SAKHA: A Chatbot-Based Support System for Legal and Mental Health Aid to Women & Children Cybercrime Victims

Srishti Gupta, Indian Institute of Technology Patna, India srishti_2021cs38@iitp.ac.in
Ashwani Kumar, Indian Institute of Technology Patna, India ashwani_1911cs06@iitp.ac.in
Sourav Kumar Dandapat, Indian Institute of Technology Patna, India sourav@iitp.ac.in
Pankaj Kumar, All India Institute of Medical Sciences Patna cppankajkumar13@gmail.com
Meha Jain, All India Institute of Medical Sciences Patna jainmeh@gmail.com
Tulika Shankar, Gujarat National Law University shankar.tulika14@gmail.com

Abstract In a digital world where cybercrimes are on the rise, women and children in India face unique challenges that are often overlooked by existing solutions. The globally available platforms fail to address the nuances of Indian laws and the cultural stigma surrounding sensitive issues like mental health and legal aid. Our chatbot bridges this gap, offering a safe, confidential, and culturally relevant space for victims of cybercrime to seek help. Tailored specifically to the Indian context, the chatbot provides legal guidance based on local laws, empowering users to understand their rights and encouraging them to seek help. It also offers compassionate mental health counselling, addressing the emotional toll of cybercrime without the fear of judgment or taboo. By providing a seamless blend of legal aid and psychological support, this chatbot ensures that women and children affected by cybercrime have a reliable, accessible, and non-threatening resource to turn to.

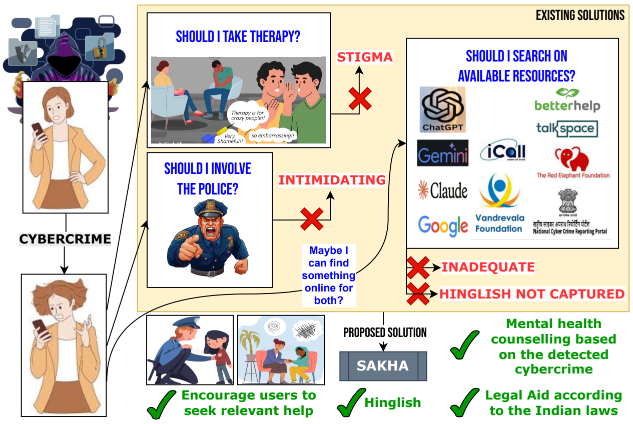

EXISTING SOLUTIONS
CYBERCRIME
SHOULD I TAKE THERAPY?
Therapy is for crazy people!!
Very Shameful!!
so embarrassing!!
STIGMA
SHOULD I INVOLVE THE POLICE?
INTIMIDATING
Maybe I can find something online for both?
SHOULD I SEARCH ON AVAILABLE RESOURCES?
ChatGPT
betterhelp
talkspace
Gemini
iCall
The Red Elephant Foundation
Claude
Google
Vandrevala Foundation
राष्ट्रीय साइबर अपराध रिपोर्टिंग पोर्टल
National Cyber Crime Reporting Portal
INADEQUATE
HINGLISH NOT CAPTURED
PROPOSED SOLUTION
SAKHA
Mental health counselling based on the detected cybercrime
Encourage users to seek relevant help
Hinglish
Legal Aid according to the Indian laws

Figure 1: Overview of challenges faced by cybercrime victims, highlighting the significant psychological distress experienced due to stigma, intimidation, and lack of adequate support, motivating the need for an integrated mental health and legal assistance framework.

1 Introduction

Cybercrime is a global challenge that affects individuals across all countries and demographics. Broadly, it refers to illegal activities carried out using computer systems or the internet (Sukhai, 2004). While the prevalence and nature of cybercrime may vary depending on the region, common vulnerabilities, such as lack of digital literacy, limited awareness of legal recourse, and psychological trauma, affect victims worldwide. However,

https://doi.org/10.3384/nejlt.2000-1533.2026.6068

the challenges faced by cybercrime victims in India differ in some important ways. In recent years, the rapid expansion of the internet has led to an alarming rise in cybercrime in India. The National Crime Records Bureau (NCRB), India, posted (Delhi, 2024) on 7^{th} February 2024, the statistical data on crimes in its publication "Crime in India" (Gogia et al., 2024) that concluded a total of 50035 in 2020, 52974 cases in 2021, and 65893 in 2022. Figure 2 depicts the State/Union Territory-wise details of cases registered under cybercrimes.

These crimes range from cyberstalking, online harassment, identity theft, and financial fraud to more severe forms of exploitation and abuse. The psychological, emotional, and financial toll of such crimes is often significant on Indian women and children (Shree, 2025), leaving victims feeling isolated, distressed, and without recourse. Legal frameworks, although present, are underutilized, and victims often lack the knowledge or resources to navigate the complex legal system. Additionally, significant cultural stigmas surrounding mental health counselling and legal reporting further hinder victims from seeking help. These challenges are compounded by the linguistic diversity of India, where users may be more comfortable communicating in regional languages or in a mix of Hindi and English (Hinglish), yet many platforms and resources are not designed to accommodate these needs.

The victims of cybercrimes often experience significant psychological distress. In India, there is a significant social stigma surrounding mental health counselling, with many people viewing it negatively or as a sign of weakness (Gaiha et al., 2020). While some platforms aim to support victims of cybercrimes, the focus is usually not specifically on cybercrime-related trauma. Instead, these platforms cater to broader mental health issues, which may not address the unique psychological burdens of cybercrime victims. They do not provide immediate, real-time guidance to prevent or address further cybercrime-related risks. At the same time, legal frameworks, although available, remain underutilized due to the lack of awareness, accessibility, and the intimidating process of filing complaints. The significant lack of digital literacy and understanding of legal rights among many users prevents them from seeking help effectively. Moreover, despite the availability of legal resources, there is often no integration with psychological counselling services.

To address these pressing challenges, we have developed a rule-based chatbot software tool, सखा **(SAKHA**[1] meaning 'friend' in Hindi) designed to provide preliminary mental health counselling and legal aid to women and children who have experienced cybercrimes. This tool combines advancements in artificial intelligence, natural language processing, and psychological assessment techniques to create an accessible, empathetic, and informative platform that supports victims in real time. At the core of the tool's functionality is its integration with the **Depression Anxiety Stress Scales** (DASS-21) (Lovibond and Lovibond, 1995), a well-established psychological assessment tool to measure the severity of depression, anxiety, and stress in individuals, as recommended by our in-house mental health counselor. Based on the outcomes, the tool offers practical suggestions for mental health improvement, including stress-relief techniques, self-care practices, and resources for further support. This feature is crucial as cybercrime victims often suffer from mental health issues such as anxiety and depression, which require immediate attention and care. In parallel, the tool is trained to identify various forms of cybercrimes based on user-provided details. After identifying the type of cybercrime, the tool provides the user with tailored advice on how to protect themselves from further incidents. Additionally, the tool integrates legal information specific to India's cyber laws, offering real-time, contextually relevant legal advice on the Indian Penal Code (IPC), the Information Technology Act, and specific laws that protect against online harassment, cyberstalking, and digital exploitation. It also encourages users to report their experiences to relevant authorities by informing victims about their legal rights and the steps they can take. The tool aims to not only provide emotional support but also empower users to take action in seeking justice and protection. If the user wishes to voice a concern, we encourage them to consult a qualified professional after each module, i.e., if they would like their grievance to be addresses by a professional via WhatsApp, as any additional requests or concerns can be addressed through professional guidance only.

To systematically evaluate SAKHA's capabilities and value to users, we investigate the following research questions:

RQ1: Can Hinglish cybercrime complaints be accurately classified into crime types using our proposed dataset and model?

RQ2: Can user complaints be reliably mapped to the correct Indian legal provisions?

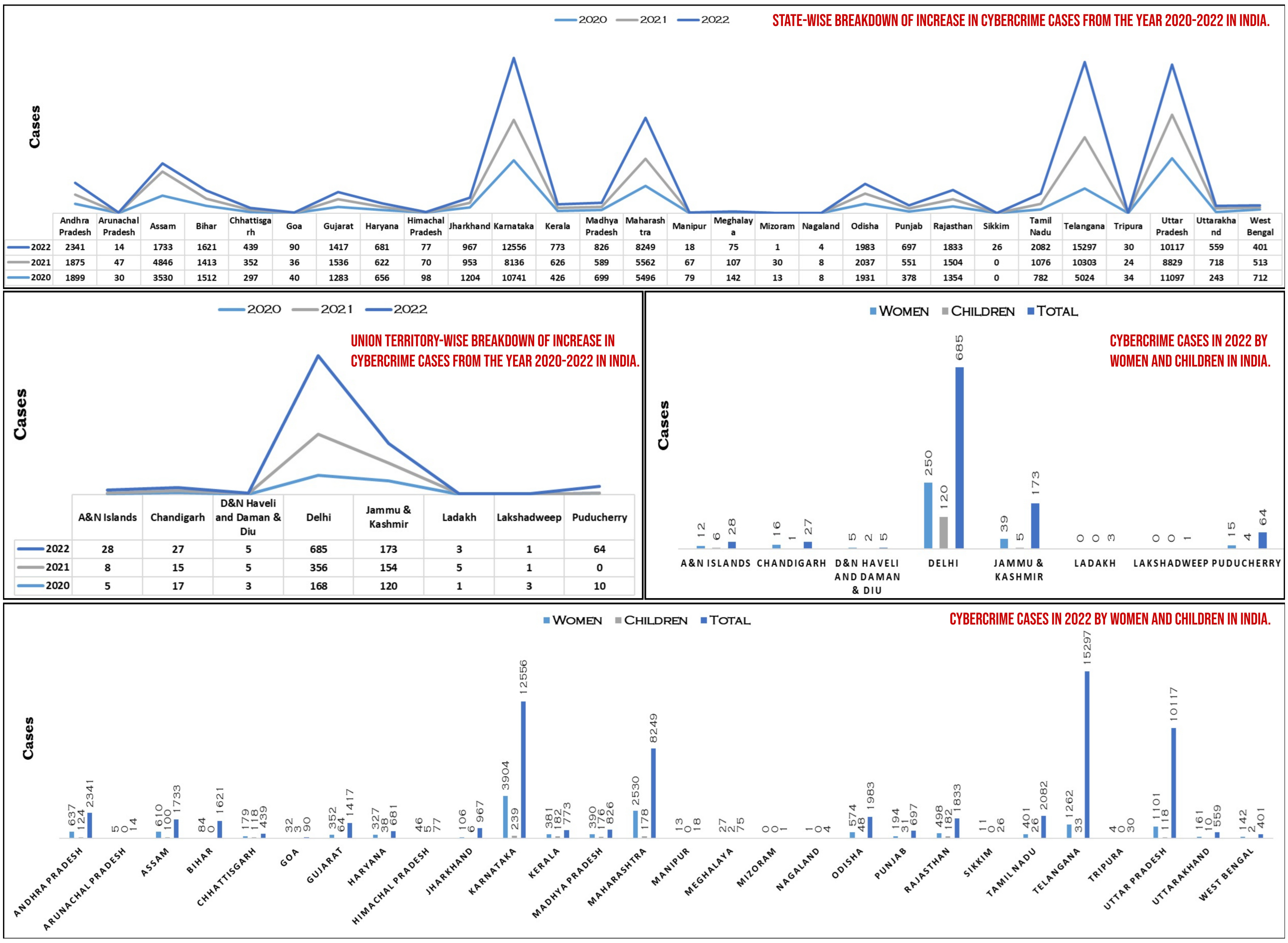
RQ3: Do the legal suggestions produced by the system encourage users to file a complaint?

