Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Agentic AI Drafting of Multilingual Stalking Complaints and the Missing Duty to Preserve What Women Report to the Police

Dr. Jakey Khan

jakey.khan@slsh.edu.in

Symbiosis International University

Dr. Jayashri Rajbangshi

Gauhati University

Dr. Hamza Khan

Symbiosis International University

Research Article

Keywords: agentic artificial intelligence, police intake, technology-facilitated stalking, gender-based violence, linguistic access, comparative criminal procedure, AI auditing

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11768494/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Title Page

Title: Agentic AI Drafting of Multilingual Stalking Complaints and the Missing Duty to Preserve What Women Report to the Police

Corresponding Author:

Name: Dr. Jakey Khan, Assistant Professor of Law, Affiliation: Symbiosis Law School, Hyderabad under Symbiosis International (Deemed University), City: Hyderabad, State: Telangana, Country: India, Email ID: jakey.khan@slsh.edu.in, Orc. Id.: https://orcid.org/0000-0002-2778-6272

Author:

Name: Dr. Jayashri Rajbangshi, Assistant Professor of Law, Affiliation: J.B. Law College Guwahati under Gauhati University, City: Guwahati, State: Assam, Country: India, Email ID: jayashri1499@gmail.com, Orc. Id.: https://orcid.org/0009-0000-6433-1296

Author:

Name: Dr. Hamza Khan, Affiliation: Symbiosis Law School, Hyderabad under Symbiosis International (Deemed University), City: Hyderabad, State: Telangana, Country: India, Email ID: hamza.khan@slsh.edu.in, Orc. Id.: https://orcid.org/0009-0007-6403-7864

Declarations

Competing interests: The authors declare no relevant financial or non-financial competing interests.

Authors' contributions: Jakey Khan conceptualised the study, designed the comparative coding frame and the VANI audit protocol, performed the coding and analysis, and wrote the original draft. Jayashri Rajbangshi contributed to the comparative analysis of European, Inter-American and international human rights instruments and decisions, and reviewed and edited the manuscript. Hamza Khan contributed to the analysis of Indian criminal procedure and the literature review, and reviewed and edited the manuscript. All authors read and approved the final manuscript.

Data availability statement: The coding workbook with pinpoint citations for every score (ESM 1) and the VANI protocol package, comprising the codebook, development fixtures, versioned prompts, data dictionary, run and annotation templates and scoring script (ESM 2), are provided as Supplementary Information. The synthetic fixtures are proposed for release under CC0 1.0. This manuscript reports no experimental outputs and no data from human participants.

Funding: The current research has not received any funding support from governmental, private or third-sector agencies.

Ethical approval: Not applicable. This study analyses publicly available legal texts and judicial decisions and uses only synthetic scenarios; no human participants, personal data or police records were involved.

Clinical trial number: Not applicable.

Consent to participate: Not applicable.

Consent to publish: Not applicable.

Agentic AI Drafting of Multilingual Stalking Complaints and the Missing Duty to Preserve What Women Report to the Police

Abstract

During a police trial in Tasmania, an AI drafting tool recorded that a woman alleged earlier threats against her life; the recording showed she had said there were none. Such reversals matter most when women report technology-facilitated intimate-partner stalking, because protection depends on whether the record carries the pattern, attribution and uncertainty of the account. Hyderabad Police now convert spoken complaints in ten languages into draft First Information Reports, yet no public evidence shows what survives that conversion. This article asks which safeguards govern the passage from first disclosure to officer-reviewed record when agentic AI performs transcription, translation, extraction and drafting. Structured comparative coding of eleven legal and governance instruments from India, the European Union, the Council of Europe, California, Utah, the CEDAW Committee and NIST, across ten safeguard dimensions, finds that none expressly requires testing whether a draft preserves the complainant's account. Instruments securing complainant participation (language access, read-back, copies) and instruments regulating AI authorship (disclosure, certification, draft retention) form separate clusters with no overlap. Coding ten judgments and treaty-body decisions identifies eight factual dimensions treated as material, including course of conduct, threats to third persons and uncertainty about identity. The article converts these findings into a standard of reviewable fidelity and into VANI, a pre-specified audit protocol for matched English, Telugu and Telugu–English accounts that compares direct drafting, source-linked drafting and an agentic verification pass. Certification without source access assigns responsibility for a record without enabling anyone to detect what it lost.

Keywords agentic artificial intelligence · police intake · technology-facilitated stalking · gender-based violence · linguistic access · comparative criminal procedure · AI auditing

1. Introduction

During a 2025 trial of an AI report-writing tool by Tasmania Police, the system produced a family-violence narrative stating that a woman had alleged that a man with a concealed firearm had previously threatened her life. When an officer checked the body-worn camera recording, the woman had in fact said something close to the opposite: that there had been no threats. The officer attributed the reversal to overlapping speech [1]. The same evaluation reported that the tool roughly halved writing time, that it sometimes surfaced coercive behaviour officers had missed, and that its performance declined when partners, children or other witnesses were present, where it occasionally misidentified roles and omitted key details [1, 2]. The episode compresses the problem this article addresses. A fluent institutional record can reverse a denial, assign a statement to the wrong speaker, or convert a suspicion into an assertion, and each of these changes alters what an officer will investigate, what a supervisor will see, and what a court may later be asked to rely on.

The stakes are sharpest when the account concerns technology-facilitated intimate-partner stalking. Such complaints are typically cumulative rather than episodic. A woman may report that she asked a former partner to stop contacting her, that messages continued on later days, that she saw unfamiliar activity on an account and suspects him, and that a relative heard an earlier threat. The legal significance lies less in any one sentence than in the relationships among them: the refusal followed by renewed contact, the link between online and offline conduct, the difference between what she saw and what she infers [3–5]. A drafting system can preserve every individual sentence and still erase that structure.

Reported deployments make this an institutional question rather than a speculative one. In May 2026 Hyderabad City Police launched AI-CopWriter, a mobile application reported to record, transcribe and translate complaints in ten Indian languages in real time and to convert spoken complaints into First Information Report (FIR) drafts within seconds [6]. Trade reporting attributes the system to a private supplier working with the police IT cell [7]. Separately, the National Crime Records Bureau describes FIR-field extraction and summarisation functions within the Crime and Criminal Tracking Network and Systems (CCTNS 2.0) [8]. An Indian engineering study reports a smart FIR-drafting system tested on 2,500 FIR records and 500 multilingual audio complaints, with metrics for legal-section prediction accuracy and drafting time [9]. None of these public sources reports how often a complainant's allegation is omitted, reversed or reattributed between the spoken account and the draft, or whether any such rate differs between English, Telugu and the code-switched Telugu–English that many Hyderabad complainants actually speak.

These systems are increasingly agentic. A single request can trigger a sequence of model-driven steps (speech recognition, translation, fact extraction, narrative drafting and, in more elaborate designs, self-verification) that runs with little human intervention between steps and whose output is consequential for a third party [10]. That is precisely the configuration in which the collection's concern with structural gender inequality becomes concrete: if the pipeline performs worse on women's code-switched speech, on cumulative narratives, or on allegations involving several actors, the inequality is encoded at the first point at which the state hears the complaint [11].

This article asks three questions. First, what do the legal and governance instruments that regulate police intake and AI-assisted police records actually require of the passage from first disclosure to official record? Second, which facts in a stalking or

intimate-partner complaint do courts and treaty bodies treat as material to an adequate official response? Third, what review standard and audit design would make the preservation of those facts verifiable before an officer adopts a record?

