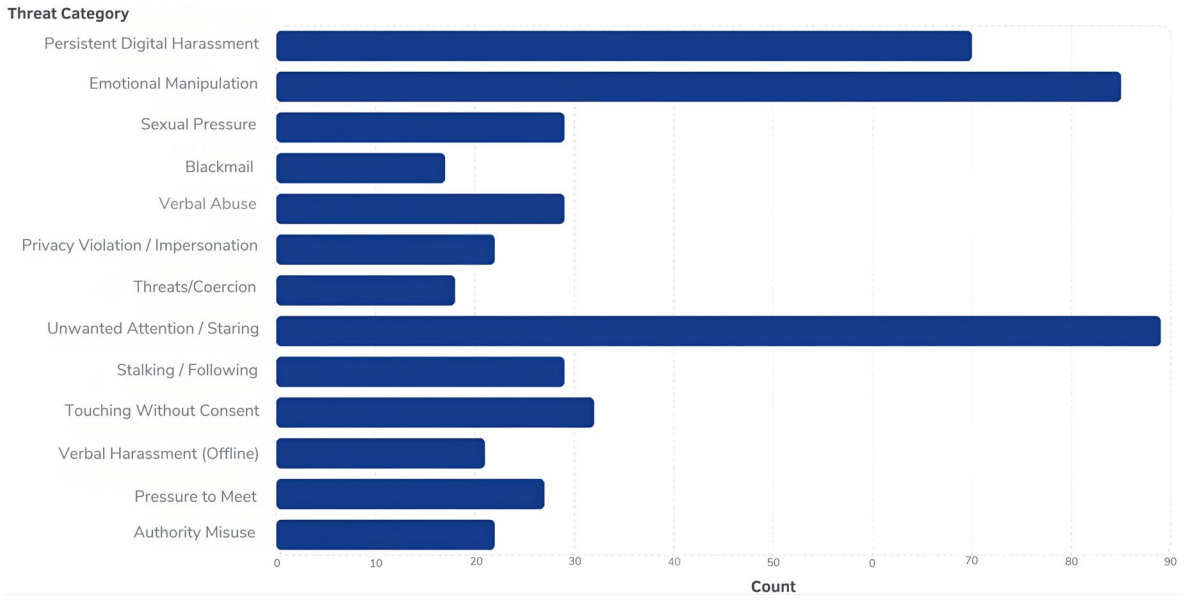

Fig. 5. Threat category counts.

As shown in Fig. 5, authority-based coercion was the most common type of threat. Public-space intimidation, threats, coercion, and physical boundary violations were the next most frequent types. Most cases involved ongoing online harassment, while incidents where online threats turned into real-world problems happened much less often. The variety of threats highlights the need for a single system that can handle both digital security and physical safety through one decision-making process.

The distribution of severity levels, illustrated in Fig. 6, shows how severe the cases are. The majority of cases demonstrate that most individuals experience moderate severity because they keep facing emotional and behavioral issues while they remain safe from physical harm. High-severity cases are less frequent but typically involve situations that may require urgent intervention or external support.

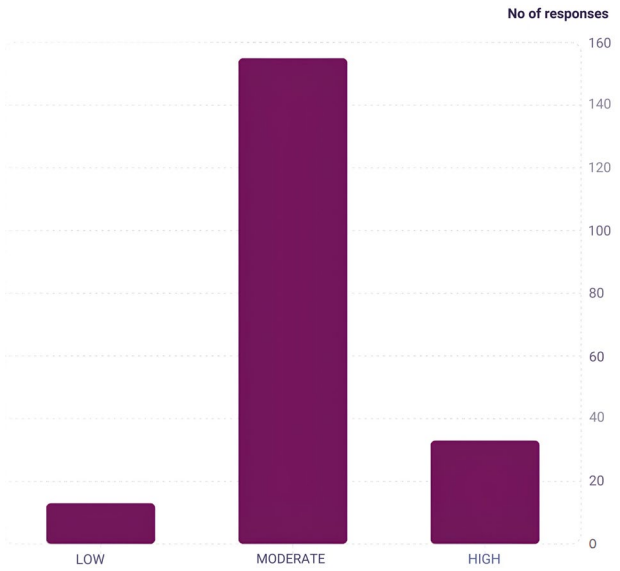

Fig. 6. Severity level distribution.

