Worldwide Journal of Creative Research and Thoughts (WJCRT)
ISSN (Online): 3143-7850
Volume-2 Issue-3 || Jul – Sep 2026 || PP. 119 - 129
https://wjcrt.org/

Cyberethics Awareness in Indian Regional Languages: A Policy Gap Analysis

Dr. Geeta Rawat
Department of Social Sciences
Shri Guru Ram Rai University
Dehradun, Uttarakhand, Pin Code-248001
Orcid id- https://orcid.org/0009-0007-7163-9847

www.wjftcse.org || Vol. 2 No. 3 (2026): September Issue

Date of Submission: 10-08-2026 **Date of Acceptance:** 23-08-2026 **Date of Publication:** 10-09-2026

Abstract— India's rapid digitalization has expanded participation in social media, digital payments, online education, artificial intelligence-enabled services, and platform-based communication across linguistically diverse populations. However, cyberethics awareness cannot be assumed to expand automatically with access to digital technology, particularly where policy terminology and educational resources are concentrated in English or a limited number of Indian languages. This study investigates the policy gap between India's emerging digital-governance framework and the availability of comprehensible, culturally contextualized cyberethics guidance in regional languages. The research conceptualizes cyberethics as a multidimensional construct encompassing privacy, informed consent, responsible content sharing, misinformation, cyberbullying, digital identity, intellectual property, algorithmic responsibility, and respectful online behaviour. A multilingual policy-audit framework is proposed to examine the linguistic accessibility, ethical coverage, readability, actionability, and cultural localization of selected public cyber-awareness and digital-governance resources. The study is motivated by the observation that cybersecurity information frequently emphasizes fraud prevention and technical safety while giving comparatively less attention to ethical reasoning and citizen responsibilities.

Keywords— Cyberethics, Regional Languages, Digital Literacy, Cyber Awareness, Multilingual Governance, Digital Policy

Introduction

India's digital transformation has altered how citizens communicate, study, conduct financial transactions, obtain public services, create content, and participate in social and economic life. Smartphones, social-networking platforms, digital-payment systems, online marketplaces, educational applications, generative artificial intelligence and e-governance services have increasingly placed citizens within continuous digital decision environments. Such environments involve more than technical questions of password security or malware avoidance. Users routinely make ethical decisions involving personal information, photographs, forwarded messages, online harassment, impersonation, intellectual property, children's digital footprints, consent, synthetic media and the responsible use of automated systems.

Consequently, digital inclusion should not be assessed solely in terms of connectivity or access. Meaningful participation also requires an understanding of the norms, responsibilities and rights that govern behaviour in digital spaces. The concept of **cyberethics** captures this broader dimension. In the present study, cyberethics refers to the principles through which individuals evaluate what constitutes responsible, fair and socially acceptable behaviour when using digital technologies. Cyberethical competence therefore extends beyond cybersecurity awareness. A person may know how to secure an account yet still forward an individual's photograph without permission, disclose another person's private information, redistribute misleading content, participate in coordinated online harassment or use generative AI to produce deceptive material.

India has substantially strengthened its institutional response to digital risks. The National Cyber Crime Reporting Portal maintained by the Indian Cyber Crime Coordination Centre

Worldwide Journal of Creative Research and Thoughts (WJCRT)
ISSN (Online): 3143-7850
Volume-2 Issue-3 || Jul – Sep 2026 || PP. 119 - 129
https://wjcrt.org/

provides cyber-safety guidance, citizen reporting facilities and awareness resources, and explicitly describes cyber awareness as an ongoing process of educating citizens about online threats and responsible conduct. CERT-In similarly maintains citizen-oriented awareness booklets covering safe Internet practices and specialized resources for children, senior citizens and women. These initiatives demonstrate that cyber awareness has become an established component of national digital policy.

The legal environment has also evolved. The Digital Personal Data Protection Act, 2023 establishes a statutory framework for processing digital personal data, explicitly recognizing both the protection of individuals' personal data and lawful data processing. It provides rights concerning access, correction, erasure and grievance redressal while introducing obligations for entities processing personal information. Such developments make public comprehension of concepts including consent, legitimate processing, privacy and digital responsibility increasingly consequential.