The article makes four contributions. It reports a structured comparative coding of eleven instruments across ten safeguard dimensions, showing that none expressly requires a substantive fidelity test and that participation safeguards and AI-authorship safeguards occupy separate, non-overlapping clusters. It derives, from ten judgments and treaty-body decisions, an eight-dimension taxonomy of material facts that a draft must carry. It translates both findings into a standard of reviewable fidelity with a common factual layer and jurisdiction-specific overlays. And it specifies VANI (Verifiable Account-to-Narrative Integrity), a pre-specified multilingual audit protocol, with development fixtures, codebook, prompts and scoring code released for replication. Section 2 positions the study against the closest literature. Section 3 describes the comparative coding method. Section 4 reports results. Section 5 develops the review standard and Section 6 the VANI protocol. Section 7 discusses implications for agentic AI governance and gender equity, and Section 8 concludes.

2. Background and related work

2.1 The first report as a gendered gate

Research on technology-facilitated abuse has shown how intimate-partner abusers exploit ordinary devices and accounts, often through access enabled by a shared history rather than technical sophistication [3], and how platforms and institutions struggle to recognise digital conduct as part of a pattern of coercive control [4, 5]. These studies converge on a point that matters for drafting: the harm is constituted by

accumulation, persistence and the link between online and offline conduct, so a record that atomises the account into unrelated incidents misdescribes it.

In India the first report is also a site of institutional gatekeeping. Under the Bharatiya Nagarik Suraksha Sanhita, 2023 (BNSS), information about a cognizable offence given orally must be reduced to writing, read over to the informant and signed, and a copy must be given free of cost to the informant or victim [12]. Where a woman reports specified offences, including stalking under section 78 of the Bharatiya Nyaya Sanhita, 2023 (BNS), the information must be recorded by a woman police officer or any woman officer [12, 13]. Under the predecessor Code, the Constitution Bench in *Lalita Kumari* held registration mandatory where information discloses a cognizable offence, while recognising categories, including matrimonial and family disputes, in which a preliminary inquiry could be appropriate [14]. BNSS section 173(3) now allows a preliminary enquiry, with the permission of an officer not below the rank of Deputy Superintendent, for cognizable offences punishable with three years or more but less than seven years [12]. Stalking on a first conviction is punishable with imprisonment of up to three years [13], so whether a draft presents a complaint as repeated, unwanted conduct after a clear indication of disinterest, or as a family quarrel, may influence whether the information is registered at once or enquired into first.

Empirical work confirms that institutional design at this gate changes outcomes for women. A randomised evaluation across 180 police stations found that women's help desks increased the registration of gender-based violence cases, particularly where female officers staffed them [15]. Research on all-women police stations similarly treats the point of reporting as decisive for women's access to justice [16]. An AI intake tool is another design choice at the same gate. Its effect on registration, and on what

is registered, is an empirical question that the existing deployments have not answered.

2.2 Evidence on AI-assisted police reporting

The empirical record on AI-drafted police reports remains thin and predominantly North American. A field experiment found that an AI report-writing tool did not significantly reduce writing time [17], while the Tasmanian evaluation reported the reverse on speed and documented role confusion and omissions in complex family-violence encounters [1, 2]. Ferguson argues that AI-generated narratives can manufacture "generative suspicion", shaping downstream discretion and accountability because police narratives are relied upon long after they are written [18]. A university project independently simulated a proprietary workflow and catalogued hallucinated content across repeated runs [19].

The closest methodological work is Aiersilan and Wuernikebai, who propose a six-axis framework (FACETS-PR) for AI-drafted reports, including source fidelity, asymmetric recall of exculpatory content, and a paired-dialect probe contrasting Standard American English with African American English renderings of the same content [20]. Run on four small open-weight models over synthetic body-worn camera transcripts, their instrument found low retention of quoted evidentiary utterances, an inculpatory–exculpatory recall asymmetry in some models, and, against prediction, a tendency of drafts to hedge rather than harden the transcript's epistemic stance. They also observe that California's statute compels disclosure, certification and retention but says nothing about the substantive quality of the draft [20]. Indian technical work has, by contrast, emphasised section prediction and efficiency [9], and policy analysis has advocated

multilingual speech-to-text with read-back for smart FIRs without testing whether read-back of a generated draft improves fidelity [21].

2.3 Multilingual and code-switched language technology

Indian-language translation and speech infrastructure has improved substantially [22, 23], and multilingual summarisation benchmarks exist [24]. Yet multilingual large language models remain markedly weaker on code-switched input than their monolingual performance suggests [25]. Machine translation literature in legal and medical settings documents that errors in negation, modality and pronoun reference are both common and consequential, and that such errors tend to be borne by people with the least ability to detect them [26]. These are exactly the features (denial, certainty, actor) on which an intimate-partner stalking complaint turns. Atomic-fact evaluation offers a measurement template, scoring generated text against a reference by decomposing it into minimal propositions [27], but such methods score precision (whether output claims are supported) rather than the recall of material facts on which omissions depend.

2.4 Agentic AI governance and the limits of human oversight

Governance work in this collection proposes lifecycle governance, equity-by-design and human-in-the-loop principles for agentic systems, while noting that the evidence base is geographically narrow and that high-level principles are rarely translated into implementable mechanisms [11]. Survey evidence from the collection also links the legal and policy environment to lower cyber-violence [28], reinforcing the importance of institutional response. A wider literature cautions that human oversight requirements can provide a false sense of security, because people are often unable to perform the oversight functions such policies assume [29], and that automation bias leads

reviewers to accept system output, including errors of omission [30]. Applied to police drafting, the concern is that a signature certifies a review that the reviewer may lack the information to conduct.

2.5 Research gap

Table 1 positions this study against the closest work. Existing studies either measure AI-drafted report quality in a monolingual, officer-observed setting [20], evaluate efficiency and section prediction in Indian FIR systems [9], analyse downstream doctrinal harm [18], or propose agentic governance principles at a general level [11]. No study identified in the search connects the legal duties governing a complainant's first disclosure, across jurisdictions, to a measurable standard for what an AI-drafted record must preserve, and none designs a matched multilingual test for first-person complaints of technology-facilitated stalking in a Global South deployment context.

Table 1 Positioning of this study against the closest existing work

Study	Input and setting	Languages	What is assessed	Contribution of this study
Aiersilan and Wuernikebai [20]	Synthetic body-worn camera transcripts; officer-observed scenes (US)	Standard American English and African American English	Fidelity, exculpatory recall, suspicion inflation; four open-weight models	Complainant's first-person disclosure; cross-language and code-switched variants; tests a verification step; comparative legal derivation of material facts
Mathivanan et al. [9]	FIR records and multilingual audio complaints (India)	Multilingual (not specified in abstract)	Legal-section prediction accuracy, drafting time	Measures preservation of allegations rather than section labels; language-gap analysis
Ferguson [18]	Doctrinal analysis of AI police narratives (US)	Not applicable	Downstream suspicion, discretion and accountability	Moves the analysis upstream to intake and makes it measurable
Tasmania Police evaluation [1, 2]	Body-worn camera audio, family-violence incidents (Australia)	English	Time saved, errors and recovered risk details	Controlled, replicable test with released fixtures and codebook
Jain and Gupta [11]	Systematic review and bibliometric analysis	Not applicable	Lifecycle and equity-by-design governance principles for agentic AI	Operationalises human oversight and bias auditing for a specific agentic public-service pipeline

3. Methods

3.1 Design

The study uses functional comparative institutional analysis implemented through structured qualitative coding of legal texts and decisions [31]. Functional comparison selects authorities by the role they play at a given procedural point rather than by formal equivalence. An Indian FIR, an EU written acknowledgement of a complaint, a Tasmanian family-violence report and a United States officer narrative have different legal effects; the comparison asks what each system requires of the passage from a person's account to the official text, and what it does not.