RQ4: Does the counselling module provide responses that participants find empathetic and helpful?

RQ5: Do users perceive SAKHA as usable and supportive in simulated complaint scenarios?

These questions guide the design and evaluation of the system, ensuring that SAKHA not only performs well technically but also creates a meaningful impact for users. Guided by these research questions, our key contributions are as follows:

[1] https://github.com/srishtigupta253/SAKHA.git



	Andhra Pradesh	Arunachal Pradesh	Assam	Bihar	Chhattisgarh	Goa	Gujarat	Haryana	Himachal Pradesh	Jharkhand	Karnataka	Kerala	Madhya Pradesh	Maharashtra	Manipur
2022	2341	14	1733	1621	439	90	1417	681	77	967	12556	773	826	8249	18
2021	1875	47	4846	1413	352	36	1536	622	70	953	8136	626	589	5562	67
2020	1899	30	3530	1512	297	40	1283	656	98	1204	10741	426	699	5496	79

	Meghalaya	Mizoram	Nagaland	Odisha	Punjab	Rajasthan	Sikkim	Tamil Nadu	Telangana	Tripura	Uttar Pradesh	Uttarakhand	West Bengal
2022	75	1	4	1983	697	1833	26	2082	15297	30	10117	559	401
2021	107	30	8	2037	551	1504	0	1076	10303	24	8829	718	513
2020	142	13	8	1931	378	1354	0	782	5024	34	11097	243	712

2020
2021
2022
UNION TERRITORY-WISE BREAKDOWN OF INCREASE IN CYBERCRIME CASES FROM THE YEAR 2020-2022 IN INDIA.
Cases

	A&N Islands	Chandigarh	D&N Haveli and Daman & Diu	Delhi	Jammu & Kashmir	Ladakh	Lakshadweep	Puducherry
2022	28	27	5	685	173	3	1	64
2021	8	15	5	356	154	5	1	0
2020	5	17	3	168	120	1	3	10

Figure 2: State-wise and union territory-wise increase in reported cybercrime cases in India from 2020 to 2022, with a breakdown of cases filed by women and children in 2022. *Source of statistics: NCRB report (Gogia et al., 2024).*

1. We design and implement a **legal-intent mapping module** that links user-generated cybercrime complaints in Hinglish to the relevant provisions of the Indian IT Act and IPC. This provides victims with transparent, statute-level information regarding potential legal recourse.

2. We develop an **integrated mental health counselling module** that leverages conversational strategies to provide supportive responses, thereby enabling stigma-free access to preliminary psychological assistance through dialogue.

3. We develop a **novel Hinglish cybercrime complaint dataset**, together with a transformer-based **crime-type classification mechanism** capable of handling code-mixed text, spelling variance, and class imbalance.

4. We conduct **extensive experiments and user studies** (100 participants) to evaluate the effectiveness of the proposed system. Results demonstrate (*i*) improved classification accuracy on Hinglish complaints, (*ii*) reasonable precision/recall in statute mapping, and (*iii*) positive usability ratings and psychological acceptability of the counselling interface.

2 Related Work

Cybercrimes, particularly those targeting women and children, are a growing concern globally. As technology continues to evolve, so do the methods employed by perpetrators of cybercrimes. Existing solutions that combine mental health support and legal guidance for cybercrime victims are limited and often fragmented, with several notable drawbacks that hinder their effectiveness, particularly in the context of India.

2.1 Mental Health Support for Cybercrime Victims

In India, online mental health support systems like **iCALL** (Sriram et al., 2012), **Vandrevala Foundation**

(VF) (Vandrevala and Hiranandani-Vandrevala, 2008), and **The Red Elephant Foundation** (REF) (Jayakumar, 2013) are emerging, providing psychological support and helplines for those affected by trauma, including cybercrime. These resources are not contextualized to handle issues uniquely tied to cybercrime. Furthermore, in the international context, several online platforms offer psychological support for individuals dealing with various types of trauma. For example, platforms such as 7 **Cups** (Moriarty, 2013), **BetterHelp** (Matas and Bragonier, 2013), and **Talkspace** (Frank and Frank, 2012) provide virtual counselling services via text, audio, or video chat. These platforms utilize licensed therapists to help individuals work through their mental health concerns, which may include the effects of online harassment or cyberbullying. However, they often require users to engage in paid services or lengthy waiting periods for access to a counselor. These platforms are often inaccessible to large segments of the population, particularly vulnerable groups such as women and children in lower-income or rural areas. They are tailored to the cultural and societal nuances of specific regions, especially in countries like India, where stigma surrounding mental health can discourage individuals from seeking help. However, these platforms focus solely on psychological support and fail to incorporate legal counselling or guidance on how victims can take legal action against their abusers—a critical need for those impacted by cybercrimes. Additionally, cultural differences and the frequent requirement to engage in paid services, along with long waiting periods for counselor access, further limit their accessibility, particularly for Indian women and children.

2.2 Legal Aid Solutions for Cybercrime Victims

In India, the **National Cyber Crime Reporting Portal** (Government, 2023) has been developed to assist victims in reporting cybercrimes directly to law enforcement. Additionally, organizations like **Cyber Peace Foundation** (CPF) (Kumar, 2005) and **Tata Consultancy Services' Cybersecurity Campaign** (TCSCC) (Limited, 2020) have taken steps to educate and empower victims of cybercrimes. However, the significant lack of digital literacy and understanding of legal rights among many users and the intimidating process of filing a complaint prevents them from seeking help effectively. Despite the availability of legal resources, there is often no integration with psychological counselling services. Victims who report a cybercrime may still be left to manage the mental health effects of their trauma without any immediate, supportive interventions. Internationally, several online resources, such as **Cyber Civil Rights Initiative** (CCRI) (Jacobs, 2012) and **Cybercrime Support Network** (CSN) (Judge, 2018), offer legal advice and support for victims of cybercrime, particularly in cases of online harassment, revenge porn, and identity theft. These platforms often provide information on reporting cybercrimes, the laws protecting victims, and what steps to take next. These platforms may not be aware of the specific laws in different countries, making their advice less relevant for victims outside the platform's jurisdiction. In the case of India, where the legal system surrounding cybercrimes is rapidly evolving, international platforms may not provide information on the latest Indian cyber laws or how to report cybercrimes in the country effectively. To battle the mentioned shortcomings, also shown in Table 1, SAKHA is intended to be the first point of contact for victims, before they seek professional counselling or legal assistance.

Platform	Counselling	Legal Aid	Indian Content	Hinglish	Type	Resource
iCALL	✓	✗	✓	✓	General	-
VF	✓	✗	✓	✓	General	-
REF	✓	✗	✓	✓	General	-
7 Cups	✓	✗	✗	✗	General	-
BetterHelp	✓	✗	✗	✗	General	-
Talkspace	✓	✗	✗	✗	General	-
NCRB	✗	✓	✓	✓	Cybercrime	-
CPF	✗	✓	✓	✓	Cybercrime	-
TCSCC	✗	✓	✓	✓	Cybercrime	-
CCRI	✗	✓	✗	✗	Cybercrime	-
CSN	✓	✗	✓	✓	Cybercrime	-
ChatGPT	✓	✓	✗	✗	General	High
Gemini	✓	✓	✗	✗	General	High
Deepseek	✓	✓	✗	✗	General	High
Claude	✓	✓	✗	✗	General	High
Llama	✓	✓	✗	✗	General	High
SAKHA	✓	✓	✓	✓	Cybercrime	Low

Table 1: Comparison of existing works.

2.3 Cybercrime Classification in Code-Mixed Datasets

Previous work by (Sedik and Romadhony, 2023) proposed a crime information extraction system for Indonesian news using Named Entity Recognition (NER) and Support Vector Machines (SVM) with TF-IDF features. While effective for monolingual settings, the reliance on handcrafted features and fixed crime categories limits adaptability, particularly in multilingual or code-mixed contexts. (Islam et al., 2022) applied supervised models to classify Bangla crime news using CountVectorizer and TfidfVectorizer. These shallow feature extraction methods fail to capture semantic nuances, especially in linguistically complex or code-mixed texts, and lack portability across languages. (Rahma and Romadhony, 2021) developed a rule-based system combining dependency parsing and ontology-driven classification. Despite accurate entity extraction, the approach suffers from rigidity and poor generalization to dynamic, informal, or code-mixed complaints due to language-specific rule dependencies. (Prabhu et al., 2023) introduced a Cyber Complaint Automation System employing RAKE for keyword extraction and ERNIE for classification. The use of external knowledge sources enhances accuracy, but the system is constrained by predefined crime categories, class imbalance, and lack of multilingual or code-mixed support. (Pongpaichet et al., 2024) proposed CAMELON for Thai news-based crime classification using BiLSTM with Thai2Vec and transformer models like WangchanBERTa, mBERT, and XLM-RoBERTa. XLM-RoBERTa achieved the best results; however, the focus on news articles and absence of privacy-preserving mechanisms limit real-world deployment in complaint systems. Despite progress, most approaches rely on traditional feature extraction (e.g., TF-IDF, bag-of-words), rule-based methods, or monolingual models that underperform on code-mixed data. Additionally, class imbalance and privacy concerns remain underexplored, limiting their applicability in real-world, multilingual cybercrime reporting systems.

3 Dataset

The primary objective of this data set is to allow the chatbot to accurately classify and identify the different types of cybercrimes that victims might report during their interaction. The step-by-step process of dataset creation follows.