The lower end presents two distinct patterns, which show that some cases require users to obtain a better understanding, while other cases only involve slight discomfort that needs no further development. Fixed-response systems, which handle emergencies and basic reassurance needs, show insufficient capacity to handle these distribution patterns. The system needs a severity-aware approach, setting responses according to the needs of the specific situation and active threats at that time.

C. Severity–Threat Relationship Analysis

Two methods were used to check the real security level of the framework. The analysis looked at how different threat types affected three main fear-based factors that determine severity levels. The research team found out how common each threat type was by looking at the percentage of yes answers for each main threat category. They chose the main threat categories for the severity analysis because these were more stable and easier to understand than the online (O1–O7) and offline (F1–F6) indicators. This method allows for comparing different types of threats by looking at how behavior patterns with multiple threats are similar or different.

The threat categories illustrated in Fig. 7 reveal distinct patterns of severity that form unique severity profiles. The methods of authority-based coercion and threats-and-coercion retain their highest levels of fear and control loss due to the psychological impact of power imbalances and intimidation. Physical boundary violations generate significant fear as individuals experience intense distress from isolated incidents. The analysis of persistent digital harassment shows that people who experience continuous minor violations tend to withdraw from social contact. The observed patterns validate both the weighting system and the structural organization of the heuristic severity index because they show that real-world experiences of severity exist beyond the observable.

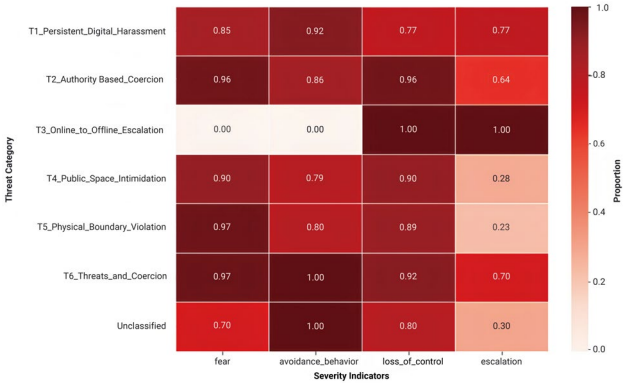

Fig. 7. Severity indicator prevalence across dominant threat categories.

A small group of cases remained unclassified because experiential severity indicators were present without enough explicit threat behaviors. The research methodologies of these cases remain intact because they have not yet reached the point of definite safety assessment.

D. Baselines and Comparison Justification

The proposed framework is assessed by testing it against a base instruction-following language model to determine how severity-based control and structured guidance methods affected the system. The baseline system operates with Mistral-7B-Instruct as its core model but lacks any of the additional features, which include threat taxonomy and severity estimation, guidance mapping, and behavioral constraints. The system creates responses based on user input because it lacks any method to establish organized control over its operations. The study maintained all aspects of the model design, including its structure and size, because it wanted to see how the framework design affected results. The baseline system demonstrates a standard operating procedure that organizations follow to use general-purpose conversational systems without implementing extra safety features.

E. Behavioral and Safety Metric Evaluation

The assessment examines both behavioral safety alignment and non-predictive results, which the

framework uses to assess non-predictive results. The team evaluated outputs through a predetermined set of criteria to verify both safety and proper usage. The criteria examined whether participants directly showed their personal limits through their answers without accusing anyone, whether their response level matched the emergency's actual danger, and whether the system prevented itself from forming guesses about what people meant or what would happen next. The researchers assessed whether the guidance material maintained its correct application for the specific severity level that had been established.

The proposed framework achieved all testing requirements because it maintained consistent performance across both medium and high-risk testing conditions. The baseline LLM demonstrated a tendency to use speculative reasoning, while it showed irregular boundary recognition, and it occasionally escalated situations too far. The model behavior study shows that model performance changes when system developers follow the procedures for specific operational requirements, but they need to use more advanced methods to establish systems that can operate safely during critical situations.