3.2 Selection of instruments and decisions

Instruments were eligible if they (a) govern the recording of a complaint or report of crime, (b) regulate the generation, disclosure or retention of AI-assisted law-enforcement records, or (c) set data-quality or risk-assessment duties that apply to police records of gender-based violence. Eleven instruments met these criteria across six normative orders: India (BNSS section 173 and Bharatiya Sakshya Adhiniyam (BSA) section 63) [12, 32]; the European Union (Directives 2012/29/EU, 2016/680 and 2024/1385 and Regulation 2024/1689) [33–36]; the Council of Europe (Istanbul Convention) [37]; the United States (California Penal Code section 13663 and Utah Code sections 53-25-601 to 602) [38, 39]; the United Nations CEDAW Committee (General Recommendations 33 and 35, coded together as treaty-body guidance) [40, 41]; and voluntary technical guidance (NIST AI 600-1) [42]. Legal force differs across these sources and is recorded as a coding attribute rather than collapsed.

Ten decisions were selected for the materiality analysis because each addresses the adequacy of an official response to domestic, gender-based or digitally mediated

violence, or the integrity of a statement as it passes through an institutional process: four judgments of the European Court of Human Rights on domestic violence and cyberviolence [43–46], one further cyberviolence judgment [47], a merits report of the Inter-American Commission [48], a judgment of the Inter-American Court [49], views of the CEDAW Committee [50], and two common-law authorities concerning interpretation and the circumstances in which a statement is made [51, 52]. *Bater-James and Mohammed* [53] was reviewed for privacy and proportionality constraints on access to complainant devices and is discussed in Section 5 rather than coded, because it addresses access to evidence rather than the content of a record.

3.3 Coding frame

Each instrument was coded against ten safeguard dimensions grouped in three clusters. Complainant participation comprises language access at the point of complaint (LA), read-back or complainant verification of the record's content (RB), and a copy or acknowledgement to the complainant (CP). Record integrity comprises attribution, meaning a distinction between facts and assessments or between categories of persons (AT); consideration of history, pattern and risk context (HX); and correction or rectification (CR). AI authorship and traceability comprises disclosure of AI involvement (DI), named reviewer certification (RC), and retention of drafts, versions or logs (TR). The tenth dimension, substantive fidelity measurement (FM), asks whether the instrument requires the accuracy of the generated or recorded text to be tested against its source. Each cell was scored 2 (express requirement), 1 (partial, conditional or indirect) or 0 (not addressed), with a pinpoint provision recorded for every non-zero score.

Each decision was coded for whether eight factual dimensions were central to its reasoning about the adequacy of the official response: course of conduct or repetition (PC); chronology, including the timing of official action (CH); links between online and offline conduct (XL); attribution and identity uncertainty (ID); threats, including threats to third persons (TH); prior reports, withdrawals and history (PR); stereotyping or minimisation of the account (ST); and language or interpretation quality (LQ). A dimension was coded as central only where the decision's reasoning turned on it, not where it merely appeared in the facts.

3.4 Transparency and reliability

Coding was performed by the first author on the official text of each instrument and decision, accessed in November 2026. Targeted searches in Scopus, Google Scholar, HeinOnline, JSTOR and Factiva on 16 November 2026 informed the positioning analysis; the search log and screening decisions are reported in the supplementary materials. Because coding was conducted by a single coder, every score is reported with its pinpoint in Supplementary Information (ESM 1) so that any reader can re-code it. Conditional obligations, notably those of the EU AI Act, which depend on a system's intended purpose and the Act's application timetable, were scored as partial rather than express. The authors used AI assistant like Scopus AI, Manupatra AI, ResearchRabbit and Consensus for literature retrieval support and organizing the same; all legal sources, codes and citations were checked by the authors, who take responsibility for the content.

4. Results

4.1 No instrument requires substantive fidelity testing

Figure 2 and Table 2 report the coding. None of the eleven instruments expressly requires that the accuracy of a drafted or recorded complaint be tested against the complainant's account. Three address the matter partially. The Law Enforcement Directive imposes general duties of accuracy and requires competent authorities to take reasonable steps to verify data quality before transmission [34]. The AI Act requires appropriate accuracy for high-risk systems, but whether a drafting tool is high-risk depends on its intended purpose, and Article 6(3) offers a qualified route out of the high-risk category for some Annex III systems [36]. NIST AI 600-1 recommends pre-deployment testing for confabulation but is voluntary [42].

The instruments written specifically for AI-assisted police records are the clearest illustration. California requires a disclosure on each report, an officer signature verifying review, retention of the first AI draft for as long as the final report, and an audit trail identifying the user and the source audio or video [38]. Utah requires an agency policy, a report-level disclaimer, and author certification that the report has been reviewed for accuracy [39]. Both create a record of who produced and approved a report. Neither specifies what the reviewer must be able to see, nor requires any measurement of how often drafts omit, reverse or misattribute what was said.