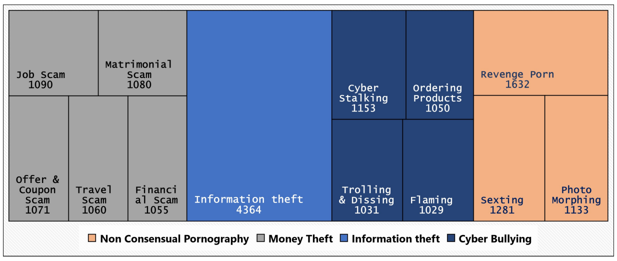
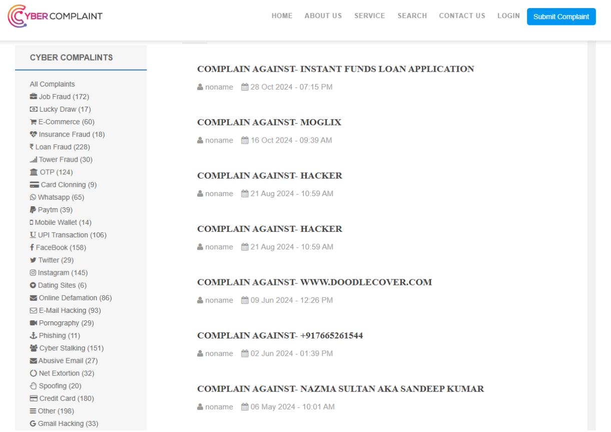


Figure 3: Distribution of the number of complaints in each category of the dataset collected from 2020 – 2025.

3.1 Data Collection and Annotation

To construct the dataset for this study on cybercrimes, we leverage the publicly available cyber complaint forum[2] that aggregates reports from Indian victims, spanning a diverse array of cybercrime incidents (Figure 4). These complaints typically detail the grievances of the victims, with some providing context about the perpetrators. To build the dataset, data were collected systematically from the forum using automated web scraping implemented with BeautifulSoup. To ensure balanced representation across crime categories, an equal number of complaints was collected per category. The source platform has no direct affiliation with law enforcement, and its terms permit republication for research purposes.



CYBER COMPALINTS

All Complaints
Job Fraud (172)
Lucky Draw (17)
E-Commerce (60)
Insurance Fraud (18)
₹ Loan Fraud (228)
Tower Fraud (30)
OTP (124)
Card Clonning (9)
Whatsapp (65)
Paytm (39)
Mobile Wallet (14)
UPI Transaction (106)
FaceBook (158)
Twitter (29)
Instagram (145)
Dating Sites (6)
Online Defamation (86)
E-Mail Hacking (93)
Pornography (29)
Phishing (11)
Cyber Stalking (151)
Abusive Email (27)
Net Extortion (32)
Spoofing (20)
Credit Card (180)
Other (198)
Gmail Hacking (33)

COMPLAIN AGAINST- INSTANT FUNDS LOAN APPLICATION
noname 28 Oct 2024 - 07:15 PM

COMPLAIN AGAINST- MOGLIX
noname 16 Oct 2024 - 09:39 AM

COMPLAIN AGAINST- HACKER
noname 21 Aug 2024 - 10:59 AM

COMPLAIN AGAINST- HACKER
noname 21 Aug 2024 - 10:59 AM

COMPLAIN AGAINST- WWW.DOODLECOVER.COM
noname 09 Jun 2024 - 12:26 PM

COMPLAIN AGAINST- +917665261544
noname 02 Jun 2024 - 01:39 PM

COMPLAIN AGAINST- NAZMA SULTAN AKA SANDEEP KUMAR
noname 06 May 2024 - 10:01 AM

Figure 4: Snapshot of Cyber Complaint Forum.

3.2 Data Pre-Processing

This phase focuses on preprocessing the dataset by anonymizing user identities. Personal information, such as the victim's name, phone number, Aadhaar number, and other identifying details, has been removed. All cases were subsequently reviewed manually to confirm that no personally identifiable information remained. Additionally, specific elements like crime dates, monetary amounts lost, and any other sensitive data are also excluded. Finally, informal language is standardized by creating a comprehensive vocabulary that maps slang terms to their formal equivalents (Table 2), ensuring consistency and clarity throughout the dataset.

Slang	Equivalent	Slang	Equivalent
ain't	am not	Hella	extremely
doesn't	does not	Legit	legitimate
how'd'y	how do you	On point	perfect
y'all	you all	Cringy	embarrassing
i'll	I shall	Meh	unimpressive

Table 2: Sample slang vocabulary for conversion to standard terms.

3.3 Data Annotation

The final step involves manually annotating the dataset with multiple labels (mentioned in Appendix A) that correspond to each complaint, ensuring that no relevant cybercrime type is overlooked when a user logs in. To achieve this, we apply the aforementioned definitions to each entry in the dataset. These annotations are then reviewed and validated by our onboard legal help to ensure accuracy and to confirm that no pertinent labels have been omitted. Table 3 summarizes the key statistics of the resulting dataset. The corpus comprises [14, 849]

[2] https://cybercomplaint.in/index.html

Statistics	Value
Total complaints	14, 849
Collection period	2020 − 2025
Number of categories	16
Min labels per complaint	2
Max labels per complaint	6
Avg. labels per complaint	3
Avg. complaint length (tokens)	316

Table 3: Dataset Statistics.

complaints in total, collected from 2020 − 2025 across 16 cybercrime categories. Since complaints frequently involve overlapping crime types, each entry carries between two and six labels, with an average of three labels per complaint. The average complaint length is approximately 316 tokens, reflecting the detailed, narrative nature of victim-authored reports.

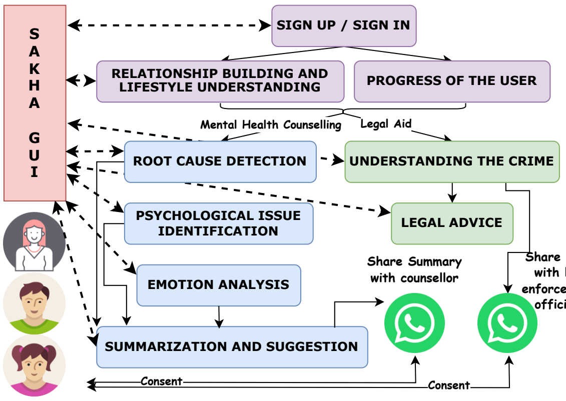

S A K H A G U I
SIGN UP / SIGN IN
RELATIONSHIP BUILDING AND LIFESTYLE UNDERSTANDING
PROGRESS OF THE USER
Mental Health Counselling
Legal Aid
ROOT CAUSE DETECTION
UNDERSTANDING THE CRIME
PSYCHOLOGICAL ISSUE IDENTIFICATION
LEGAL ADVICE
EMOTION ANALYSIS
Share Summary with counsellor
Share case with law enforcement official
SUMMARIZATION AND SUGGESTION
Consent
Consent

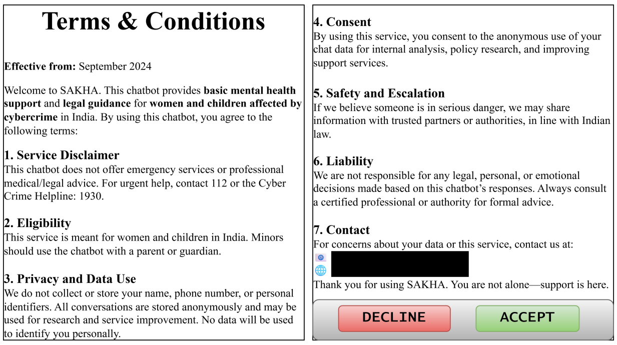
Figure 5: Phase-wise architecture of SAKHA.

4 Methodology

The entire process, as shown in Figure 5 and illustrated with a case study in Figure 13, can be broken down into several key subcomponents, each playing a vital role in ensuring the system operates smoothly and effectively, once the user signs up/in the bot. Importantly, we don't store the user's real name. Instead, we ask them to provide a nickname, which allows us to refer to the victim in a more personalized and tailored manner. As shown in Figure 6, we also obtain the user's consent to use their data in anonymized form for purposes such as research, and data analysis.

4.1 Relation Building and Lifestyle Understanding

As recommended by our onboard professionals, effective therapeutic interventions are grounded in thoroughly understanding the client's background and current state before addressing sensitive topics to make them comfortable. A series of well-tailored rule-based questions (Appendix B) are presented to achieve this and gradually familiarize the user with the platform and its interface. This collects demographic and situational information, and various lifestyle factors, such as sleep patterns, physical activity, eating habits, etc. In addition to lifestyle factors, the bot also inquires about cultural elements, such as religion, the type of locality the user resides in, etc, to gain insights into the users' cultural backgrounds, experiences, and worldviews.



Terms & Conditions

Effective from: September 2024

Welcome to SAKHA. This chatbot provides **basic mental health support** and **legal guidance** for **women and children affected by cybercrime** in India. By using this chatbot, you agree to the following terms:

1. Service Disclaimer
This chatbot does not offer emergency services or professional medical/legal advice. For urgent help, contact 112 or the Cyber Crime Helpline: 1930.

2. Eligibility
This service is meant for women and children in India. Minors should use the chatbot with a parent or guardian.

3. Privacy and Data Use
We do not collect or store your name, phone number, or personal identifiers. All conversations are stored anonymously and may be used for research and service improvement. No data will be used to identify you personally.

4. Consent
By using this service, you consent to the anonymous use of your chat data for internal analysis, policy research, and improving support services.

5. Safety and Escalation
If we believe someone is in serious danger, we may share information with trusted partners or authorities, in line with Indian law.

6. Liability
We are not responsible for any legal, personal, or emotional decisions made based on this chatbot's responses. Always consult a certified professional or authority for formal advice.

7. Contact
For concerns about your data or this service, contact us at:
[redacted]
Thank you for using SAKHA. You are not alone—support is here.

Figure 6: Terms & Conditions of SAKHA.

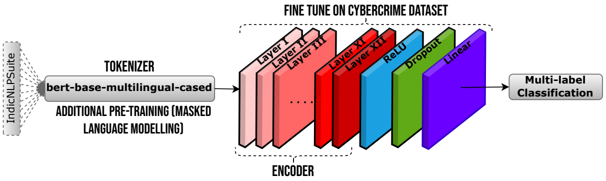

FINE TUNE ON CYBERCRIME DATASET
IndicNLPSuite
TOKENIZER
bert-base-multilingual-cased
ADDITIONAL PRE-TRAINING (MASKED LANGUAGE MODELLING)
Layer I
Layer II
Layer III
Layer XI
Layer XII
ReLU
Dropout
Linear
ENCODER
Multi-label Classification

Figure 7: Crime identification Architecture.