The study analyzed how different models escalated situations by recording their frequency of producing urgent guidance that was unnecessary during less serious events. Fig. 8 shows that the baseline model had a much higher rate of creating unnecessary emergencies. The proposed framework produced better results because it gave proper emergency directions for dangerous situations while it restricted urgent responses to critical emergencies.

The system-level evaluation demonstrates how severity-aware control systems function as vital components for system operation. We performed an ablation study to assess how each system element contributed to performance by disabling particular control functions and maintaining the original language model. In Fig. 9, the baseline instruction-following LLM and the new severity-aware framework show different performance results, which researchers evaluated through five safety metrics: boundary recognition, severity alignment, non-judgmental tone, escalation control, and response consistency.

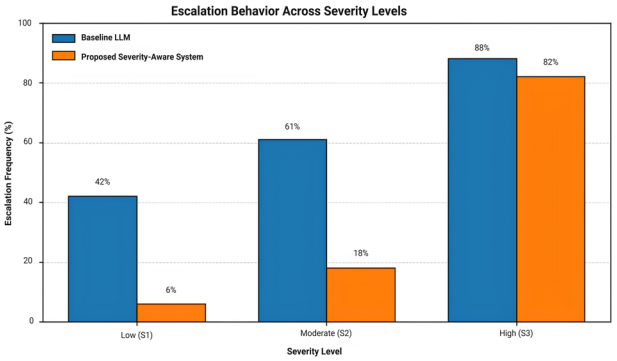

Fig. 8. Comparison of escalation frequency between the baseline LLM and the proposed framework.

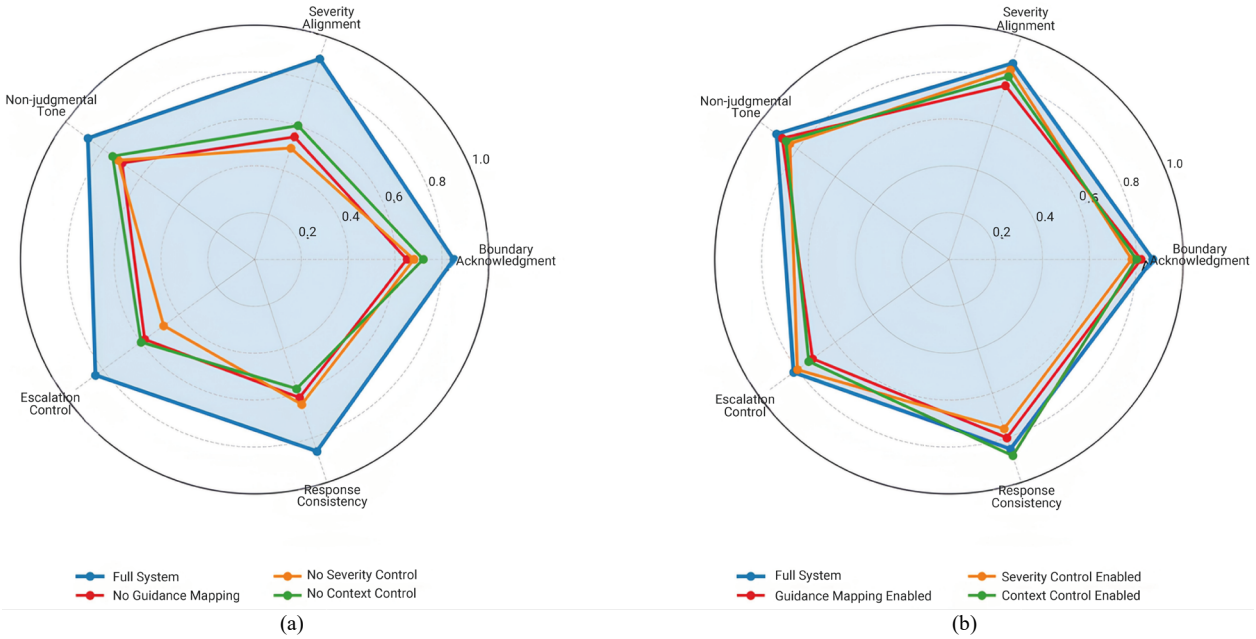

Fig. 9. Performance results: (a) Baseline LLM under control ablation and (b) Proposed framework with enabled control components.