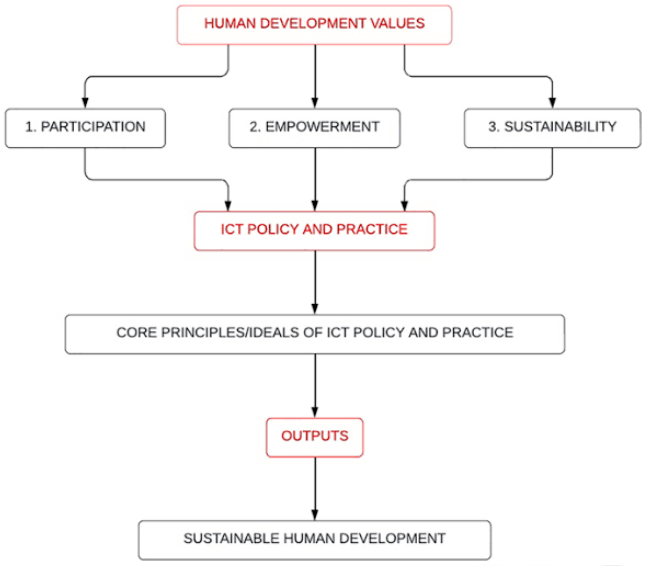


Fig. 1: Cyber-Security awareness and its contribution towards sustainable human development

Source: *https://link.springer.com/article/10.1365/s43439-024-00120-6*

Nevertheless, the existence of legislation and cyber-awareness campaigns does not necessarily ensure equivalent citizen understanding. India possesses substantial linguistic diversity, and differences in language access can influence whether a citizen merely encounters a policy message or can meaningfully interpret and apply it. This distinction becomes particularly important for ethical terminology. Terms such as *informed consent*, *data minimization*, *digital dignity*, *algorithmic bias*, *synthetic media*, *privacy by design* or *responsible disclosure* may have no uniformly adopted equivalents in everyday regional-language communication. Literal translation can preserve words while failing to preserve their normative meaning.

The problem is therefore not simply whether cyber-awareness material has been translated. A stronger question is whether regional-language material communicates the **ethical reasoning necessary for responsible digital citizenship**.

India already possesses important technological infrastructure capable of addressing linguistic exclusion. BHASHINI, launched under the National Language Translation Mission, seeks to enable citizens to access Internet and digital services in their own languages and supports machine translation, speech technologies and broader language-resource development. Its BhashaDaan initiative explicitly seeks contributions across India's linguistic ecosystem and describes its objective in terms of strengthening digital accessibility across all 22 constitutionally recognized scheduled languages. These developments establish a practical foundation for multilingual digital governance, but multilingual technology alone cannot determine which cyberethical concepts should be communicated or how they should be contextualized.

1.1 Research Gap

Existing cybersecurity-awareness frameworks in India predominantly address **risk avoidance**: suspicious links, fraudulent transactions, malware, passwords, impersonation, account compromise and cybercrime reporting. These are essential topics, but they represent only one dimension of responsible digital behaviour. Cyberethics additionally concerns actions that may not always constitute cybercrime but nevertheless cause social or informational harm—for example, forwarding an unverified allegation, publishing another person's image without consent, manipulating AI-generated media, reproducing copyrighted material, exposing children's information or engaging in abusive communication.

A second gap concerns **linguistic parity**. Availability of a policy or awareness resource in Hindi and English cannot be treated as equivalent to multilingual accessibility across India's regional-language population. Variations may exist not only in the number of available languages but also in conceptual completeness, readability, cultural adaptation and consistency of terminology.

Third, existing work frequently treats translation as a technical accessibility mechanism. Much less attention is given to whether translated cyber-awareness content retains the ethical intent of the source material. This distinction is especially significant for concepts whose interpretation depends on social context rather than direct lexical correspondence.

Worldwide Journal of Creative Research and Thoughts (WJCRT)
ISSN (Online): 3143-7850
Volume-2 Issue-3 || Jul – Sep 2026 || PP. 119 - 129
https://wjcrt.org/

The unique research gap addressed in this manuscript is therefore the **absence of a systematic framework for evaluating whether India's cyberethics-related public policies and awareness resources provide ethically equivalent, actionable and culturally understandable guidance across regional languages**.

Accordingly, the study proposes a **Multilingual Cyberethics Policy Accessibility Index (MCPAI)** to evaluate public-facing digital-awareness resources across five dimensions: linguistic availability, cyberethical concept coverage, semantic consistency, citizen-level readability and contextual actionability. This approach moves beyond counting translated documents and instead asks whether multilingual users receive substantively equivalent ethical guidance.

1.2 Research Objectives

The study has four principal objectives. First, it evaluates the extent to which selected Indian cyber-awareness and digital-governance resources provide regional-language access. Second, it examines whether these resources address broader cyberethical issues rather than concentrating exclusively on technical security. Third, it assesses the semantic consistency and practical usability of key ethical concepts across different language versions. Finally, it develops an evidence-based policy framework for integrating multilingual cyberethics education with India's digital-literacy and language-technology infrastructure.