4.2 Root Cause Detection

This phase of mental health counselling seeks to answer **What** is the nature of the crime?, **Who** might be responsible? and whether the incident has been formally **reported**. To minimize cognitive overload, the bot first allows the user to express themselves openly by posing the question, `What brings you here?` This prompt lets the user describe their situation freely, offering insights into the crime and potential suspect identity. This input could be text or audio converted to transcript using Whisper (Radford et al., 2023), which is first cleaned (Section 3.2) and then processed using a multi-label classification model to predict the most likely types of cybercrimes. As illustrated in Figure 7, our architecture decomposes the Hinglish classification task into three core stages: (*I*) tokenizer vocabulary adaptation, (*II*) domain-adaptive pretraining via masked language modeling (MLM), and (*III*) supervised fine-tuning for multi-

Complaint	Label	Culprit
I am [redacted], a student of the [redacted] staying at [redacted]. This is our exam time so we are trying to focus on our studies. At this time I am getting unknown phone calls from an unknown number. The same thing is happening to my friends. At first, I thought it was known to us so I tried to call back but no answer from that side. So we blocked the number. After blocking the number he starts sending abusive texts to my phone number and threatening me	Cyber Bullying Cyber Stalking Flaming	Unknown
The website called lifefriendship has taken my [redacted] in the name of friendship and dating and not picking the call and saying still they need money. The person incharge for this is Ms.Nisha and her number is 9614636976.	Money Theft Matrimonial/Dating scam	Ms.Nisha
I received a call on my telephone number from a female Akansha Sharma from telephone number 7042433301 for renewal of my driving license to which I agreed. I sent her all details of my license and paid amounts of Rs [redacted], [redacted] and [redacted] to M B Holidayss for renewal of my license. Next she said that there are challans pending against my car [redacted]. Though I refused that there are no challans pending she insisted that I pay immediately else my car will be black list.	Money Theft Financial scam Information Theft	Akansha Sharma

Table 4: Sample complaints taken from Cyber Complaint forum and their respective multi-label tags.

label classification. In Stage *I*, we augment the vocabulary of the pretrained BERT multilingual tokenizer[3] (Devlin et al., 2019) using an existing corpus of unlabeled Hinglish and transliterated Hindi. Frequent subword units are statistically extracted and merged into the tokenizer's vocabulary using WordPiece, reducing [UNK] token frequency and improving semantic segmentation of Romanized Hindi words. In Stage *II*, we continue pretraining a transformer encoder (BERT model selected based on results in Table 6) using MLM on the same unlabeled Hinglish corpus. In Stage *III*, we fine-tune the MLM-adapted encoder on our curated cybercrime classification dataset. The classification head consists of a dense projection over the $[CLS]$ token followed by a ReLU non-linearity, dropout, and a final linear layer for multi-label prediction. Once the potential cybercrimes are classified, the bot validates these predictions by asking the user confirmation questions. For the "Who" aspect, we utilize Coreference Resolution to identify the mentioned real-world entity. These identified entities are then presented to the user as interactive buttons, allowing the user to select which individuals might be responsible for the distress they described. We also provide a "Someone unknown" option and allow users to type in the culprit if no other options are suitable.

4.3 Psychological Issue Identification

Based on the suggestion of our onboard counselors, we evaluate if the user is experiencing any underlying psychological issue using the DASS-21 questionnaire (Lovibond and Lovibond, 2011). It is a set of three self-report scales designed to measure the emotional states of depression, anxiety, and stress. The user must read each statement and choose a number 0-NEVER, 1-SOMETIMES, 2-OFTEN, or 3-ALMOST ALWAYS, which indicates how much the statement applied to them over the past week. Once the user gives their input for each question, pre-defined metrics for the score help declare the severity of each disorder.

4.4 Emotion Analysis

Our on-board mental health professionals recommend helping clients recognize their negative emotions as a key step in therapy. To support this, SAKHA uses the Linguistic Inquiry and Word Count (LIWC) tool (Pennebaker et al., 2007) to systematically analyze the user's responses to the open-ended question in Section 4.2. LIWC detects and categorizes the negative emotions expressed, which SAKHA then uses to provide tailored insights and guidance aimed at helping users understand and manage these emotions effectively.

4.5 Legal Aid

This phase focuses on understanding the crime to display laws that apply to the crime. If the user is receiving mental health counselling and seeks legal aid, their input from *Root Cause Detection* can be transferred to the *Legal Aid* module, after consent. If they start with legal aid, we ask the open-ended question. With the assistance of our legal experts, we have developed a specialized database that maps keywords to the relevant sections from Indian law pertaining to cybercrime. We systematically identify the user's intent and map it to corresponding legal provisions. During the process of identifying applicable laws, we focus on detecting the precise intents associated with the text. Transformer models could be employed for feature extraction in this case but due to their computational overhead (Table 5), we leverage Doc2Vec, to effectively identify words in the

[3] https://huggingface.co/google-bert/bert-base-multilingual-cased

user's input that align with specific legal intents.

Model	Accuracy	Response Time (sec)
GLOVE (Pennington et al., 2014)	0.68	30
Word2Vec (Mikolov et al., 2013)	0.71	47
Doc2Vec (Le and Mikolov, 2014)	0.89	49
Transformer (Vaswani et al., 2017)	0.91	137

Table 5: Results of the intent classification task.

4.6 Summarization & Suggestion

The final phase of the bot focuses on concisely summarizing the user's details for easier reference. If the user chooses to continue consulting with a professional in person, they can simply present this summary. They could also forward this summary via WhatsApp to a linked professional. Along with summarization, another goal of this phase is to provide the user with actionable solutions to address their concerns. The bot generates a tailored list of recommendations based on:

1. The **platform** where the crime occurred and how they could seek help from there.
2. The detected **crime-related** suggestions with information about how to avoid such a crime in the future, like how to block the perpetrator, and best practices for documenting the incident, such as saving screenshots.
3. Pre-defined suggestions by our counselors for
   (a) The detected **negative emotions** in the user's text fron Section 4.4.
   (b) Based on the **severity** of the disorder from Section 4.3. Say her symptoms are mild, then she is on the safer side, and no such suggestions are provided. In case she shows severe symptoms, she is advised to quickly consult a doctor, but in moderate cases, we display basic suggestions provided by the on-board counselors that have proven to be of help to other patients.
4. A list of **relevant legal sections** to reassure users that the law supports them and encourages filing a complaint
5. A list of relevant **evidence**[4] to reassure that the wrong that happened to them can be used as evidence to make their case strong.
6. Detailed procedure on how to **file a complaint** online and offline to government bodies.
7. Details of **NGOs** are mentioned to promote an optimistic outlook and empower users to feel more in control.

4.7 Progress of the User

This phase allows users to resume interactions, addressing unresolved issues or new concerns. The bot focuses on previously identified problems for users returning to continue their sessions. Specifically, the bot revisits the *Root Cause Detection* and *Psychological Issue Identification* phases to assess any progress or changes in the user's condition. Additionally, the bot re-evaluates key lifestyle factors to monitor for improvements or deterioration in the user's progress over time.

5 Evaluation and Results

In this Section, we analyze the different research questions listed in Section 1 to depict the effectiveness of SAKHA.

5.1 Automatic Evaluation

5.1.1 Baselines and Setup

In order to develop the model for crime classification, we compare various models based on their accuracy and response time when deployed on a server. We compare the following state of the art (SOTA) Hinglish models by fine-tuning them on our curated crime dataset: 1) MuRIL (Khanuja et al., 2021) 2) XLM-R (Conneau et al., 2020) 3) IndicBERT (Kakwani et al., 2020) 4) mBERT (Libovický et al., 2019). Alongside these, we employ our proposed pipeline from Section 4.2 by using the following models as encoders: 1) BERT (Devlin et al., 2019), 2) DeBERTa (He et al., 2021), 3) RoBERTa (Liu et al., 2019), 4) T5 (Raffel et al., 2020), 5) XLNet (Yang et al., 2020), 6) DistilBERT (Sanh et al., 2019), 7) ELECTRA (Clark et al., 2020). To conduct a fair comparison, we use the following hyperparameters: $8:1:1$ (train:val:test) ratio to obtain a classification accuracy, CrossEntropyLoss using the `AutoModelForSequenceClassification` with `problem_type=multi_label_classification` and AdamW optimization, 12 epochs, 16 batch size, and a learning rate of $2e-05$. The models were fine-tuned using an NVIDIA Tesla T4 GPU (16 GB VRAM, 12 GB RAM) and later deployed on PythonAnywhere for inference. Since PythonAnywhere is CPU-only (2 vCPUs, 3 GB RAM), all response times reported in Table 1 correspond to CPU execution.

5.1.2 Model Selection

To implement the Hinglish cybercrime multi-label classification task, we initially fine-tune several SOTA trans-

[4] This list of relevant evidence for each crime is generated with the help of our onboard legal expert.

Dataset	Model	Parameters	Micro			Macro			Subset	Label	Response	
			Precision	*Recall*	*F1-Score*	*Precision*	*Recall*	*F1-Score*	Accuracy	Accuracy	Time	
MLM: IndicNLPSuite, Fine-tune:Our	BERT	≈ 111˜112M	**0.8971**	**0.8456**	**0.8689**	<u>0.8519</u>	<u>0.7455</u>	<u>0.7879</u>	**0.6572**	**0.877**	<u>127</u>	→Best trade off
	DeBERTa	≈ 140˜141M	0.5494	0.5392	0.6597	0.3723	0.2902	0.3122	0.0907	0.528	328	
	RoBERTa	≈ 126˜127M	0.8154	0.5956	0.6884	0.3534	0.3384	0.3374	0.1331	0.661	277	
	T5	≈ 221M+	0.8451	0.6933	0.7617	0.7002	0.4551	0.4856	0.4311	0.703	249	
	XLNet	≈ 111˜112M	<u>0.8866</u>	<u>0.8424</u>	<u>0.8656</u>	**0.8612**	**0.7589**	**0.7974**	<u>0.6402</u>	0.680	311	
	DistilBERT	≈ 67˜68M	0.8379	0.4628	0.5962	0.2606	0.2224	0.2354	0.4227	0.736	**104**	
	ELECTRA	≈ 111˜112M	0.5676	0.1942	0.2894	0.0355	0.0625	0.0453	0.4017	<u>0.776</u>	273	
Fine-tune: Our	MuRIL	≈ 236.01M	0.7265	0.2485	0.3703	0.0930	0.1001	0.0933	0.0286	0.582	348	
	XLM-R	≈ 270.01M	0.8069	0.5171	0.6303	0.3102	0.2663	0.2658	0.5386	0.611	350	
	mBERT	≈ 177.01M	0.8829	0.7363	0.8030	0.6089	0.5252	0.5460	0.4363	0.684	333	
	IndicBERT	≈ 12.01M	0.7107	0.4801	0.5730	0.5472	0.4723	0.5069	0.4531	0.715	328	

Table 6: Comparison of classification performance across models. *Note:* ***bold*** *= best in column;* <u>*underline*</u> *= second-best.*

former models on our curated crime-specific dataset. Results in Table 6 depict BERT and XLNet clearly outperform the other models, delivering the strongest and most balanced results across both micro and macro metrics. Figure 12 presents a category-wise performance breakdown for all evaluated models, enabling a more granular analysis of classification behavior across different cybercrime types. **BERT** achieves the best overall performance, with the highest micro F1, strong macro F1, and the top subset and label accuracy, making it the most reliable choice for this task. XLNet closely follows and slightly leads in macro F1, indicating better handling of rare labels. BERT is chosen over XLNet in our work as it achieves the best trade-off, yielding the highest classification performance while maintaining the lowest response time during deployment. Models such as mBERT and T5 show moderate performance but struggle with macro recall, while RoBERTa, DistilBERT, and XLM-R exhibit high precision but low recall, suggesting bias toward frequent labels. ELECTRA, MuRIL, and DeBERTa perform the weakest overall, with very low macro scores and poor exact-match accuracy, highlighting challenges in capturing the full multi-label distribution.