The baseline model in Fig. 9(a) shows a major performance drop when its control elements are disabled. The system shows inconsistent performance results because all baseline components function at 0.90 severity alignment and 0.84 escalation control. The system performance shows a major decline when users disable guidance mapping and severity control, which results in escalation control dropping to 0.58 and 0.48 while severity alignment decreases to 0.55 and 0.50. The baseline system maintains its ability to engage in general conversation but shows unpredictable patterns, especially when handling unexpected situations and understanding the limits of safe interaction.

The framework demonstrated in Fig. 9b establishes superior and constant performance across all assessment criteria. The complete system achieves consistently high scores, including 0.91 for non-judgmental tone and 0.87 for boundary recognition. The system achieves total performance through control components which each provide essential support: guidance mapping increases non-judgmental tone to 0.88, while severity control prevents excessive escalation at 0.80, and context management maintains consistent response patterns at 0.88. The ablation results demonstrate that system performance decreases at a slower pace after component removal because the modular design protects system stability. Overall, the explicit severity modeling with guidance mapping and context control functions is an essential element for sustaining framework stability while ensuring fairness and ethical compliance.

F. Qualitative Case Comparison

The study used the same set of story-based scenarios to evaluate the two systems in a structured manner. In the initial setup, users responded in a more natural, conversational way, reflecting how they typically communicate with others. Although the responses appeared fluent, they often implied actions without stating them clearly, and instances of boundary crossing were not explicitly acknowledged.

In contrast, the proposed framework generated responses that were more safety-oriented and clearly identified both safety concerns and boundary violations. It also handled responsibility in a more balanced way, without placing excessive burden on the user. Table II presents the evaluation criteria used to compare the models and highlights improvements that are unlikely to be due to chance.

TABLE II. QUALITATIVE COMPARISON BETWEEN BASELINE INSTRUCTION-FOLLOWING LLM AND THE PROPOSED SEVERITY-AWARE SYSTEM

7 (pts.)	Evaluation Criterion	Baseline LLM (Mistral-7B-Instruct, No Severity Control)	Proposed Severity-Aware System
1	Response structure	Free-form narrative	Fixed, safety-oriented sections
2	Severity adaptation	Implicit or absent	Explicitly conditioned on risk level
3	Intent inference	Often speculative	Explicitly avoided
4	Boundary validation	Indirect or ambiguous	Direct and explicit
5	Emotional reassurance	Inconsistent	Consistently present
6	Escalation guidance	Uncalibrated	Severity-gated and conditional
7	Victim-blaming risk	Present in phrasing	Explicitly constrained

G. User-Centered Evaluation and Statistical Validation

For this research, a user-centered evaluation was done. It involved 30 participants who tested both systems to see how trustworthy they found them and how much value they perceived. The participants judged the responses based on five criteria: emotional appropriateness, severity alignment, clarity, non-judgmental tone, and trustworthiness. They used a five-point Likert scale to rate each of these.

The participants had consistent views on all the dimensions because their average scores were above 4.2 and the standard deviations were low, indicating consistent responses (see Table III). A Wilcoxon signed-rank test was used because the same participants tested both systems. The proposed framework showed statistically significant improvements over the baseline LLM in emotional appropriateness, severity alignment, and non-judgmental tone ($p < 0.01$). It also showed significant improvements in clarity and trust ($p < 0.05$), as shown in Table III. The proposed framework received 76.7% of direct participant preferences, showing that severity-aware alignment increases perceived value and trust, as found through both qualitative and quantitative methods.