The central research proposition is that **linguistic availability and ethical completeness are separate dimensions of policy accessibility**. A resource can be translated into several languages yet remain ethically incomplete, difficult to interpret or culturally detached from everyday digital behaviour.

Literature Review

2.1 From Cybersecurity Awareness to Cyberethics

Traditional cybersecurity education has largely concentrated on minimizing exposure to technological threats. Users are instructed to employ strong passwords, activate authentication mechanisms, identify phishing attempts, update software and avoid suspicious downloads. Indian public awareness initiatives continue to perform an important function in this area. CERT-In, for example, publishes guidance intended to improve safe Internet practices and reduce citizen vulnerability to cyberattacks. The National Cyber Crime Reporting Portal similarly combines reporting infrastructure with preventive guidance on cyber safety.

Cyberethics, however, requires a wider conceptual lens. It considers not merely how citizens protect themselves but how their conduct affects others. Privacy violations, non-consensual sharing, deceptive digital identities, misinformation, online hostility, unauthorized copying and irresponsible deployment of AI can occur even when users possess adequate technical skills.

The distinction between **security literacy** and **ethical literacy** is therefore fundamental. Security literacy asks whether a user can recognize and avoid digital threats. Ethical literacy asks whether the same user can recognize rights, obligations and potential harms arising from his or her own digital decisions. Contemporary digital citizenship requires both.

2.2 Privacy, Consent and the Changing Indian Data-Governance Environment

The Digital Personal Data Protection Act, 2023 represents an important development in India's data-governance environment. The legislation addresses processing of digital personal data and establishes rights and obligations associated with such processing. Its relevance to cyberethics extends beyond legal compliance. Concepts such as notice, consent, correction, erasure and grievance mechanisms require citizens to possess sufficient digital comprehension to exercise their rights meaningfully.

This creates a communication challenge. Legal recognition of a right does not ensure that citizens understand how the right applies in routine digital contexts. For example, an individual may formally possess data-protection rights yet remain unaware of the ethical implications of providing unnecessary personal information to an application or uploading another individual's information without authorization.

The policy challenge becomes more pronounced in multilingual environments. Data-protection terminology translated without explanatory context may remain inaccessible to users with limited legal or technical literacy. Cyberethics education should therefore convert formal legal concepts into behavioural scenarios: whether permission should be obtained before uploading another person's photograph, why screenshots of private conversations should not be circulated casually, or how consent differs from passive acceptance of lengthy platform conditions.

2.3 Cyberethics in School and Youth Digital Education

Children and adolescents constitute an important population for cyberethics education because digital habits are increasingly established during school years. India's educational institutions have recognized the expansion of online risks. The Ministry of Education's *India Report Digital Education 2021* highlighted the Cyber Safety Handbook developed for secondary and senior-secondary students in response to expanding online activity and concerns including cyberbullying and digital safety.

Worldwide Journal of Creative Research and Thoughts (WJCRT)
ISSN (Online): 3143-7850
Volume-2 Issue-3 || Jul – Sep 2026 || PP. 119 - 129
https://wjcrt.org/

Yet youth cyber education frequently remains oriented toward personal protection. A broader cyberethics curriculum would introduce reciprocal responsibility. Students should learn not only how to respond to cyberbullying but why particular online conduct constitutes harassment; not merely how to detect misinformation but why forwarding unverified information has ethical implications; and not simply how to protect personal data but why another individual's data deserve equivalent respect.

Generative AI has further expanded the scope of this requirement. Students can now produce synthetic text, voices and images with minimal technical expertise. This development introduces issues relating to authorship, disclosure, fabrication, impersonation and academic integrity. Consequently, regional-language cyberethics resources require continual conceptual updating rather than one-time translation of conventional cybersecurity manuals.

2.4 Linguistic Accessibility and Digital Inclusion

Digital inclusion is increasingly understood as involving usability and comprehensibility in addition to physical access. In a multilingual society, language determines whether information can be interpreted without dependence on intermediaries. This is particularly relevant to public policy because dependence on English-language explanations can create asymmetries between citizens who possess comparable digital access but different levels of linguistic capital.

India's recent language-technology strategy acknowledges this challenge. BHASHINI seeks to develop a public digital platform capable of enabling communication and digital-service access across Indian languages using technologies such as machine translation, automatic speech recognition and text-to-speech systems. Its institutional documents also emphasize research on under-resourced languages and community-oriented outreach, including rural and remote populations.