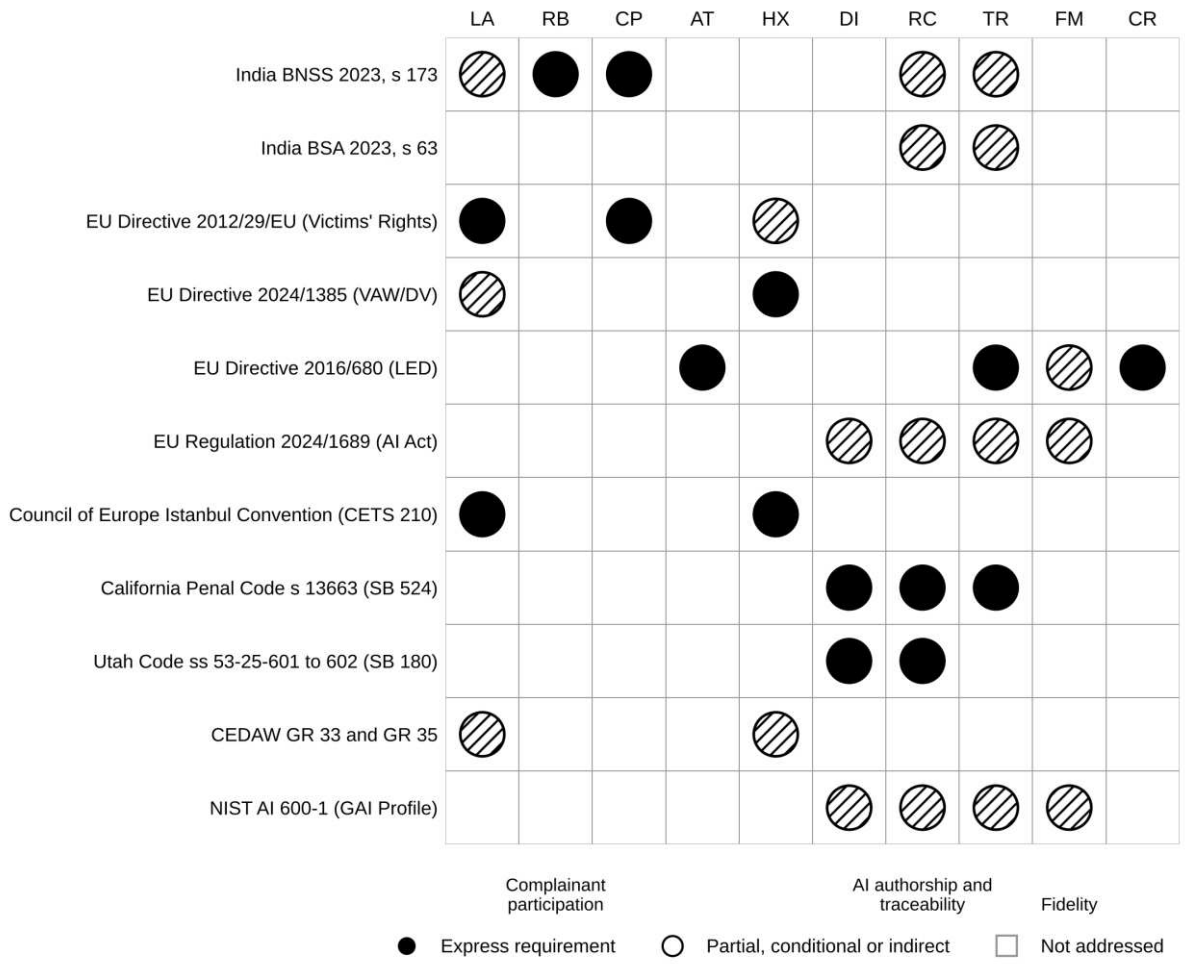

Fig. 2 Coding of eleven legal and governance instruments against ten safeguard dimensions. Filled circle, express requirement; hatched circle, partial, conditional or indirect; empty cell, not addressed. LA language access; RB read-back; CP copy or acknowledgement; AT attribution; HX history and risk context; DI disclosure of AI use; RC reviewer certification; TR retention and traceability; FM substantive fidelity measurement; CR correction

Table 2 Coding of eleven instruments against ten safeguard dimensions

Safeguard dimension	Express	Partial	Instruments with an express requirement
Language access at complaint (LA)	2	3	EU Directive 2012/29/EU; Council of Europe Istanbul Convention
Read-back / complainant verification (RB)	1	0	India BNSS 2023, s 173
Copy or acknowledgement to complainant (CP)	2	0	India BNSS 2023, s 173; EU Directive 2012/29/EU
Attribution: facts vs assessments, categories of persons (AT)	1	0	EU Directive 2016/680
History, pattern and risk context (HX)	2	2	EU Directive 2024/1385; Council of Europe Istanbul Convention
Disclosure of AI involvement (DI)	2	2	California Penal Code s 13663; Utah Code ss 53-25-601 to 602
Named reviewer certification (RC)	2	4	California Penal Code s 13663; Utah Code ss 53-25-601 to 602

Safeguard dimension	Express	Partial	Instruments with an express requirement
Retention of drafts, versions or logs (TR)	2	4	EU Directive 2016/680; California Penal Code s 13663
Substantive fidelity testing against source (FM)	0	3	None
Correction or rectification (CR)	1	0	EU Directive 2016/680

Express: the instrument states the requirement. Partial: the requirement is conditional (for example on high-risk classification), indirect, limited to a sub-class of persons, or voluntary. Pinpoints for every score are given in ESM 1.

4.2 Participation and AI-authorship safeguards do not overlap

Across the 110 instrument-dimension cells, 15 were coded as express requirements, 18 as partial and 77 as not addressed. No instrument addressed more than five of the ten dimensions in any form; the highest, BNSS section 173, does so through participation safeguards written for a human recorder. The second finding is structural. Safeguards that secure the complainant's participation in her own record, namely language access, read-back and a copy, appear in five instruments (BNSS, the Victims' Rights Directive, the 2024 Directive, the Istanbul Convention and the CEDAW recommendations). Express AI-disclosure duties appear in two (California and Utah), and conditional or voluntary disclosure in two more (the AI Act and NIST). No instrument contains both an express participation safeguard and an express AI-disclosure safeguard. The bodies of law that regulate AI drafting are written around the officer as author and the defendant as the person later affected; they do not address the complainant's language or her opportunity to correct the record. The bodies of law that secure the complainant's language and participation were written before AI drafting and say nothing about a generated intermediary.

India illustrates the consequence. BNSS section 173 contains the strongest participation safeguard in the set: oral information must be read over to the informant, signed, and copied to her [12]. It is the only instrument with an express read-back duty. But the provision presumes a human recorder and a record that the informant can

understand. Where a tool generates an English draft from Telugu–English speech, read-back requires a back-translation that is itself machine-mediated, and neither the statute nor any public operating procedure for AI-CopWriter identified in this study specifies how errors introduced at that step are to be detected [6, 7]. The BSA's certification regime for electronic records addresses the integrity of devices and processes when a record is tendered in evidence [32]; it does not address whether a derived narrative faithfully represents the oral account from which it was produced.

4.3 What the jurisprudence treats as material

Table 3 reports the materiality coding. Course of conduct was central in seven of the ten decisions, chronology and threats in six each, and prior reports and stereotyping in five each. The European domestic-violence cases tie the adequacy of the state's response to the history known to authorities, including earlier complaints and withdrawals [43, 44], and require that risk assessment be autonomous, proactive and comprehensive while leaving the assessment itself to the competent authorities [45]. The Inter-American and CEDAW decisions connect institutional tolerance, delay and stereotyped responses with state responsibility [48–50].

Two dimensions appear only in the most recent cases. Links between online and offline conduct, and attribution under uncertainty, were central in *Buturugă* and *Volodina (No. 2)*, which require authorities to examine cyberviolence in the context of the domestic violence with which it is connected and address conduct such as impersonation, tracking and the misuse of images [46, 47]. Language and interpretation quality appears only in *Tran*, a Canadian case on interpretation for an accused [51], and the identification of who said what, in what circumstances, is central to the United States confrontation cases [52].

Table 3 Factual dimensions central to the reasoning of ten decisions on official responses to gender-based and digitally mediated violence

Decision	PC	CH	XL	ID	TH	PR	ST	LQ
Opuz v Turkey (ECtHR 2009)	●	●			●	●	●	
Talpis v Italy (ECtHR 2017)	●	●			●	●	●	
Kurt v Austria [GC] (ECtHR 2021)	●				●	●		
Buturugă v Romania (ECtHR 2020)	●		●	●	●			
Volodina v Russia (No 2) (ECtHR 2021)	●	●	●	●	●			
Maria da Penha v Brazil (IACHR 2001)	●	●				●	●	
González et al. (Cotton Field) v Mexico (IACtHR 2009)		●					●	
González Carreño v Spain (CEDAW 2014)	●				●	●	●	
R v Tran (SCC 1994)								●
Davis v Washington; Hammon v Indiana (USSC 2006)		●		●				
Decisions in which central (n = 10)	7	6	2	3	6	5	5	1

PC course of conduct or repetition; CH chronology including timing of official response; XL online–offline link; ID attribution and identity uncertainty; TH threats including to third persons; PR prior reports, withdrawals and history; ST stereotyping or minimisation of the account; LQ language or interpretation quality. ● = central to the reasoning. Sources: [43–52]

4.4 The gaps align

Read together, the two analyses show that the factual dimensions most exposed to AI drafting error are the least protected. The Tasmanian evaluation and the FACETS-PR study identify role confusion, omission and changes in epistemic stance as characteristic failures [1, 20]; the code-switching literature predicts heightened risk for negation and reference in mixed-language input [25, 26]. These map onto attribution and identity uncertainty (ID), cross-channel links (XL) and denials within the threat dimension (TH). Yet ID and XL are the dimensions that have entered jurisprudence most recently, and no coded instrument other than the Law Enforcement Directive expressly requires a record to distinguish fact from assessment. The safeguards that

do exist attach at the ends of the pathway (language access at the start, signature and retention at the end) and leave the transformations between them unexamined (Fig. 1).