TABLE III. STATISTICAL COMPARISON BETWEEN BASELINE AND SEVERITY-AWARE SYSTEM

5 (pts.)	Evaluation Dimension	Test	*p*-value	Significance
1	Emotional appropriateness	Wilcoxon signed-rank	<0.01	Significant
2	Severity alignment	Wilcoxon signed-rank	<0.01	Significant
3	Clarity of guidance	Wilcoxon signed-rank	<0.05	Significant
4	Non-judgmental tone	Wilcoxon signed-rank	<0.01	Significant
5	Trust in future use	Wilcoxon signed-rank	<0.05	Significant

H. System-Level Performance Metrics Comparison

The comparison between the baseline instruction-following LLM and the proposed severity-aware framework shows all framework performance metrics through model evaluation in Fig. 10. The proposed framework demonstrated all behavioral and safety metrics, with higher accuracy ranging from 65% to 95% in three tests: severity alignment (89.2% vs. 54.3%), boundary acknowledgment (93.6% vs. 61.5%), and non-judgmental response rate (96.1% vs. 58.1%). The proposed system performs better because it decreases failure-related problems, which include constraint violations and context drift, thus enhancing system stability while maintaining ethical standards. The results show that performance improves when severity is explicitly modeled together with guidance control and context management, which creates a better outcome than standard instruction-following methods.

I. Severity-Aware Decision-Support Algorithm

The proposed severity-aware decision-support framework follows the workflow outlined in Algorithm 1. The process begins by extracting threat-related and experiential severity indicators from user-provided textual descriptions of safety-related situations. These indicators

are then mapped to a discrete severity level ranging from low (S1) to critical (S4). This severity level acts as a control signal throughout the interaction. Based on this, the framework constructs a severity-aware prompt and sends it to a constrained large language model to generate guidance appropriate to the situation.

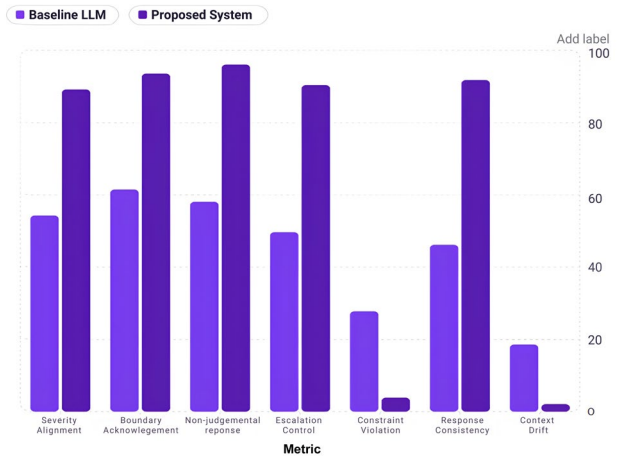

Fig. 10. Comparative system-level performance.

An example of an algorithm can be found as Algorithm 1:

Algorithm 1. Severity-Aware Safety Guidance Generation

Input: User textual description U
Output: Structured safety guidance R
Extract threat and severity indicators from U
Compute severity level $L \in \{S1, S2, S3, S4\}$
while user interaction continues, do
    Build severity-aware prompt using U and L
    Generate constrained LLM response
end while
if $L = S3$ or $L = S4$ then
    Include safety actions and escalation options
else
    Include reassurance and boundary guidance
end if
    Format response into structured sections
Return R

The algorithm distinguishes between high-severity cases (S3–S4), which require immediate safety actions and possible escalation, and lower-severity cases (S1–S2), which focus more on reassurance and boundary-setting guidance. The generated responses are organized into structured sections to maintain clarity and ensure an appropriate emotional tone. Through this design, the framework maintains ethical alignment while supporting interpretable behavior and consistent performance across different severity levels.

J. Limitations and Future Work

There are several limitations to the current research. First, the dataset used in this study was collected within a particular socio-cultural context and, therefore, may be difficult to generalize. Second, the heuristic severity model uses dichotomous variables and fixed weights to represent users' experiences and, hence, may fail to reflect nuances of perception and emotion. Third, the framework described above does not implement any sort of context-dependent self-calibration, meaning that it cannot adapt to changes in cultural or situational parameters. Fourth, the use of a bounded memory approach for contextual conversation may impair the ability of the agent to establish a continuous context over longer periods of time.