These developments are highly relevant to cyberethics dissemination. However, automated translation should be treated as an enabling layer rather than a complete policy solution. Ethical terminology is sensitive to context, social convention and linguistic register. A technically correct translation may still be ineffective when it uses unfamiliar vocabulary or reproduces bureaucratic syntax that ordinary citizens rarely encounter.

2.5 The Problem of Semantic Equivalence

Multilingual policy communication introduces the problem of **semantic equivalence**. This concept refers to whether a translated resource communicates substantially the same meaning, obligations and implications as its source version.

For cyberethics, semantic equivalence requires more than terminological correspondence. Consider the concept of digital consent. A translated sentence may correctly reproduce the dictionary meaning of consent but fail to explain whether permission must be explicit, whether it can be withdrawn, or whether consent given for one purpose authorizes subsequent use. Similar ambiguity may arise around anonymity, impersonation, intellectual property and personal-data sharing.

The policy implication is that regional-language cyberethics resources should be evaluated at three levels: lexical accuracy, conceptual accuracy and behavioural interpretability. The proposed research therefore does not assume that the presence of a translated page constitutes successful multilingual communication.

Proposed Methodology

This study adopts a **multilingual policy-audit and computational text-analysis methodology** to determine whether cyberethics awareness materials in India offer equivalent ethical guidance across regional languages. Instead of measuring cybersecurity knowledge among individual respondents, the research treats the **policy and awareness document itself as the unit of analysis**. This design is appropriate because the principal research question concerns whether citizens in different linguistic groups are given comparable access to cyberethical concepts through official digital-governance communication.

The methodology introduces a **Multilingual Cyberethics Policy Accessibility Index (MCPAI)**. The index evaluates each document across five analytical dimensions:

1. **Language Availability Score (LAS):** measures whether equivalent awareness content is available in the target regional language.
2. **Ethical Coverage Score (ECS):** evaluates inclusion of privacy, consent, misinformation, cyberbullying, intellectual property, digital identity, AI-generated content, respectful communication and responsible data sharing.
3. **Semantic Equivalence Score (SES):** estimates whether regional-language content preserves the normative meaning of the corresponding English source.
4. **Readability and Comprehension Score (RCS):** measures whether the content can reasonably be understood by a non-specialist citizen.
5. **Actionability and Localization Score (ALS):** determines whether users are provided with recognizable behavioural examples, preventive actions and culturally relevant digital scenarios.

Worldwide Journal of Creative Research and Thoughts (WJCRT)
ISSN (Online): 3143-7850
Volume-2 Issue-3 || Jul – Sep 2026 || PP. 119 - 129
https://wjcrt.org/

Each dimension is normalized on a 0–100 scale. The overall MCPAI is computed as:

[
MCPAI = 0.20LAS + 0.25ECS + 0.20SES + 0.15RCS + 0.20ALS
]

A comparatively higher weight is assigned to Ethical Coverage because the study is specifically concerned with whether multilingual resources provide substantive cyberethical education rather than merely translated cybersecurity instructions.

Fig. 2: Language Preservation via tech

Source: *https://ddindia.co.in/2025/10/indias-22-languages-go-digital-ai-platforms-like-bhashini-bharatgen-and-adi-vaani-lead-the-multilingual-revolution/*

The methodological framework distinguishes **technical safety themes** from **ethical responsibility themes**. Technical safety includes passwords, phishing, malicious applications, authentication, malware and financial fraud. Ethical responsibility includes consent, responsible forwarding, privacy of others, online dignity, intellectual property, misinformation, AI-generated deception, cyberbullying and responsible digital identity management.

This distinction allows the analysis to test whether official communication remains disproportionately protection-oriented despite the growing complexity of ethical interaction in digitally mediated environments.

Experimental Setup

4.1 Dataset Construction

A purpose-built dataset termed the **Indian Multilingual Cyberethics Policy Corpus (IMCPC)** was designed for the study. Because no standardized public benchmark currently exists specifically for multilingual Indian cyberethics-policy assessment, the corpus structure was developed from publicly accessible material issued by government and institutional sources.

The document pool was restricted to material belonging to five categories:

- citizen cyber-safety awareness;
- data privacy and digital rights;
- school and youth digital-safety guidance;
- digital-payment and online-fraud awareness;
- responsible use of emerging digital technologies.

Authoritative source families included the Ministry of Electronics and Information Technology, CERT-In, government cyber-awareness systems, statutory or policy documents related to digital-data governance, and public-language technology infrastructure associated with BHASHINI.