5.1.3 Run of the App (RQ1, RQ2)

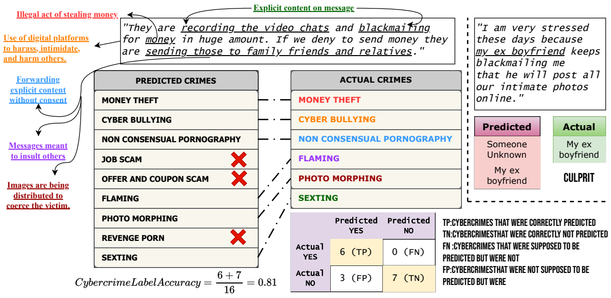
The bot functionalities were tested automatically using **Selenium** scripts, which allowed for the automated input of test data and the simulation of user interactions. This method enabled the systematic and repeatable execution of various bot functions, allowing for comprehensive performance evaluation across different use cases and scenarios. The results were compiled into a detailed report that covered several critical aspects of the bot's performance:

1. The accuracy of the crime classification model (Table 7) was determined by comparing the predicted crime categories against the correct, expected outputs. An example is illustrated in Figure 8. In response to **RQ1**, the results show that the label accuracy of the model in classifying the correct crime types for the test set was 0.87.

Criteria	Average Accuracy
Crime Identification	0.8772
Culprit Identification	0.9412

Table 7: Test set results of more than 300 cases.

2. The bot's ability to correctly identify potential culprits was evaluated. The culprit identification accuracy was measured by comparing the names or entities identified by the bot against the user's actual culprit. We observe the culprit identification accuracy was 0.94. Moreover, with only 10% of cases involving an unknown culprit, it reinforces our point that a known perpetrator discourages filing complaints.



"They are recording the video chats and blackmailing for money in huge amount. If we deny to send money they are sending those to family friends and relatives."

"I am very stressed these days because my ex boyfriend keeps blackmailing me that he will post all our intimate photos online."

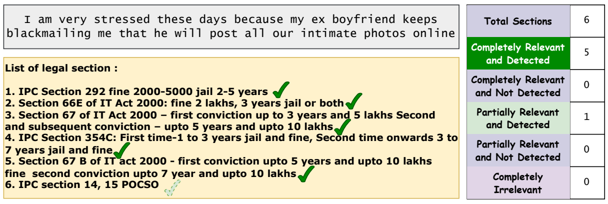
$CybercrimeLabelAccuracy = \frac{6+7}{16} = 0.81$

TP:CYBERCRIMES THAT WERE CORRECTLY PREDICTED
TN:CYBERCRIMESTHAT WERE CORRECTLY NOT PREDICTED
FN :CYBERCRIMES THAT WERE SUPPOSED TO BE PREDICTED BUT WERE NOT
FP:CYBERCRIMESTHAT WERE NOT SUPPOSED TO BE PREDICTED BUT WERE

Figure 8: Results of crime & culprit identification.

3. To assess the relevance and appropriateness of the legal advice and suggestions provided by the bot, a legal expert reviewed the sections of the law presented to the user in relation to the identified cybercrimes. When obtaining these automated results, we ask the legal experts to check if the

sections have been appropriately identified or not using the criteria mentioned in Figure 9. In response to **RQ2**, the results show that none of the test cases show ‘Completely irrelevant’ and ‘Completely relevant and not detected’ sections, depicting user complaints are mapped reliably to the correct Indian legal provisions.



```
I am very stressed these days because my ex boyfriend keeps
blackmailing me that he will post all our intimate photos online
```

List of legal section :

1. IPC Section 292 fine 2000-5000 jail 2-5 years ✓
2. Section 66E of IT Act 2000: fine 2 lakhs, 3 years jail or both ✓
3. Section 67 of IT Act 2000 – first conviction up to 3 years and 5 lakhs Second and subsequent conviction – upto 5 years and upto 10 lakhs ✓
4. IPC Section 354C: First time-1 to 3 years jail and fine, Second time onwards 3 to 7 years jail and fine ✓
5. Section 67 B of IT act 2000 - first conviction upto 5 years and upto 10 lakhs fine second conviction upto 7 year and upto 10 lakhs ✓
6. IPC section 14, 15 POCSO ✓

Total Sections	6
Completely Relevant and Detected	5
Completely Relevant and Not Detected	0
Partially Relevant and Detected	1
Partially Relevant and Not Detected	0
Completely Irrelevant	0

Figure 9: Results after evaluation of legal sections.

5.2 Human Evaluation (RQ3, RQ4, RQ5)

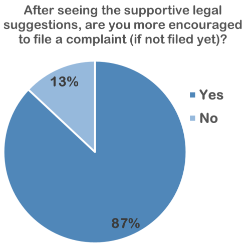


Figure 10: Analysis addressing **RQ3**.

The complete run of the entire app was conducted to analyze the usability from the perspective of the end user, ensuring that the system operates intuitively and meets user expectations. A human evaluation was performed as part of this process, involving a series of structured tests with real users. We asked 100 women and children (70 : 30) aged 13 – 50 to test our app and answer a questionnaire that would give us their feedback. Participants were recruited from our college and screened for basic fluency in Hindi/English and the ability to read Hinglish text. Participation was voluntary. All participants (or parents/guardians for minors) provided written informed consent (Table 12 and 10 in Appendix C) prior to participation; minors additionally provided assent (Table 11 in Appendix C). We anonymized all inputs: personal identifiers were removed prior to storage and analysis, and data were kept on encrypted storage accessible only to the study team. The goal was to gather direct feedback on the app’s functionality, ease of use, and overall user experience. Participants in the evaluation were asked to navigate through all phases of the app, from the initial sign-up process to the final phase, where suggestions are provided. They were given new complaints from the forum and asked to impersonate the cybercrime victim to test our app and answer a questionnaire (Table 13 in Appendix C) that would give us their feedback. After collecting detailed feedback, the results were analyzed to identify common pain points, usability challenges, and areas for improvement. Figure 11 depicts the obtained results. The insights gathered from this evaluation were instrumental in refining the app’s user interface, conversational flow, and overall design. The users seemed satisfied with the bot and its response time. Additionally, our suggestions regarding the platforms and detailed steps of filing complaints were well appreciated, thereby addressing **RQ4** and **RQ5**.

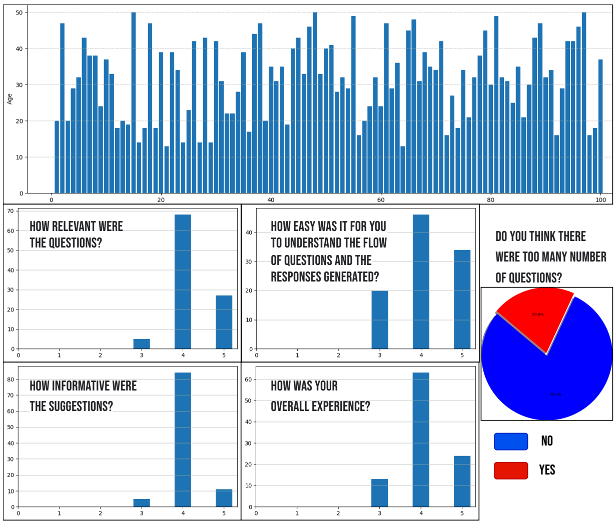


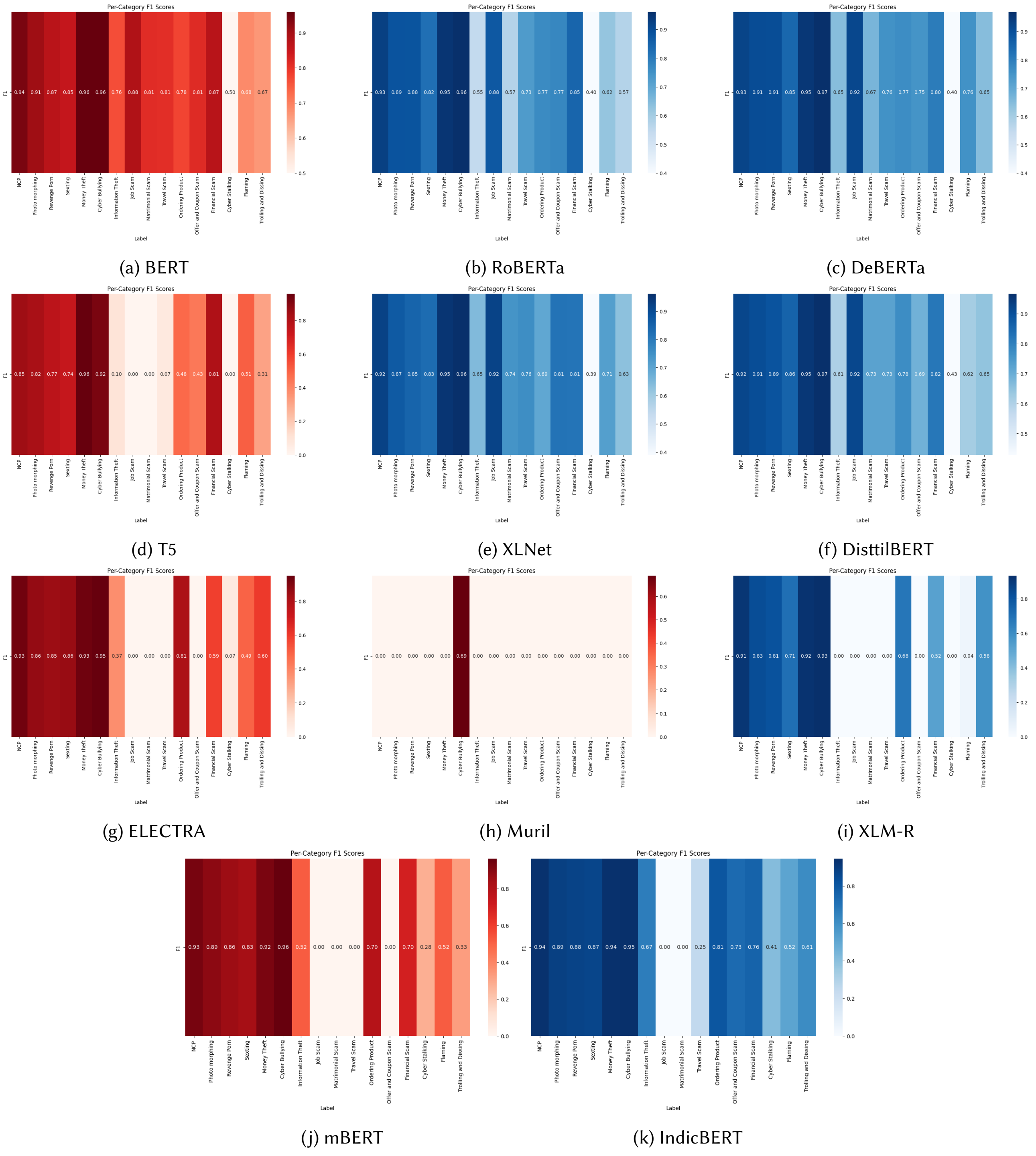
Figure 11: Results of Human Evaluation questionnaire.