As part of future work, the severity-based, context-sensitive conversational decision-making framework will be extended into a full-scale, real-world application by integrating the proposed framework with an existing hardware-based women's safety system. In addition, severity-based context sensitivity can be further enhanced by making the system adaptive to different contexts and enabling the parameterization of severity mappings. For example, the same indicators, such as fear, authority misuse, and boundary violations, could have somewhat different interpretations across various cultures. This is an issue that will need to be addressed in future research.

V. Conclusion

This research presents a framework that evaluates women's safety both online and offline by integrating a decision-making framework that merges deterministic control methods with a restricted Large Language Model (LLM). The method consists of four main components, which include threat perception and severity assessment, together with guidance strategy control and language implementation. The system applies structured threat classification together with heuristic severity assessment to create safety profiles that use actual indicators for emergency response evaluation. The framework uses severity-based control signals to control various aspects of response development, response delivery, and decision-making for escalation times. The LLM serves as a language generation module whose output stays within specific boundaries while using parameter-efficient fine-tuning methods. Users can maintain consistent behavior throughout different interactions because the system effectively manages conversational context.

The results show enhanced alignment of generated guidance through distributional analysis, together with behavioral auditing, controlled baseline comparisons, and user-centered statistical evaluation. The findings show that conditioning the system on severity levels leads to better user responses, together with higher user trust. More important than everything else, the study shows that organizations must establish definite control mechanisms for their LLM deployments in situations that require safe performance.

This work introduces a practical framework in which LLMs operate as controlled language generation components within women's safety systems. The system enables context-sensitive and proportionate guidance, which solves the current emergency technology system weaknesses. The study demonstrates that structured control combined with large language models enables safe decision-making, which maintains safety and explanation capabilities in real-world emergencies.

Data Availability Statement

The dataset used in this study was collected by the authors through a trauma-informed survey with informed consent and anonymization procedures. The dataset is publicly available in the APERTA repository at https://doi.org/10.48623/aperta.286830 for research and reproducibility purposes.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

C.J.R.: Conceptualization, Methodology, Software, Data curation, Formal analysis, Visualization, Writing—original draft. M.M.: Supervision, Conceptualization, Methodology, Validation, Writing—review & editing. V.I.M.: Investigation, Resources, Data curation, Writing—review & editing. I.T.J.S.: Investigation, Resources, Data curation, Writing—review & editing. All authors had approved the final version.

Acknowledgment

The authors would like to express their sincere gratitude to all participants who generously shared their experiences, making this research possible. Their willingness to contribute sensitive insights was essential to the development and evaluation of the proposed women's safety decision-support framework. The authors also acknowledge the support of peers and colleagues who provided constructive feedback during the design and refinement of the survey instrument and system methodology. The authors also sincerely thank Karunya Institute of Technology and Sciences for providing the academic environment and opportunity to carry out this research work.