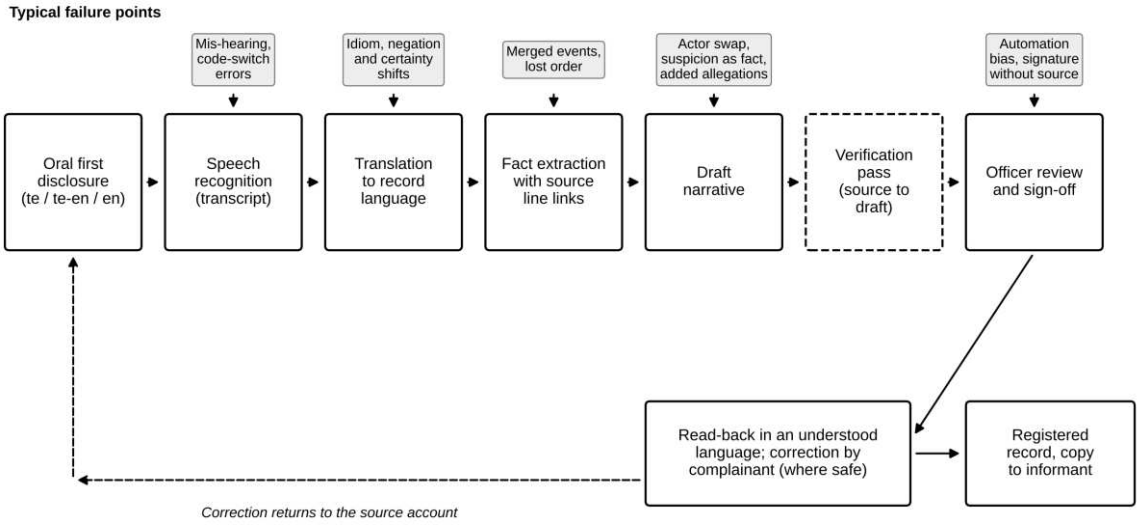

Fig. 1 Source-to-record pathway for agentic AI drafting of a multilingual complaint, showing typical failure points (shaded labels above each step) and the read-back and correction loop. The dashed verification box marks the optional agentic verification stage tested in VANI Condition C. Safeguards found in the coded instruments attach at the beginning (language access) and end (sign-off, read-back, retention) of the pathway; none requires measuring what happens to the account between them

4.5 A field check: the Tasmanian evaluation

The coding frame can be checked against the one field evaluation of AI drafting in family-violence matters for which official material is public. Each failure reported from the Tasmanian trial maps onto a dimension in Table 3. Confusion of names and roles concerns attribution (ID). Missing a threat, assault or risk factor concerns threats and course of conduct (TH, PC). The recorded allegation of prior threats, where the woman had said there were none, is a reversed denial within the threat dimension, the error type that VANI classifies as critical [1]. Conversely, the reported capacity of the tool to surface coercive behaviour that officers had missed concerns history and pattern (PR,

PC), showing that drafting assistance can improve as well as degrade materiality on the same dimensions [1, 2].

The case also shows which safeguard actually detected the most serious error. The evaluation describes mandatory human verification and supervisor oversight as the basis of accuracy [1]. Yet the reversed denial was identified when an officer returned to the recording, not when a draft was signed. In the terms of the coding frame, detection depended on retention and access to the source (TR) used for a comparison of source and draft (FM), not on certification (RC). It is precisely that comparison which no coded instrument requires. The evaluation also concluded that no legislative change was needed for the tool to be used [1], which is consistent with the finding in Section 4.1 that existing law, even where it addresses AI drafting, is silent on fidelity.

4.6 Applying the frame to the Hyderabad deployment

Applying the same ten dimensions to the publicly reported features of AI-CopWriter shows how little of the relevant design is visible. Language access (LA) is the deployment's central promise: launch reporting describes real-time transcription and translation in ten languages and reduced dependence on interpreters [6]. Reviewer involvement (RC) is implied by reporting that officers finalise drafts [6, 7]. For every other dimension (read-back of a generated draft in the complainant's language, disclosure of AI use on the record, retention of the first draft and intermediate translations, attribution conventions, correction procedure and any fidelity testing by language), no public information was found. BNSS section 173 supplies read-back, signature, copy and a woman officer for stalking complaints by women [12], so the statutory frame is stronger than in most coded jurisdictions on participation. What is missing is any public mechanism connecting that participation to the generated text.

A public-records request for the operating procedure, supported-language list, version history and dated aggregate correction figures by language would allow the frame to be applied in full; no such request had been made at the time of writing.

5. From findings to a standard: reviewable fidelity

5.1 Two directions of review

The results support a principle of reviewable fidelity. A legally material proposition in the draft should be traceable to an attributable passage of the first account, and a legally material proposition in that account should remain visible to the reviewer. The first requirement constrains unsupported additions; the second addresses omissions. A source link beside every sentence of a draft cannot reveal that the draft never mentioned a prior threat. Review therefore has to run in both directions: from draft to source for support, and from source to draft for coverage. Fidelity in this sense is a relation between source and record, not a judgment about truth. A faithful record may contain an allegation that another person disputes, or an inference the complainant cannot substantiate, provided it records it as hers and at the certainty she expressed. The Law Enforcement Directive's distinction between facts and personal assessments is the closest existing legal expression of this idea [34].

5.2 A common layer and local overlays

Materiality has two layers. The common layer, derived from Table 3, comprises distinct episodes, relative order, reported objections, threats and the persons threatened, the complainant's source of knowledge, attribution, denial and uncertainty. A local overlay then adds the elements relevant to a given jurisdiction. In India, the repeated-contact limb of BNS section 78 makes the complainant's clear indication of disinterest, and contact that follows it, an element of the offence, while monitoring of electronic

communication is a separate limb [13]. Under the 2024 EU Directive, the same expressed boundary is relevant to the individual assessment of protection needs [35]. Table 4 shows how one source proposition can be retained as a factual label and evaluated under more than one framework without importing one jurisdiction's offence definition into another.