Question	Corresponding Contributions
RQ1	Novel Hinglish cybercrime complaint dataset and transformer-based crime-type classification module; recision/recall-based classification performance.
RQ2	Legal-intent mapping module linking complaints to relevant provisions of the Indian IT Act and IPC; quantitative evaluation.
RQ3	Legal suggestion component and user study assessing users’ perceived encouragement to file formal complaints.
RQ4	Integrated mental health counselling module; user study evaluating perceived empathy and helpfulness of system responses.
RQ5	End-to-end evaluation of SAKHA through simulated complaint scenarios, including usability and perceived supportiveness measures.

Table 8: Mapping between research questions and paper contributions

6 Conclusion

This paper presented a chatbot-based tool designed to provide preliminary mental health counselling and legal aid to women and children affected by cybercrimes in India. The tool integrates a multi-label classification



(a) BERT

(b) RoBERTa

(c) DeBERTa

(d) T5

(e) XLNet

(f) DisttilBERT

(g) ELECTRA

(h) Muril

(i) XLM-R

(j) mBERT

(k) IndicBERT

Figure 12: Per-Category F1-Score Heatmaps for Multi-label Cybercrime Detection.

model trained on a custom dataset to detect cybercrime types, enabling personalized support and legal guidance. Additionally, it uses the DASS-21 questionnaire to assess mental health and offer tailored improvement suggestions. By combining psychological support with legal advice, the tool addresses a critical gap in current solutions, empowering victims with the knowledge and resources to navigate both emotional and legal challenges. Future work will focus on enhancing the model's accuracy and expanding its capabilities to support a broader range of cybercrimes.

Limitations

Despite promising results, our system exhibits a few recurring failure modes. In Table 9, we summarize the most salient categories (illustrative anonymized examples follow) and their likely cause, with a few more listed below.

1. The system does not attempt to verify the truthfulness of the information provided by users. Elements such as “who” are extracted for linguistic and contextual analysis only, not for legal attribution.
2. Currently, there is a lack of privacy techniques that could avoid identification of the user by the description of the case.
3. Our system does not validate or adjudicate the truth of a complaint, nor does it identify offenders in any authoritative sense.

Ethics Statement

The ethical approval for this study was obtained from the Department of Psychiatry, All India Institute of Medical Sciences (AIIMS), Patna, India (Approval No. AIIMS/Pat/Psy/2025/54). All participants were informed about the purpose and procedures of the study, and written informed consent was obtained from each participant prior to participation.

Acknowledgment

This work is supported through the project, Centre of Excellence in Cyber Crime Prevention against Women and Children- AI-based Tools for Women and Children Safety, from the Ministry of Home Affairs, Government of India. We also wish to thank Vaibhav Panday, Maheeth Reddy, Nishant Reddy Musukula, Kaushal Raj, and Utkarsh Singh for designing the GUI of the module.

References

Clark, Kevin, Minh-Thang Luong, Quoc V Le, and Christopher D Manning. 2020. Electra: Pre-training text encoders as discriminators rather than generators. *arXiv preprint arXiv:2003.10555*.

Conneau, Alexis, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Delhi, PIB. 2024. Increase in cyber crimes. https://pib.gov.in/PressReleaseIframePage.aspx?PRID=2003505.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding.

Frank, Oren and Roni Frank. 2012. Talkspace. https://www.talkspace.com/.

Gaiha, Shivani Mathur, Tatiana Taylor Salisbury, Mirja Koschorke, Usha Raman, and Mark Petticrew. 2020. Stigma associated with mental health problems among young people in india: a systematic review of magnitude, manifestations and recommendations. *BMC psychiatry*, 20(1):538.

Gogia, Vivek, Sanjay Mathur, Vinay Kumar Pandey, Divya Singh, M. Rajakumar, Suresh Chand Bohra, D.C. Pandey, Hemant Kr. Soni, M. Suresh Kumar, Jitender Singh Rawat, and Rajesh Kumar. 2024. Crime in india 2022. *National Crime Records Bureau, Government of India.*

Government, Indian. 2023. National cybercrime reporting portal. https://cybercrime.gov.in/.

He, Pengcheng, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2021. Deberta: Decoding-enhanced bert with disentangled attention.

Islam, Nusrat, Rokeya Siddiqua, and Sifat Momen. 2022. Machine learning techniques applied to bangla crime news classification. In *2022 IEEE 2nd Conference on Information Technology and Data Science (CITDS)*, pages 130–135.

Jacobs, Holly. 2012. Cyber civil rights initiative. https://cybercivilrights.org/.

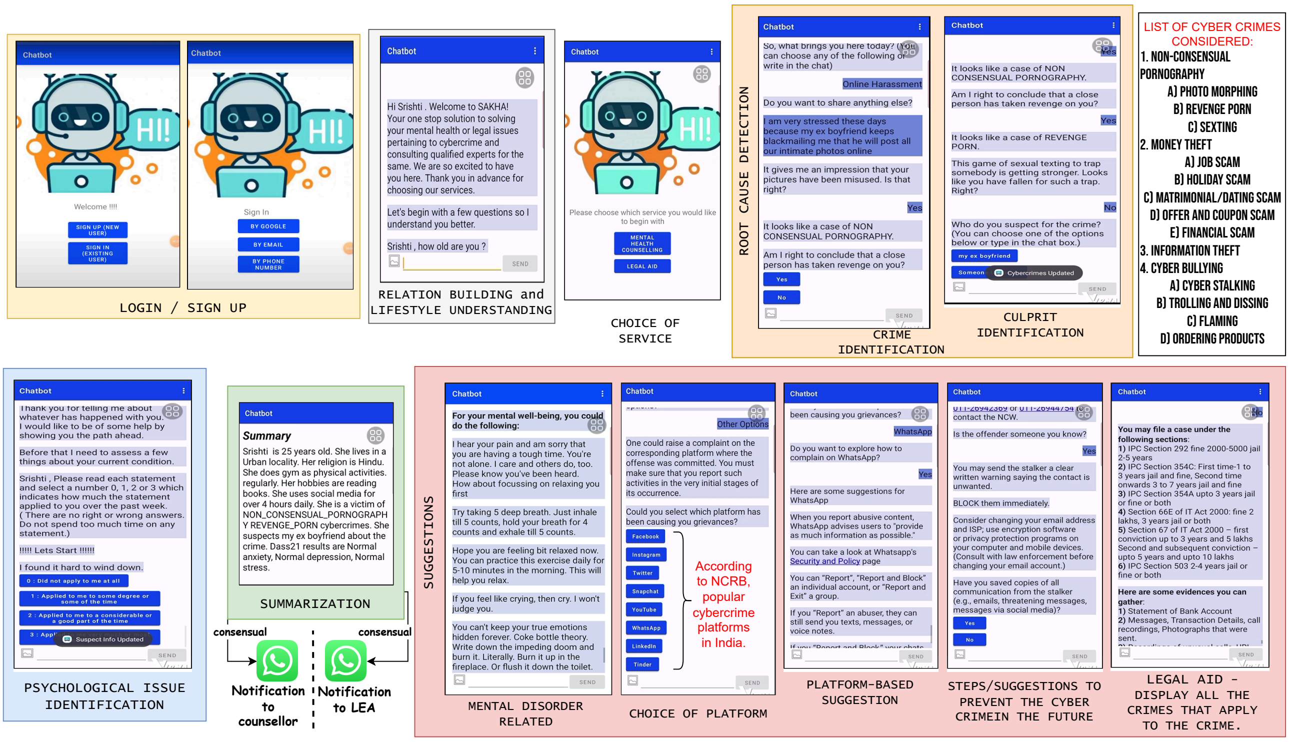
Jayakumar, Kirthi. 2013. The red elephant foundation. https://www.girlsnotbrides.org/our-partnership/member-directory/the-red-elephant-foundation/.

Judge, Kristin. 2018. Cybercrime support network. https://fightcybercrime.org/.

Kakwani, Divyanshu, Anoop Kunchukuttan, Satish Golla, Gokul N.C., Avik Bhattacharyya, Mitesh M. Khapra, and Pratyush Kumar. 2020. IndicNLPSuite: Monolingual Corpora, Evaluation Benchmarks and Pre-trained Multilingual Language Models for Indian Languages. In *Findings of EMNLP*.

User text	Gold	Predicted	Error type	Notes
I'm fine, don't worry (used to hide distress)	Neutral	Distress / Anxiety	LIWC false negative	Surface lexical signals misleading.
🔊: “she is blackmailing me” became “she is back mailing me”	Non-consensual pornography	Information theft	ASR transcription error → downstream misclassification	Numeric and named-entity corruption also common in ASR outputs.
Partial DASS-21 answers (user skipped items)	Moderate severity	Low severity	DASS scoring — partial responses	Missing items lead to under-estimation of severity.
“Wow, kya mast service hai iss so-called ‘loan app’ ki — they threatened to leak my contacts for fun! Must be their way of saying ‘thank you’ for timely repayments.”	Information Theft, Cyber Bullying	Money Theft, Financial Scam, Cyber Bullying	Hinglish sarcasm	Model is misled by Hinglish expressions like “kya mast service hai” and sarcastic praise.

Table 9: Recurrently occurring errors we noticed from SAKHA.



Hi Srishti . Welcome to SAKHA! Your one stop solution to solving your mental health or legal issues pertaining to cybercrime and consulting qualified experts for the same. We are so excited to have you here. Thank you in advance for choosing our services.

Let's begin with a few questions so I understand you better.

Srishti , how old are you ?

Chatbot
SEND
RELATION BUILDING and LIFESTYLE UNDERSTANDING
Chatbot
HI!
Please choose which service you would like to begin with
MENTAL HEALTH COUNSELLING
LEGAL AID
CHOICE OF SERVICE
ROOT CAUSE DETECTION
Chatbot

So, what brings you here today? (You can choose any of the following or write in the chat)

Online Harassment

Do you want to share anything else?

I am very stressed these days because my ex boyfriend keeps blackmailing me that he will post all our intimate photos online

It gives me an impression that your pictures have been misused. Is that right?