References

[1] World Health Organization, *Violence Against Women Prevalence Estimates, 2018*, Geneva, Switzerland: WHO Press, 2021.
[2] K. M. Devries, J. Y. T. Mak, C. García-Moreno *et al*., "The global prevalence of intimate partner violence against women," *Science*, vol. 340, no. 6140, pp. 1527–1528, 2013. doi: 10.1126/science.1240937
[3] Pew Research Center, *The State of Online Harassment*, Washington, DC, USA: Pew Research Center, 2021.
[4] N. Henry and A. Powell, "Technology-facilitated sexual violence: A literature review," *Violence Against Women*, vol. 24, no. 5, pp. 567–602, 2018. doi: 10.1177/1077801218756130
[5] D. K. Citron, *Hate Crimes in Cyberspace*, Cambridge, MA, USA: Harvard University Press, 2014.
[6] Plan International, *Free to Be Online? Girls' and Young Women's Experiences of Online Harassment*, Plan International Report, 2020.
[7] L. Kelly, *Surviving Sexual Violence*, Cambridge, U.K.: Polity Press, 1988.
[8] FRA – European Union Agency for Fundamental Rights, *Violence Against Women: An EU-Wide Survey*, Luxembourg: Publications Office of the European Union, 2014.
[9] S. Hinduja and J. W. Patchin, "Cyberbullying and self-esteem," *J. School Health*, vol. 85, no. 11, pp. 714–722, 2015. doi: 10.1111/josh.12302
[10] UN Women, *Online and ICT-Facilitated Violence Against Women and Girls*, New York, NY, USA: UN Women Report, 2022.
[11] T. B. Brown, B. Mann, N. Ryder *et al*., "Language models are few-shot learners," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901, 2020.
[12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423
[13] A. Vaswani, N. Shazeer, N. Parmar *et al*., "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, pp. 5998–6008, 2017.
[14] A. Radford, J. Wu, R. Child *et al*., *Language Models are Unsupervised Multitask Learners*, OpenAI Technical Report, 2019.
[15] R. Bommasani, D. A. Hudson, E. Adeli *et al*., "On the opportunities and risks of foundation models," arXiv preprint, arXiv:2108.07258, 2021.
[16] W. X. Zhao, K. Zhou, J. Li *et al*., "A survey of large language models," *ACM Computing Surveys*, vol. 55, no. 12, pp. 1–35, 2023. doi: 10.1145/3571730
[17] A. Chowdhery, S. Narang, J. Devlin *et al*., "PaLM: Scaling language modeling with pathways," arXiv preprint, arXiv:2204.02311, 2022.
[18] OpenAI, "GPT-4 technical report," arXiv preprint, arXiv:2303.08774, 2023.
[19] J. Wei, X. Wang, D. Schuurmans *et al*., "Chain-of-thought prompting elicits reasoning in large language models," *Advances in Neural Information Processing Systems*, vol. 35, pp. 24824–24837, 2022.
[20] P. Liang, R. Bommasani, T. Lee *et al*., "Holistic evaluation of language models," arXiv preprint, arXiv:2211.09110, 2022.
[21] United Nations, *Transforming Our World: The 2030 Agenda for Sustainable Development*, New York, NY, USA: United Nations, 2015.
[22] E. G. Krug, L. L. Dahlberg, J. A. Mercy *et al*., *World Report on Violence and Health*, Geneva, Switzerland: World Health Organization, 2002.
[23] S. Sarkar and S. Das, "Technology-based interventions for women safety," *International Journal of Engineering & Technology*, vol. 7, no. 2, pp. 236–240, 2018.
[24] M. Gurjar and A. Bhardwaj, "Smart wearable devices for women safety," *Procedia Computer Science*, vol. 167, pp. 1872–1879, 2020. doi: 10.1016/j.procs.2020.03.208
[25] Government of India, *Criminal Law (Amendment) Act, 2013*, New Delhi, India: Government of India, 2013.
[26] J. Kaur and S. Garg, "Addressing domestic violence against women: An unmet need," *Journal of Family Violence*, vol. 29, no. 6, pp. 593–603, 2014. doi: 10.1007/s10896-014-9629-8
[27] OECD, *Gender-Based Violence and the Role of Public Policy*, Paris, France: OECD Publishing, 2019. doi: 10.1787/d1e5f3ce-en
[28] R. Bhattacharyya, "Understanding the spatialities of sexual assault against Indian women in India," *Gender, Place & Culture*, vol. 22, no. 10, pp. 1340–1356, 2015. doi: 10.1080/0966369X.2014.969684
[29] National Crime Records Bureau, *Crime in India 2021*, New Delhi, India: Ministry of Home Affairs, Government of India, 2022.
[30] S. Turgay and S. Özyurt, "Dynamic ergonomic risk assessment with REBA and fuzzy multi-criteria decision making: Addressing repetitive movements," *Spectrum of Decision Making and Applications*, vol. 4, no. 1, pp. 1–17, 2025. doi: 10.31181/sdmap4157
[31] K. Ullah, N. Rehman, and A. Ali, "A robust circular complex intuitionistic fuzzy framework for optimizing agricultural robot decision-making under uncertain environmental conditions," *Spectrum of Decision Making and* Applications, pp. 1–42, 2026. doi: 10.31181/sdmap41202770