Table 4 Common factual layer and jurisdiction-specific overlays for selected material dimensions

Dimension	Example source proposition	Common-layer requirement	India overlay	EU overlay
Course of conduct (PC)	“I told him on Monday to stop contacting me. He messaged on Tuesday and again on Thursday.”	Two later contacts recorded as distinct and as following the refusal	Clear indication of disinterest and repeated contact: BNS s 78(1)(i) [13]	Persistence despite an expressed boundary as a risk indicator: Directive 2024/1385 art 16 [35]
Identity uncertainty (ID)	“I saw unfamiliar activity on my account. I suspect him but do not know who did it.”	Observation, suspicion and stated uncertainty each preserved	Monitoring of electronic communication as a separate limb: BNS s 78(1)(ii) [13]	Facts distinguished from personal assessments: Directive 2016/680 art 7 [34]
Threats and denials (TH)	“My sister said she heard him threaten me. There was no earlier threat to me.”	Source of knowledge and the denial preserved	Information enabling the officer to apply BNSS s 173 thresholds [12]	Individual assessment of protection needs [33, 35]
Cross-channel link (XL)	“The last message mentioned the market where he had stopped me last week.”	Link between online message and offline encounter retained as a proposition	Facts relevant to both s 78 limbs recorded together	Cyberviolence examined in context of domestic violence [46]

5.3 Correction, safety and the officer's responsibility

A complainant-facing correction step and an official sign-off address different problems. A complainant can identify a mistranslated idiom or a misassigned speaker; the officer remains responsible for what the official record contains and for handling disagreements. Correction should be offered in a language the complainant understands and through a channel that is safe for her, which may not be her own device if that device is the one she believes is being monitored. Urgent protective action must not wait for a draft to be finalised; the Grand Chamber's insistence that

risk assessment remain with competent authorities supports keeping any model out of risk scoring [45].

5.4 Privacy and proportionality

Traceability has privacy costs. A retained recording or transcript can reveal a survivor's location or safe contact details. California's draft-retention rule illustrates the accountability value of preserving transformations [38], but data minimisation and disclosure law constrain how such records should be kept and shared [34]. *Bater-James* illustrates the proportionality limits on access to complainant-held digital material [53]. Source verification requires access to the passage that generated a disputed assertion; it supplies no basis for collecting unrelated device contents. The standard proposed here therefore separates access-controlled originals and versions from routine circulation of the final narrative, and records corrections with their authors.

6. The VANI audit protocol

6.1 Purpose and scope

VANI (Verifiable Account-to-Narrative Integrity; *vāṇi* means speech or voice in Telugu and Sanskrit) is a pre-specified protocol for testing whether an AI drafting workflow preserves the material propositions of a first-person complaint of technology-facilitated intimate-partner stalking across matched language variants. It extends the measurement logic of FACETS-PR [20] in three directions: from officer-observed scenes to a complainant's own disclosure; from a dialect pair to cross-language and code-switched variants; and from describing errors to testing whether a source-linked, agentic verification step reduces them. The protocol isolates drafting from speech recognition by using text input in its primary analysis. This article reports the design.

6.2 Corpus

The corpus consists of independently authored synthetic complaints, each with English (en), Telugu (te) and code-switched Telugu–English (te-en) variants sharing a stable case identifier and line-level alignment. Case types follow the materiality taxonomy: repeated contact after an expressed refusal; suspected electronic monitoring with explicit uncertainty about the actor; linked online and offline incidents; two actors with a specific denial; account ownership and quoted content; and a non-stalking control alleging a single contact. Each case includes qualifying facts that weaken an inference as well as facts that support one. Bilingual reviewers independently verify semantic equivalence across variants before the evaluation set is frozen, and development cases used to refine prompts are kept separate from evaluation cases. Two development cases with all three aligned variants are released with this article (ESM 2). No police, survivor or student records are used.

6.3 Conditions

All conditions use the same model version, decoding settings, source text and output language (English). Condition A produces a direct draft. Condition B produces a single-pass draft in which each assertion is linked to numbered source lines. Condition C is an agentic two-stage workflow: the model first extracts a fact list with source links and review flags, then drafts a narrative, removes assertions without source support, restores relevant omitted details with their source line, and flags unresolved ambiguity. The A–B contrast estimates the effect of requiring links; the B–C contrast estimates the effect of the verification stage. An optional compute-matched Condition D, an unlinked second pass, tests how much of any C advantage is due to an additional inference step alone. Run order is randomised and every attempt, including failures,

is logged with model digest, runtime version, prompt and input hashes, and parameters.

6.4 Annotation

Two bilingual legal reviewers independently assign each gold proposition one primary code: preserved, omitted, contradicted, wrong actor, wrong time, or certainty changed. Where several errors coexist, the primary code follows a fixed precedence (wrong actor, wrong time, certainty changed, contradicted). Unsupported additions are recorded under their own identifiers, and the same transformation is never counted both as a changed gold proposition and as an addition. Reviewers are blinded to condition where feasible; source-link accuracy is scored in a separate view because visible links reveal the workflow. Agreement is reported before adjudication, and adjudicated labels are used for analysis. Table 5 gives coding examples.

Table 5 Illustrative coding decisions under the VANI codebook (hypothetical drafts)

Source account	Draft error	Primary code and consequence
“I asked him to stop on Monday. He contacted me again on Thursday.”	Reports the Thursday contact without the objection	Objection proposition: omitted. Sequence proposition checked separately
“I suspect he accessed the account. I did not see him do it.”	States that he accessed the account	Certainty changed (not also an unsupported addition). Denial of observation scored separately
“My sister said she heard him threaten me.”	States that the complainant heard the threat	Wrong actor (source of knowledge)
“I received one message.”	States that messages arrived repeatedly	Contradicted (frequency); counted once
“There were no threats before this.”	States that he had previously threatened her	Contradicted (denial reversed); critical error

6.5 Outcomes and analysis

The primary outcome is the proportion of gold material propositions preserved, per case, language variant and condition. Secondary outcomes are unsupported assertions per gold proposition, the proportion of cases with at least one critical error

(pre-defined as a change in actor, in a denial, or in an explicit threat), and the proportion of initial errors corrected in Condition C without introducing new ones. The case is the unit of inference: repeated runs are summarised within case, and paired differences between conditions are estimated with case-clustered uncertainty intervals. The protocol reports the language gap within each condition (te and te-en relative to en) and whether verification narrows it, so that a pooled metric cannot conceal weaker performance on code-switched input. Invalid outputs are counted and sensitivity analyses address failed drafts and differences in output length.

A model-only benchmark can show whether a workflow flags and revises errors on the tested corpus. Whether officers or complainants can detect and correct those errors in practice requires a separate reviewer study with appropriate ethics approval and access arrangements.

7. Discussion

7.1 Responsibility without detection

The central finding is that current law allocates responsibility for AI-assisted police records without equipping anyone to detect what a record has lost. Certification identifies a reviewer; retention makes later inspection possible; language access lets a complainant participate. Each supplies something the others cannot, but none measures what the drafting process did to the account. The human-oversight literature predicts that a signature obtained under these conditions will often ratify rather than review [29, 30]. For agentic pipelines the problem compounds, because each intermediate step (transcript, translation, fact list) is a further transformation that a reviewer reading only the final English narrative cannot see.

7.2 Why this is a gender-equity question

The failure modes identified here are not evenly distributed. Complaints of intimate-partner stalking are cumulative, multi-actor and laden with uncertainty about who controls an account; they are frequently made by women speaking in a language other than the language of the record; and in India they are made at a gate where institutional discretion over registration has long been gendered [15, 16]. A system that performs well on short, single-actor, monolingual English accounts and less well on long, code-switched, multi-actor ones would reproduce existing inequality while appearing neutral, and aggregate accuracy figures would hide it. Measuring fidelity separately for each language variant and each material dimension is therefore the operational form of the equity-by-design principle that governance frameworks in this collection advocate [11]. It also aligns with Sustainable Development Goal targets on eliminating violence against women and girls (SDG 5.2) and ensuring equal access to justice (SDG 16.3).