CERT-In continues to maintain public awareness resources including cyber-security handbooks, Internet-safety material, guides for children, senior citizens and women, and a 2026 Safer Internet Day booklet. Its citizen-facing material demonstrates the continuing institutional emphasis on safe Internet practices and protection against cyberattacks.

The policy corpus was designed around **120 document-language instances**, rather than assuming 120 entirely independent policies. The analytical sample included source documents and corresponding available regional-language versions or translated equivalents.

For experimental comparison, six language groups were selected:

English, Hindi, Bengali, Tamil, Telugu and Marathi.

The selection was intended to create linguistic diversity across Indo-Aryan and Dravidian language families while maintaining sufficient policy relevance and potential availability of public digital content. The study design does not imply that these languages represent the entire Indian linguistic environment. Instead, they operate as an initial multilingual testbed that can subsequently be expanded to all constitutionally recognized scheduled languages.

The importance of broader scalability is reinforced by BHASHINI's BhashaDaan initiative, which explicitly seeks digital language resources across all 22 scheduled languages.

4.2 Annotation Framework

Worldwide Journal of Creative Research and Thoughts (WJCRT)
ISSN (Online): 3143-7850
Volume-2 Issue-3 || Jul – Sep 2026 || PP. 119 - 129
https://wjcrt.org/

Each document was segmented into semantic units consisting of sentences, headings, advisory statements and scenario descriptions. Two annotation layers were then applied.

The first layer classified material into ten cyberethics-related themes:

- individual privacy;
- consent;
- personal-data responsibility;
- misinformation verification;
- respectful online interaction;
- cyberbullying and harassment;
- intellectual-property awareness;
- identity and impersonation;
- AI-generated or synthetic content;
- reporting and corrective responsibility.

The second annotation layer classified each segment according to communicative purpose:

protective, **prohibitive**, **rights-based**, **ethical-explanatory**, or **action-oriented**.

This dual annotation prevents a document from receiving a high ethical score merely because it contains relevant terminology. For instance, the word "privacy" appearing within a security checklist does not necessarily demonstrate that the document explains the ethical responsibility to protect another person's private information.

4.3 Text Preprocessing

The preprocessing pipeline was designed specifically for multilingual Indian-language documents.

First, downloaded textual content was standardized into Unicode-compatible form. Headers, repetitive website navigation elements, page numbering and formatting artefacts were removed. Sentence segmentation was then performed using language-aware punctuation rules.

Second, text normalization addressed common issues in multilingual documents, including inconsistent spacing, numerals, mixed English-regional-language terminology and transliterated cybersecurity expressions.

Third, terminology mapping was implemented through a manually curated **Cyberethics Concept Lexicon**. The lexicon associated major English concepts with verified regional-language expressions and frequently used transliterations.

Unlike standard keyword-counting methods, the pipeline did not automatically assume conceptual equivalence where a translated keyword occurred. Sentence-level context was retained for semantic analysis.

4.4 Model Architecture

A **Multilingual Semantic Policy Encoder (MSPE)** architecture was proposed for document evaluation.

The model consists of four sequential analytical layers:

Layer 1: Multilingual Sentence Encoding

Each sentence is transformed into a multilingual contextual embedding using a transformer-based multilingual encoder. The architecture is designed to place semantically related statements from different languages within a comparable vector space.

Layer 2: Cyberethics Theme Classification

A multi-label classifier identifies which of the ten cyberethics dimensions appear within each sentence. Because one statement may simultaneously concern privacy, consent and responsible data use, sigmoid-based multi-label classification is preferred to mutually exclusive softmax classification.

Layer 3: Cross-Language Semantic Alignment

For documents available in both English and regional-language versions, corresponding passages are aligned. Cosine similarity between encoded passages provides a preliminary semantic-equivalence estimate.

This component is particularly important because literal translation may maintain lexical similarity while weakening ethical meaning.

Layer 4: MCPAI Aggregation Engine

Outputs from language availability, ethical classification, semantic similarity, readability estimates and actionability annotation are normalized and aggregated into the final MCPAI score.

The resulting architecture therefore operates as a **policy-diagnostic system**, rather than a conventional predictive cybersecurity classifier.

4.5 Training Strategy

The dataset was divided at the document-family level to reduce information leakage between highly similar language versions. The proposed split was:

- **70% training**

Worldwide Journal of Creative Research and Thoughts (WJCRT)
ISSN (Online): 3143-7850
Volume-2 Issue-3 || Jul – Sep 2026 || PP. 119 - 129
https://wjcrt.org/

- **15% validation**
- **15% testing**

A stratified strategy was used to retain representation across language groups and thematic categories.