Yes

It looks like a case of NON CONSENSUAL PORNOGRAPHY.

Am I right to conclude that a close person has taken revenge on you?

Yes

It looks like a case of NON CONSENSUAL PORNOGRAPHY.

Am I right to conclude that a close person has taken revenge on you?

Yes

It looks like a case of REVENGE PORN.

This game of sexual texting to trap somebody is getting stronger. Looks like you have fallen for such a trap. Right?

No

Who do you suspect for the crime? (You can choose one of the options below or type in the chat box.)

LIST OF CYBER CRIMES CONSIDERED:

1. NON-CONSENSUAL PORNOGRAPHY
   A) PHOTO MORPHING
   B) REVENGE PORN
   C) SEXTING
2. MONEY THEFT
   A) JOB SCAM
   B) HOLIDAY SCAM
   C) MATRIMONIAL/DATING SCAM
   D) OFFER AND COUPON SCAM
   E) FINANCIAL SCAM
3. INFORMATION THEFT
4. CYBER BULLYING
   A) CYBER STALKING
   B) TROLLING AND DISSING
   C) FLAMING
   D) ORDERING PRODUCTS

Thank you for telling me about whatever has happened with you. I would like to be of some help by showing you the path ahead.

Before that I need to assess a few things about your current condition.

Srishti , Please read each statement and select a number 0, 1, 2 or 3 which indicates how much the statement applied to you over the past week. (There are no right or wrong answers. Do not spend too much time on any statement.)

!!!!! Lets Start !!!!!!

I found it hard to wind down.

0 : Did not apply to me at all
1 : Applied to me to some degree or some of the time
2 : Applied to me to a considerable or a good part of the time
3 : Appl
Suspect Info Updated
SEND
PSYCHOLOGICAL ISSUE IDENTIFICATION
Chatbot

Summary

Srishti is 25 years old. She lives in a Urban locality. Her religion is Hindu. She does gym as physical activities. regularly. Her hobbies are reading books. She uses social media for over 4 hours daily. She is a victim of NON_CONSENSUAL_PORNOGRAPHY REVENGE_PORN cybercrimes. She suspects my ex boyfriend about the crime. Dass21 results are Normal anxiety, Normal depression, Normal stress.

For your mental well-being, you could do the following:

I hear your pain and am sorry that you are having a tough time. You're not alone. I care and others do, too. Please know you've been heard. How about focussing on relaxing you first

Try taking 5 deep breath. Just inhale till 5 counts, hold your breath for 4 counts and exhale till 5 counts.

Hope you are feeling bit relaxed now. You can practice this exercise daily for 5-10 minutes in the morning. This will help you relax.

If you feel like crying, then cry. I won't judge you.

You can't keep your true emotions hidden forever. Coke bottle theory. Write down the impeding doom and burn it. Literally. Burn it up in the fireplace. Or flush it down the toilet.

Other Options

One could raise a complaint on the corresponding platform where the offense was committed. You must make sure that you report such activities in the very initial stages of its occurrence.

Could you select which platform has been causing you grievances?

been causing you grievances?

WhatsApp

Do you want to explore how to complain on WhatsApp?

Yes

Here are some suggestions for WhatsApp

When you report abusive content, WhatsApp advises users to "provide as much information as possible."

You can take a look at Whatsapp's Security and Policy page

You can “Report”, “Report and Block” an individual account, or “Report and Exit” a group.

If you “Report” an abuser, they can still send you texts, messages, or voice notes.

011-26942369 or 011-26944754 to contact the NCW.

Is the offender someone you know?

Yes

You may send the stalker a clear written warning saying the contact is unwanted.

BLOCK them immediately.

Consider changing your email address and ISP; use encryption software or privacy protection programs on your computer and mobile devices. (Consult with law enforcement before changing your email account.)

Have you saved copies of all communication from the stalker (e.g., emails, threatening messages, messages via social media)?

No

You may file a case under the following sections:
1) IPC Section 292 fine 2000-5000 jail 2-5 years
2) IPC Section 354C: First time-1 to 3 years jail and fine, Second time onwards 3 to 7 years jail and fine
3) IPC Section 354A upto 3 years jail or fine or both
4) Section 66E of IT Act 2000: fine 2 lakhs, 3 years jail or both
5) Section 67 of IT Act 2000 – first conviction up to 3 years and 5 lakhs Second and subsequent conviction – upto 5 years and upto 10 lakhs
6) IPC Section 503 2-4 years jail or fine or both

Here are some evidences you can gather:
1) Statement of Bank Account
2) Messages, Transaction Details, call recordings, Photographs that were sent.

Figure 13: A case study to depict the complete working of the bot.

Khanuja, Simran, Diksha Bansal, Sarvesh Mehtani, Savya Khosla, Atreyee Dey, Balaji Gopalan, Dilip Kumar Margam, Pooja Aggarwal, Rajiv Teja Nagipogu, Shachi Dave, Shruti Gupta, Subhash Chandra Bose Gali, Vish Subramanian, and Partha Talukdar. 2021. Muril: Multilingual representations for indian languages.

Kumar, Vineet. 2005. Cyberpeace foundation. https://www.cyberpeace.org/.

Le, Quoc and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 1188–1196, Bejing, China. PMLR.

Libovický, Jindřich, Rudolf Rosa, and Alexander Fraser. 2019. How language-neutral is multilingual bert?

Limited, TATA Consultancy Services. 2020. Data and Cybersecurity to Fight off Cyberattacks. https://www.tcs.com/what-we-do/services/cybersecurity.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach.

Lovibond, S H and P F Lovibond. 2011. Depression anxiety stress scales.

Lovibond, S. H. and P. H. Lovibond. 1995. Manual for the depression anxiety & stress scales. *Psychology Foundation of Australia*.

Matas, Alon and Danny Bragonier. 2013. Better help. https://www.betterhelp.com/.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space.

Moriarty, Glen. 2013. 7 cups. https://www.7cups.com/.

Pennebaker, James W, Roger J Booth, and Martha E Francis. 2007. Linguistic inquiry and word count: Liwc [computer software]. *Austin, TX: liwc. net*, 135.

Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

Pongpaichet, Siripen, Boonyapat Sukosit, Chitchaya Duangtanawat, Jiramed Jamjongdamrongkit, Chancheep Mahacharoensuk, Kantapong Matangkarat, Pattadon Singhajan, Thanapon Noraset, and Suppawong Tuarob. 2024. Camelon: A system for crime metadata extraction and spatiotemporal visualization from online news articles. *IEEE Access*, 12:22778–22802.

Prabhu, Anand V, MJ Jefiya, Joel Denny Joseph, Tisa Sunny, and Cerene Mariam Abraham. 2023. Cyber complaint automation system. In *2023 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA)*, pages 1–5.

Radford, Alec, Jong Wook Kim, Tao Xu, Greg Brockman, Christine Mcleavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 28492–28518. PMLR.

Raffel, Colin, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(1).

Rahma, F. and A. Romadhony. 2021. Rule-based crime information extraction on indonesian digital news. In *2021 International Conference on Data Science and Its Applications (ICoDSA)*, pages 10–15.

Sanh, Victor, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *ArXiv*, abs/1910.01108.

Sedik, Roy Rachman and ADE Romadhony. 2023. Information extraction from indonesian crime news with named entity recognition. In *2023 15th International Conference on Knowledge and Smart Technology (KST)*, pages 1–5.

Shree, S. A. Subha. 2025. Cyber crime against women in india cyber crimes: Types, patterns and prospects. *International Journal of Law Management and Humanities*, 8(2):2237–2245.

Sriram, Sujata, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2012. icall – initiating concern for all. https://www.girlsnotbrides.org/our-partnership/member-directory/the-red-elephant-foundation/.

Sukhai, Nataliya B. 2004. Hacking and cybercrime. In *Proceedings of the 1st Annual Conference on Information Security Curriculum Development*, InfoSecCD '04, page 128–132, New York, NY, USA. Association for Computing Machinery.

Vandrevala, Cyrus and Priya Hiranandani-Vandrevala. 2008. Vandrevala foundation. https://www.vandrevalafoundation.com/free-counseling.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Yang, Zhilin, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2020. Xlnet: Generalized autoregressive pretraining for language understanding.

A Cybercrimes Considered

According to the statistics released by NCRB, **Table 9A.10** and **Table 9A.11** in (Gogia et al., 2024) help us identify the most occurring crimes with women and children, respectively. Based on their data, we define the following classes of cybercrimes:

1. **Non-Consensual Pornography**: It includes all kinds of acts that involve filming, photographing, or distributing sexually explicit images or videos of someone without their consent.

   (a) ***Photo Morphing***: It refers to the use of digital manipulation techniques to alter or fabricate images, often with the intent to deceive or harm.

   > *"My photo is being misused by someone, and after posting my photo, some abusive words were written. It is making me and my family feel ashamed and disrespected in front of everyone"*

   (b) ***Revenge porn***: It describes the act of distributing sexually explicit images or videos of a person without their consent, typically with the intent to harm, humiliate, or seek revenge against that person. It usually occurs after an intimate or sexual relationship ends, and one party shares private material with others to seek retribution or cause distress to the victim.

> *"My ex named Nikhil Sharma* +91 − 7407286944 *(Whatsapp and Facebook ID) harassed me because he has my sex videos. I have too much audio from this person harassing me."*

(c) ***Sexting***: It can be stated as an act of coercion when one party pressures another into sending explicit material via messages. This can also extend to other forms of abuse when it involves minors, coercion, blackmail, or the distribution of images without the explicit, ongoing consent of the person depicted.

> *"I am getting calls and WhatsApp messages from mobile number* +91 − 7014627685 *asking for nude pictures and sexual content."*

2. **Money Theft**: The illegal act of stealing money using online methods. It can occur in various ways and is often facilitated by the use of technology, including computers, mobile devices, and the internet.

(a) ***Job scam***: Cybercriminals create fake job postings or companies that seem legitimate. They ask job seekers to pay upfront for application processing, training materials, or background checks, all under the guise of a "necessary step" in securing a job.

> *"I have got a message on WhatsApp regarding part time job for which i can earn* 1500+ *per day against which i have to perform some tasks. Initially the tasks were free but later on they asked me to deposit money to earn more commission which i did and eventually they said my account is frozen and need to pay more money to unfreeze it. I ended up paying Rs.*804000 *and lost them."*

(b) ***Holiday scam***: Fraudsters create fake travel deals that sound too good to be true. These scams promise highly discounted or luxurious holiday packages, sometimes with additional perks like flights, accommodations, and meals included. The victim is required to pay upfront to finalize the booking.