7.3 Implications for procurement and regulation

Four implications follow. First, procurement should require vendors to demonstrate, on held-out bilingual cases, material-fact recall and unsupported-assertion rates reported separately by supported language and communication mode, rather than section-prediction accuracy or time savings alone. Second, the reviewing officer's interface should display, for any disputed assertion, the underlying source passage in the original language, the translated passage, and the changed wording; demonstrating that links exist is insufficient when they point to the wrong segment. Third, read-back under BNSS section 173 should be understood, where a tool generates the draft, to require a read-back of the substance in a language the informant understands, with a recorded opportunity to correct. Fourth, statutes

modelled on California and Utah would be strengthened by a fidelity-testing requirement and a complainant-facing correction channel; and instruments modelled on the Victims' Rights Directive would be strengthened by an express provision on generated intermediaries.

7.4 Limitations

The selected jurisdictions are functionally illustrative and do not support claims about worldwide prevalence or consensus. Coding was performed by a single coder; scores are released with pinpoints to enable re-coding, and inter-coder agreement should be established in replication. The legal landscape is moving: the AI Act's high-risk provisions are subject to a staged application timetable and amendment, and the 2024 EU Directive has a transposition deadline of 14 June 2027 [35, 36]. The VANI protocol uses synthetic text and has not yet been executed; real complaints involve interruption, variation in speech and literacy, and institutional workload that a text-first design cannot capture. Public information on the Hyderabad deployment is limited to launch and supplier reporting, and no administrative records on its operating procedures or error rates were available. The Tasmanian evaluation concerns body-worn camera audio rather than complainant-initiated intake.

8. Conclusion

The passage from first disclosure to official record is where agentic AI can most quietly change whose account the state holds and what it says. The comparative analysis shows that the law currently secures the complainant's voice at the start of that passage and the officer's signature at its end, but nothing between. The jurisprudence identifies what must survive: the pattern of conduct, its chronology, threats and the persons threatened, the history of earlier reports, the link between online and offline

conduct, and the line between what a woman saw and what she suspects. Reviewable fidelity and the VANI protocol turn those requirements into questions that a vendor, an agency or a court can ask and answer before a record is adopted.

Supplementary information

ESM 1 (ESM 1 Coding Workbook.xlsx): instrument and decision coding with pinpoint provisions for every score. ESM 2 (ESM 2 VANI Protocol.zip): codebook, development fixtures, versioned prompts, data dictionary, run and annotation templates, scoring script and search log.

References

1. Gibson J. AI tool from US manufacturer Axon drafts family violence reports for Tasmania Police. ABC News. 23 Oct 2026. https://www.abc.net.au/news/2026-10-23/ai-assisted-police-report-writing-dv-domestic-violence/107297138. Accessed 17 Nov 2026.
2. Tasmania Police. Right to Information application 557/26: in-scope documents with redactions [Draft One evaluation materials]. Hobart: Tasmania Police; 2026. https://www.police.tas.gov.au/uploads/RTI-55726-IN-SCOPE-DOCUMENTS-with-redactions.pdf. Accessed 16 Nov 2026.
3. Freed D, Palmer J, Minchala D, Levy K, Ristenpart T, Dell N. “A stalker’s paradise”: how intimate partner abusers exploit technology. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. New York: ACM; 2018. https://doi.org/10.1145/3173574.3174241.
4. Dragiewicz M, Burgess J, Matamoros-Fernández A, Salter M, Suzor NP, Woodlock D, Harris B. Technology facilitated coercive control: domestic violence and the competing roles of digital media platforms. Fem Media Stud. 2018;18(4):609–625. https://doi.org/10.1080/14680777.2018.1447341.
5. Cuomo D, Dolci N. New tools, old abuse: technology-enabled coercive control (TECC). Geoforum. 2021;126:224–232. https://doi.org/10.1016/j.geoforum.2021.08.002.
6. Telangana Today. Hyderabad police launch AI-CopWriter to accept complaints. 23 May 2026. https://telanganatoday.com/hyderabad-police-launch-ai-copwriter-to-accept-complaints. Accessed 17 Nov 2026.

7. Express Computer. BCSSL deploys AI-powered multilingual FIR app with Hyderabad City Police. 6 Jul 2026. https://www.expresscomputer.in/news/bcssl-deploys-ai-powered-multilingual-fir-app-with-hyderabad-city-police/136586/. Accessed 16 Nov 2026.

8. National Crime Records Bureau. NCRB Newsletter, July–September 2026. New Delhi: Ministry of Home Affairs; 2026. p. 6. https://www.ncrb.mha.gov.in/uploads/files/NewsletterJulytoSeptember2026.pdf. Accessed 16 Nov 2026.

9. Mathivanan P, Mohanapriya D, Aravind B, Gokul Bharath V, Krishna L, Tamizhvanan S. AI-enabled smart FIR drafting and legal validation system. In: 2026 4th International Conference on Artificial Intelligence and Machine Learning Applications (AIMLA). Namakkal: IEEE; 2026. https://doi.org/10.1109/AIMLA67915.2026.11522503.

10. Chan A, Salganik R, Markelius A, Pang C, Rajkumar N, Krasheninnikov D, et al. Harms from increasingly agentic algorithmic systems. In: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. New York: ACM; 2023. p. 651–666. https://doi.org/10.1145/3593013.3594033.

11. Jain B, Gupta R. Advancing gender equity and inclusive digital futures through responsible agentic artificial intelligence governance. Discov Glob Soc. 2026;4:281. https://doi.org/10.1007/s44282-026-00624-y.

12. Government of India. The Bharatiya Nagarik Suraksha Sanhita, 2023 (Act 46 of 2023), s 173. https://www.indiacode.nic.in/bitstream/123456789/20340/1/bnss%2C_2023.pdf. Accessed 17 Nov 2026.

13. Government of India. The Bharatiya Nyaya Sanhita, 2023 (Act 45 of 2023), s 78. https://www.indiacode.nic.in/bitstream/123456789/20062/1/a2023-45.pdf. Accessed 16 Nov 2026.

14. Supreme Court of India. Lalita Kumari v Government of Uttar Pradesh. (2014) 2 SCC 1, para 120. https://indiankanoon.org/doc/10239019/. Accessed 16 Nov 2026.

15. Sukhtankar S, Kruks-Wisner G, Mangla A. Policing in patriarchy: an experimental evaluation of reforms to improve police responsiveness to women in India. Science. 2022;377(6602):191–198. https://doi.org/10.1126/science.abm7387.

16. Jassal N. Gender, law enforcement, and access to justice: evidence from all-women police stations in India. Am Polit Sci Rev. 2020;114(4):1035–1054. https://doi.org/10.1017/S0003055420000684.

17. Adams IT, et al. No man’s hand: artificial intelligence does not improve police report writing speed. J Exp Criminol. 2024. https://doi.org/10.1007/s11292-024-09644-7.

18. Ferguson AG. Generative suspicion and the risks of AI-assisted police reports. Northwest Univ Law Rev. 2025;120(2). https://scholarlycommons.law.northwestern.edu/nulr/vol120/iss2/9/.

19. University of Chicago Justice Project. Hallucinations in AI-assisted police reports [project report]. 2026. https://uchicago-justice-project.github.io/ai_police_report_landing_page/. Accessed 16 Nov 2026.

20. Aiersilan A, Wuernikebai Y. Measuring source fidelity, exculpatory recall, and generative suspicion in AI-drafted police reports. Proc AAAI Symp Ser. 2026;10(1):226–236. https://doi.org/10.1609/aaaiss.v10i1.43410.