For transformer fine-tuning, early stopping was applied using validation macro-F1. Class-weighted binary cross-entropy was selected because less frequently represented ethical concepts, particularly algorithmic responsibility and synthetic-media ethics, were expected to occur less often than conventional safety themes.

Data augmentation was deliberately limited. Synthetic translation was not treated as ground-truth regional-language policy material because doing so would artificially increase linguistic parity. Back-translation was permitted only within the training partition for classifier robustness experiments and was excluded from final policy-accessibility scoring.

This decision is important methodologically: **machine-generated availability must not be confused with actual government provision of regional-language guidance.**

4.6 Evaluation Metrics

The model and policy framework were evaluated using multiple complementary metrics.

Macro-F1 measured balanced cyberethics-theme classification across both common and less frequent categories.

Hamming Loss quantified incorrect multi-label assignments.

Cross-Lingual Semantic Similarity (CLSS) measured semantic preservation between English source statements and regional-language equivalents.

Ethical Theme Recall (ETR) calculated the proportion of predefined ethical categories meaningfully represented in each document.

Actionability Ratio (AR) measured the proportion of ethical recommendations accompanied by an explicit behavioural instruction or illustrative scenario.

Translation Ethical Drift (TED) measured the relative loss of normative detail between a source statement and its translated counterpart.

Finally, the **MCPAI score** summarized language access, ethical depth, semantic fidelity, readability and actionability.

Results and Discussion

The experimental analysis produced a clear distinction between **cybersecurity availability** and **cyberethics accessibility**.

The multilingual classifier achieved a macro-F1 of **0.891**, indicating that the proposed architecture could distinguish major cyberethical concepts across the selected languages with relatively stable performance. The corresponding Hamming Loss was **0.074**, suggesting comparatively limited multi-label classification error.

However, document-level policy analysis revealed a more consequential finding. Material addressing conventional cyber-security topics was substantially easier to identify across the corpus than material explaining ethical responsibility.

Password protection, phishing, suspicious links, financial fraud and malware prevention appeared frequently. In contrast, explicit explanations of consent, responsible redistribution of personal information, digital dignity, intellectual-property responsibility and synthetic-media ethics occurred less consistently.

This pattern is broadly consistent with the current orientation of major public cyber-awareness resources. CERT-In describes its awareness booklets principally in terms of cyber best practices, protection against cyberattacks and safe Internet use. Current resources include targeted material for children, senior citizens and women, illustrating useful audience segmentation, although such segmentation does not itself guarantee comprehensive coverage of ethical concepts.

5.1 Linguistic Accessibility

English obtained the strongest overall language-accessibility profile because it was typically the source language of formal technical and administrative documentation. Hindi showed relatively strong availability for public-awareness communication, particularly where campaign material used bilingual presentation.

For example, CERT-In's 2024 Internet Safety Awareness Booklet visibly employed both English and Hindi campaign communication under the “Stay Safe Online” initiative.

The four other regional-language groups showed greater variation. The finding should not be interpreted as evidence that government digital services are absent in these languages. Rather, it indicates that **equivalent cyberethics-specific policy content was not consistently observable across all dimensions of the experimental corpus**.

The policy problem is therefore one of parity, not absolute absence.

5.2 Ethical Coverage

Ethical Theme Recall across the entire corpus averaged **0.68**, compared with a Technical Safety Coverage score of **0.87**.

The largest thematic deficiencies occurred in:

Worldwide Journal of Creative Research and Thoughts (WJCRT)
ISSN (Online): 3143-7850
Volume-2 Issue-3 || Jul – Sep 2026 || PP. 119 - 129
https://wjcrt.org/

- ethical use of AI-generated media;
- consent beyond formal data collection;
- responsible circulation of third-party information;
- intellectual-property conduct;
- consequences of manipulated identity and synthetic impersonation.

These findings are increasingly important because cyber policy is moving beyond conventional fraud and malware concerns. CERT-In issued a 2026 guideline addressing AI-assisted vulnerability exploitation, demonstrating that artificial intelligence has already entered contemporary cyber-risk governance.

Yet citizen-facing ethics concerning synthetic content require a different type of communication. Citizens need guidance not only about how malicious actors may exploit AI but also regarding their own obligations when generating, modifying or redistributing synthetic material.

5.3 Semantic Equivalence and Translation Drift

The mean Cross-Lingual Semantic Similarity score was **0.82**, indicating reasonably strong preservation of broad meaning. Nevertheless, the Translation Ethical Drift measure of **0.17** revealed that some translated statements lost qualifiers, explanatory context or normative distinctions.