> *"I got a fraud promotional call from timestravelcare.com and asked me to purchase a holiday package. There will be assured gift that is dell laptop i5 and other item with this package for which i paid but they are not picking up my calls now."*

(c) ***Offer and coupon scam***: Cybercriminals trick individuals with fake promotional offers, coupons, and discount codes. The victim is asked to enter personal or financial details to redeem the offer or coupon, often ending up paying upfront for products that either don't exist or are of poor quality. In some cases, victims may unknowingly subscribe to recurring services or memberships, resulting in ongoing charges.

> *"I received a massage 'Dear sir/madam congratulations for Snapdeal company you are second lucky winner tata indigo car . Contact* 7321996284/6201706421*'. i call them they ask* 5600 *for this bank account first Mr.Laxmi Narayana Panda a/c-36369759723 IFSC-SBIN0006909 now they told me to transfer registration amount* 5600*. When i contact snapdeal then they told me 'Please do not respond to any Phone Call/Email/SMS claiming to offer rewards/lucky draw prizes on behalf of Snapdeal. We NEVER request our customers for unsolicited financial information or advance payments in exchange for rewards.' Now i have lost money."*

(d) ***Matrimonial/Dating scam***: Scammers create fake profiles on matrimonial/dating websites, often using stolen photos of attractive individuals. After developing a relationship with the victim, they invent a crisis or emergency and convince the victim to send money.

> *"The website called lifefriendship has taken my* [redacted] *in the name of friendship and dating and not picking the call and saying still they need money. The person incharge for this is Ms.Nisha and her number is* 9614636976.*"*

(e) ***Financial scam***: Cybercriminals often impersonate a government officials, to ask for payments for clearing different types of violations.

> *"I received a call on my telephone number* [redacted] *from a female Akansha Sharma from telephone number* 7042433301 *for re-*

> *newal of my driving license to which I agreed. I sent her all details of my license and paid amounts of Rs.1, 250, Rs.1, 250 and Rs.3, 000 to M B Holidayss for renewal of my license. Next she said that there are challans pending against my car which i had to pay."*

3. **Information Theft**: Cybercriminals steal personal information, such as credit card details, addresses, photos, and other personally identifiable information, to impersonate someone else.

   > *"Some one have misused my adhar card. Transfer parcel from Mumbai to Taiwan with my adhar card"*

4. **Cyberbullying**: It refers to the use of digital platforms to harass, intimidate, and harm others. It can involve the spreading of false rumors, threats, or name-calling.

   (a) ***Cyber Stalking***: It involves the use of digital communication tools to stalk and harass an individual. This can include repeated and unwanted monitoring, threatening, or contact that causes fear or distress.

   > *"Ms. Beauty Srivastava (Dy Manager, Star Union Di- Ichi,Mumbai Head Office) is threatening me, My son and my Paternal family to destroy us and to kill us through different means of social networking and mobilephone (Her No. is 9930950959). She is also threatening me over Whatsapp due to this we are living in very fearful position. She can any time attack on us to kill or to Damage us by any mean."*

   (b) ***Trolling and Dissing***: It refers to intentionally posting inflammatory, rude, or off-topic messages in online forums, social media, or comment sections to provoke emotional reactions in the victim. It involves belittling, mocking, or degrading others through negative comments.

   > *"I commented to support and respect humanity instead of judging anyone based on their deeds. It wasn't a direct comment on anyone. But this guy Sarthak has given me rape threats publicly."*

   (c) ***Flaming***: Flaming is similar to trolling but is usually more personal and one-on-one. It involves aggressive, hostile, or obscene messages meant to insult others. It often occurs in online forums, direct messages on various social platforms, etc.

   > *"A person named Nagarajan P. residing at Villivakkam, Chennai Tamil Nadu having contact number +91 − 8754499771 has created a new Facebook profile which is fake with no details had sent me friend requests which i have ignored. Later he sent me messages abusing me."*

   (d) ***Ordering Products***: This type of cybercrime involves using someone's identity or financial information to order products without consent or engaging in deceptive practices related to the sale of goods.

   > *"i placed the order on zepokart.com using instagram selling apple airpods at Rs.1999 , and i placed the order, i received a fake wired earphone.and i also tried to open that website , now its closed.they have intentionally done this with many people."*

B Sample Rule-Based Approach

Figure 14 illustrates a representative subset of the rule-based decision structure used in the legal aid module to generate suggestions. The figure is intended to provide a conceptual overview of how user responses guide the system toward appropriate next steps, rather than an exhaustive enumeration of all rules. The depicted flow demonstrates how high-level user intents, such as willingness to file a complaint, preferred mode of reporting (online vs. physical), and anonymity concerns, are translated into actionable guidance. Based on these branching decisions, the system either presents step-by-step instructions for filing a complaint through the relevant channel or provides supportive alternatives, such as information about anonymous reporting options or deferring formal action. The complete system includes additional rules not shown in the figure; however, the illustrated flow captures the core logic through which legal suggestions are surfaced to users in practice.

C Consent Templates

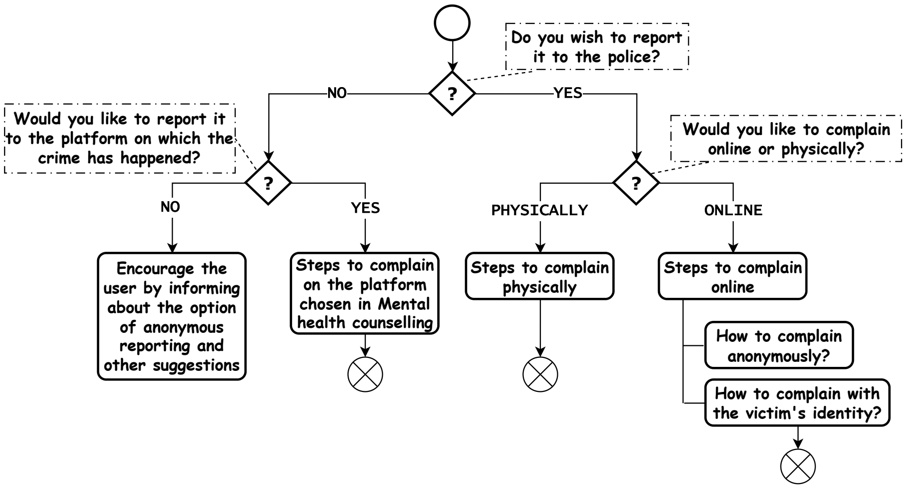

Do you wish to report it to the police?
?
NO
YES
Would you like to report it to the platform on which the crime has happened?
?
NO
YES
Would you like to complain online or physically?
?
PHYSICALLY
ONLINE
Encourage the user by informing about the option of anonymous reporting and other suggestions
Steps to complain on the platform chosen in Mental health counselling
Steps to complain physically
Steps to complain online
How to complain anonymously?
How to complain with the victim's identity?

Figure 14: Representative subset of the rule-based suggestion flow in the legal aid module, illustrating how user responses are mapped to actionable complaint guidance.

Title of Study: Evaluation of SAKHA — a chatbot for legal & mental health support to cybercrime victims

Principal Investigator: Dr. Sourav Kumar Dandapat, [Contact: sourav@iitp.ac.in]

Purpose: You are invited to take part in a study to evaluate a chatbot designed to provide initial legal and mental health guidance to victims of cybercrime. This study involves reading anonymized complaint reports, interacting with the chatbot, and completing a short questionnaire.

Procedures: If you agree to participate, you will:

1. read up anonymized forum complaints,
2. impersonate the victim described and interact with the chatbot for approximately [duration] minutes, and
3. complete a questionnaire about the chatbot's performance and your experience.

Risks and Discomforts: We anticipate minimal risk. Some scenarios may be upsetting as they describe cybercrime victims; you may skip or stop participating at any time without penalty.

Benefits: There is no direct benefit to you. Your participation will help to improve resources for cybercrime victims.

Confidentiality: All responses will be anonymized. Personal identifiers will not be stored. Data will be kept on encrypted servers accessible only to the research team.

Voluntary Participation: Your participation is voluntary. You can withdraw at any time without any penalty.

***Consent**: I have read and understood the information above, had the opportunity to ask questions, and voluntarily agree to participate.*

Name of Applicant ____ Date __________

Signature __________

Table 10: Adult Participant Informed Consent (for participants aged 18+)

I have been asked to help test a chatbot that gives information to people who report cybercrimes. I understand that:

- I will read short anonymized stories and pretend to be that person.
- I will talk to the chatbot and answer some questions.
- I can stop at any time and skip any question.

I agree to take part: ________________ (minor's signature).
Date: ________________

Table 11: Assent for participation (for persons under 18)

Title of Study: Evaluation of SAKHA — a chatbot for legal & mental health support to cybercrime victims.

Principal Investigator: Dr. Sourav Kumar Dandapat, [Contact: sourav@iitp.ac.in]

Purpose: We ask for your permission to allow your child to take part in a study that evaluates a chatbot's usability using anonymized text describing cybercrime incidents. The study is minimal risk and educational.

Procedures:

1. read anonymized complaint descriptions,
2. impersonate the described victim and interact with the chatbot for up to 20 minutes under supervision, and
3. answer a short questionnaire.

Risks: Content may be sensitive (describes cybercrime). Your child may skip any item or stop at any time.

Benefits and Compensation: No direct benefit. Your child will not receive anything for participation.

Confidentiality: Your child's responses will be anonymized. No identifying information will be included in publications.

Voluntary Participation: Participation is voluntary. You or your child may withdraw at any time without penalty.

Parental/Guardian Consent: *I have read and understood the information and give permission for my child named below to participate.*

Child's name: ________________
Parent/Guardian name: ________________
Parent/Guardian signature: ________________
Date: ________________
Parent/Guardian contact (optional): ________________

Table 12: Parental/Guardian Consent for Minor Participants (for participants aged < 18)

Study title: Evaluation of SAKHA — a chatbot for legal & mental health support
Date: _______________
Age: _______________

Instructions (to participant): You will be given an anonymized complaint (derived from public forum posts). Please read it carefully and impersonate the victim in conversation with the chatbot. Use the chatbot interface to ask questions you think a victim would ask. After interacting, please answer the questions below. You may stop at any time. If any content makes you uncomfortable, please skip it and notify the researcher.

	Not relevant at all	Slightly relevant	Relevant	Fairly relevant	Very relevant
How relevant were the questions?	○	○	○	○	○
	Too less	less	Just right	A bit too much	Too much
How informative were the suggestions?	○	○	○	○	○
	Very Bad	Bad	Okay	Good	Very Good
How was your overall experience?	○	○	○	○	○
	Very Abrupt	Abrupt	Mediocre	Smooth	Very Smooth
How easy was it for you to understand the flow of questions and the responses generated?	○	○	○	○	○
	Yes	No			
Do you think there were too many NUMBER OF QUESTIONS?	○	○			
	Yes	No			
After seeing the supportive legal suggestions, are you more encouraged to file a complaint (if not filed yet)?	○	○			

Do you have any other suggestions for us?

Table 13: Questionnaire.