21. Takshashila Institution. Smart FIRs in India's police system [policy brief]. Bengaluru: Takshashila Institution; 2026. https://takshashila.org.in/content/publications/assets/LEPF-Policy-Brief-AI-Assisted-FIR-Registration.pdf. Accessed 16 Nov 2026.

22. Gala J, Chitale PA, AK R, Gumma V, Doddapaneni S, Kumar A, et al. IndicTrans2: towards high-quality and accessible machine translation models for all 22 scheduled Indian languages. Trans Mach Learn Res. 2023. https://arxiv.org/abs/2305.16307.

23. AI4Bharat. Kathbath: speech recognition benchmark for Indian languages [dataset]. Hugging Face; 2022. https://huggingface.co/datasets/ai4bharat/Kathbath. Accessed 16 Nov 2026.

24. Hasan T, Bhattacharjee A, Islam MS, Mubasshir K, Li YF, Kang YB, et al. XL-Sum: large-scale multilingual abstractive summarization for 44 languages. In: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. ACL; 2021. p. 4693–4703. https://doi.org/10.18653/v1/2021.findings-acl.413.

25. Zhang R, Cahyawijaya S, Cruz JCB, Winata GI, Aji AF. Multilingual large language models are not (yet) code-switchers. In: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Singapore: ACL; 2023. p. 12567–12582. https://doi.org/10.18653/v1/2023.emnlp-main.774.

26. Vieira LN, O'Hagan M, O'Sullivan C. Understanding the societal impacts of machine translation: a critical review of the literature on medical and legal use cases. Inf Commun Soc. 2021;24(11):1515–1532. https://doi.org/10.1080/1369118X.2020.1776370.

27. Min S, Krishna K, Lyu X, Lewis M, Yih W, Koh PW, et al. FActScore: fine-grained atomic evaluation of factual precision in long form text generation. In: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Singapore: ACL; 2023. p. 12076–12100. https://doi.org/10.18653/v1/2023.emnlp-main.741.

28. Al-Omoush KS. Explaining cyberbullying and cyber violence in the age of AI using individual and contextual factors in Jordan. Discov Glob Soc. 2026;4:278. https://doi.org/10.1007/s44282-026-00625-x.

29. Green B. The flaws of policies requiring human oversight of government algorithms. Comput Law Secur Rev. 2022;45:105681. https://doi.org/10.1016/j.clsr.2022.105681.

30. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc. 2012;19(1):121–127. https://doi.org/10.1136/amiajnl-2011-000089.

31. Hall MA, Wright RF. Systematic content analysis of judicial opinions. Calif Law Rev. 2008;96(1):63–122.

32. Government of India. The Bharatiya Sakshya Adhiniyam, 2023 (Act 47 of 2023), s 63. https://www.indiacode.nic.in/indiacode/bitstream/123456789/20063/1/aa202347.pdf. Accessed 16 Nov 2026.

33. Directive 2012/29/EU of the European Parliament and of the Council of 25 October 2012 establishing minimum standards on the rights, support and protection of victims of crime. OJ L 315, 14.11.2012, p. 57–73. https://eur-lex.europa.eu/eli/dir/2012/29/oj/eng.

34. Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences. OJ L 119, 4.5.2016, p. 89–131. https://eur-lex.europa.eu/eli/dir/2016/680/oj/eng.

35. Directive (EU) 2024/1385 of the European Parliament and of the Council of 14 May 2024 on combating violence against women and domestic violence. OJ L, 2024/1385, 24.5.2024. https://eur-lex.europa.eu/eli/dir/2024/1385/oj/eng.

36. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), as amended. OJ L, 2024/1689, 12.7.2024. https://eur-lex.europa.eu/eli/reg/2024/1689. Accessed 16 Nov 2026.

37. Council of Europe. Convention on Preventing and Combating Violence against Women and Domestic Violence (CETS No. 210). Istanbul; 2011. https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treatynum=210.

38. California Legislature. Senate Bill 524 (2025–2026 Reg. Sess.), Stats. 2025, ch. 587, codified at Penal Code § 13663. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202520260SB524. Accessed 16 Nov 2026.

39. Utah Legislature. S.B. 180 (2025 General Session), enrolled, codified at Utah Code §§ 53-25-601 to 53-25-602. https://le.utah.gov/Session/2025/bills/enrolled/SB0180.pdf. Accessed 16 Nov 2026.

40. Committee on the Elimination of Discrimination against Women. General recommendation No. 33 on women's access to justice. CEDAW/C/GC/33. 3 Aug 2015. https://docs.un.org/en/CEDAW/C/GC/33.

41. Committee on the Elimination of Discrimination against Women. General recommendation No. 35 on gender-based violence against women, updating general recommendation No. 19. CEDAW/C/GC/35. 26 Jul 2017. https://docs.un.org/en/CEDAW/C/GC/35.

42. National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. NIST AI 600-1. Gaithersburg: NIST; 2024. https://doi.org/10.6028/NIST.AI.600-1.

43. European Court of Human Rights. Opuz v Turkey. Application no. 33401/02. Judgment of 9 June 2009. https://hudoc.echr.coe.int/?i=001-92945.

44. European Court of Human Rights. Talpis v Italy. Application no. 41237/14. Judgment of 2 March 2017. https://hudoc.echr.coe.int/eng?i=001-171994.

45. European Court of Human Rights (Grand Chamber). Kurt v Austria. Application no. 62903/15. Judgment of 15 June 2021, paras 164–177, 190. https://hudoc.echr.coe.int/eng?i=001-210463.

46. European Court of Human Rights. Buturugă v Romania. Application no. 56867/15. Judgment of 11 February 2020, paras 72–79. https://hudoc.echr.coe.int/eng?i=001-201342.

47. European Court of Human Rights. Volodina v Russia (No. 2). Application no. 40419/19. Judgment of 14 September 2021. https://hudoc.echr.coe.int/eng?i=003-7116031-9638379.

48. Inter-American Commission on Human Rights. Maria da Penha Maia Fernandes v Brazil. Case 12.051, Report No. 54/01. 16 April 2001. https://cidh.oas.org/annualrep/2000eng/ChapterIII/Merits/Brazil12.051.htm.

49. Inter-American Court of Human Rights. González et al. ("Cotton Field") v Mexico. Preliminary objection, merits, reparations and costs. Judgment of 16 November 2009. Series C No. 205. https://hist.cndh.org.mx/index.php/documento/caso-gonzalez-y-otras-campo-algodonero-vs-mexico.

50. Committee on the Elimination of Discrimination against Women. González Carreño v Spain. Communication No. 47/2012. CEDAW/C/58/D/47/2012. Views adopted 16 July 2014, paras 9.4–9.9. https://docs.un.org/en/CEDAW/C/58/D/47/2012.

51. Supreme Court of Canada. R v Tran. [1994] 2 SCR 951, 985–990. https://decisions.scc-csc.ca/scc-csc/scc-csc/en/1166/1/document.do.

52. Supreme Court of the United States. Davis v Washington; Hammon v Indiana. 547 US 813, 822, 827–832 (2006). https://www.govinfo.gov/app/details/USREPORTS-547/USREPORTS-547-813.

53. Court of Appeal of England and Wales (Criminal Division). Bater-James and Mohammed v R. [2020] EWCA Crim 790, paras 70–74. https://www.bailii.org/ew/cases/EWCA/Crim/2020/790.html.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1CodingWorkbook.xlsx
- ESM2VANIProtocol.zip