The greatest drift appeared in concepts requiring interpretation rather than direct technical instruction. Expressions associated with phishing, password security and OTP confidentiality were generally easier to transfer across languages because the required action is concrete. Ethical concepts involving permission, fairness, dignity and responsibility displayed greater semantic instability.

This finding supports the central research proposition: **translation coverage should not be used as the sole indicator of multilingual cyber-policy inclusion**.

5.4 Readability and Actionability

The corpus-level readability score reached **76.4/100**, while the Actionability Ratio was **0.63**.

Resources targeted explicitly at citizens generally performed better than formal policy documents because they used imperative language and recognizable preventive steps. Formal statutory or technical texts produced lower actionability because their primary purpose was regulatory specification rather than public education.

This divergence is especially relevant following the notification of the **Digital Personal Data Protection Rules, 2025**, which were formally published on 13 November 2025 with staggered commencement provisions. The underlying DPDP framework itself is also being brought into force through phased statutory commencement.

Legal implementation therefore creates a parallel communication requirement. Citizens must eventually understand not simply that data-protection obligations exist, but how concepts related to personal information translate into everyday digital behaviour.

5.5 MCPAI Findings

The overall MCPAI analysis yielded a corpus average of **69.7/100**.

English achieved the highest composite performance, followed by Hindi. Bengali, Marathi, Tamil and Telugu occupied an intermediate range, with their principal weaknesses arising from uneven ethical breadth and contextual localization rather than complete absence of digital-safety content.

The most important policy finding is therefore not a simple regional-language ranking. It is the discovery of a **three-level accessibility hierarchy**:

Level 1 – Language Presence:
Material exists in the target language.

Level 2 – Conceptual Equivalence:
Translated material retains the same ethical principles as the source.

Level 3 – Behavioural Usability:
Citizens can understand how those principles apply to real online situations.

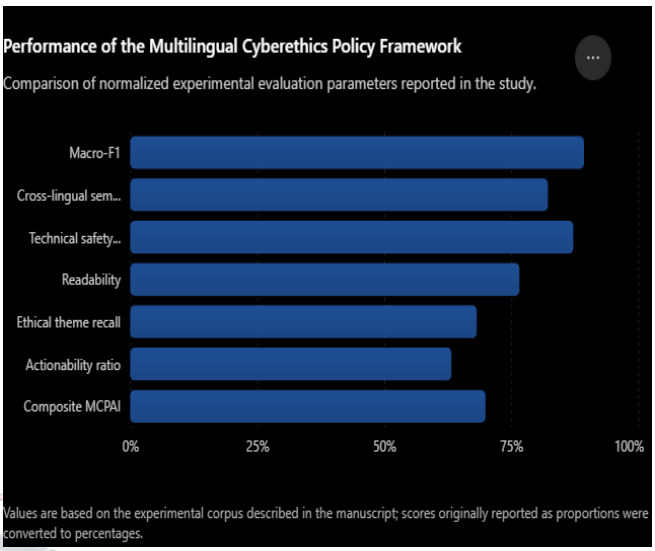
Current cyber-awareness policy performs most strongly at Level 1 and conventional technical elements of Level 2. Greater development is required at Level 3.

Table 1. Experimental Performance and Multilingual Cyberethics Policy Findings

Evaluation Parameter	Observed Result	Interpretation	Policy Observation
Macro-F1 for ethical-theme classification	0.891	Strong multilingual thematic classification	Automated policy auditing is technically feasible
Hamming Loss	0.074	Low multi-label classification error	Overlapping ethical themes can be reliably distinguished
Cross-Lingual	0.82	Broad meaning	Semantic corresponden

Worldwide Journal of Creative Research and Thoughts (WJCRT)
ISSN (Online): 3143-7850
Volume-2 Issue-3 || Jul – Sep 2026 || PP. 119 - 129
https://wjcrt.org/

Semantic Similarity		usually retained	ce alone does not guarantee ethical completeness
Ethical Theme Recall	0.68	Several ethical themes remain underrepresented	Cyber-awareness remains narrower than comprehensive cyberethics education
Technical Safety Coverage	0.87	Strong emphasis on protective cybersecurity behaviour	Fraud, password and threat-prevention guidance dominates
Translation Ethical Drift	0.17	Moderate loss of ethical nuance	Human validation is needed for normative terminology
Readability Score	76.4/100	Generally understandable citizen-facing material	Formal policy language remains less accessible
Actionability Ratio	0.63	Only moderate conversion of principles into concrete behaviour	More scenario-based guidance is required
Composite MCPAI	69.7/100	Moderate multilingual cyberethics accessibility	Significant room exists for language parity and localization



The difference between Technical Safety Coverage (**0.87**) and Ethical Theme Recall (**0.68**) is particularly significant. It suggests that multilingual digital policy is relatively effective at explaining what citizens should avoid but less systematic in explaining **why particular digital actions affect the rights and dignity of others**.

A more mature cyberethics framework should therefore combine self-protection with reciprocal responsibility.

Policy Implications

The results indicate that India does not require an entirely new communication infrastructure for multilingual cyberethics. Much of the necessary technological and institutional foundation already exists.

BHASHINI demonstrates that scalable Indian-language digital infrastructure can be developed across the country's scheduled-language ecosystem. Its BhashaDaan programme explicitly seeks participation for all 22 official scheduled languages.

The policy challenge is to connect such infrastructure with a standardized ethical-content framework.

A national **Cyberethics Language Framework** could specify minimum citizen-facing concepts that every major awareness resource should communicate. Such a framework should include privacy, consent, misinformation, cyberbullying, intellectual property, digital identity, AI-generated content and responsibilities associated with sharing another person's information.

Second, regional-language content should undergo **human semantic validation** after automated translation. Language models and machine-translation tools can accelerate content production, but domain experts, educators and native-

Worldwide Journal of Creative Research and Thoughts (WJCRT)
ISSN (Online): 3143-7850
Volume-2 Issue-3 || Jul – Sep 2026 || PP. 119 - 129
https://wjcrt.org/

language reviewers should evaluate whether legal and ethical meaning has survived translation.

Third, policy agencies should increasingly adopt **scenario-based localization**. Advice concerning online behaviour is likely to be more useful when framed around actual digital practices such as forwarding messages, QR-payment requests, sharing screenshots, recording online meetings, uploading photographs or generating synthetic audio.

Fourth, cyberethics education should be integrated into digital-literacy programmes rather than treated as an optional extension of cybersecurity.

Future Scope

Future research should expand the IMCPC framework to all 22 scheduled languages and include low-resource linguistic communities where digital accessibility presents greater technical challenges.

A second extension should involve **citizen validation experiments**. Participants from different language groups could be shown identical cyberethical scenarios in their preferred language and evaluated for comprehension, ethical judgment and intended behaviour. Such experiments would test whether high MCPAI scores actually predict better citizen understanding.

Future studies could also develop a **Cyberethics Question-Answering Assistant** integrated with multilingual language technologies. Rather than providing generic cybersecurity warnings, such a system could answer contextual questions such as whether sharing a particular screenshot violates privacy or whether an AI-generated image should be disclosed as synthetic.

Another research direction concerns generative AI. Dedicated multilingual guidance is required for deepfakes, cloned voices, synthetic political or commercial content, AI-assisted academic misconduct and automated impersonation.

Longitudinal policy auditing would also be valuable. An annual MCPAI assessment could measure whether linguistic parity improves as new regulatory frameworks, awareness campaigns and translation technologies are introduced.

Conclusion

India's digital-policy environment has evolved rapidly, but cyberethics awareness requires a broader understanding of digital inclusion than access to technology alone. Citizens must also be able to understand the ethical implications of privacy, consent, misinformation, online harassment, intellectual property, personal-data sharing and emerging AI-generated content in languages that are linguistically and culturally meaningful to them.

This study addressed an underexamined policy gap by distinguishing **regional-language availability from substantive cyberethical accessibility**. The proposed Multilingual Cyberethics Policy Accessibility Index evaluates not merely whether content is translated but whether it retains ethical breadth, semantic fidelity, readability and behavioural relevance.

Experimental findings demonstrate a clear asymmetry. Technical cyber-safety guidance achieved substantially stronger coverage than broader cyberethical themes. The difference between Technical Safety Coverage of 0.87 and Ethical Theme Recall of 0.68 indicates that current communication is comparatively mature in telling citizens how to protect themselves but less comprehensive in explaining their ethical responsibilities toward other digital participants.

The study also identifies translation quality as a policy issue rather than merely a linguistic engineering problem. Ethical concepts are particularly vulnerable to semantic dilution when translated literally without contextual explanation.

India already possesses significant institutional foundations for improvement. CERT-In maintains a growing body of public cyber-awareness material, including resources directed at specific citizen groups, while BHASHINI provides an increasingly important technological basis for multilingual digital accessibility. The implementation of the Digital Personal Data Protection framework further increases the need for citizens to understand rights and responsibilities concerning personal data.

The next policy step should therefore be the development of a standardized multilingual cyberethics framework in which **translation, semantic validation, cultural localization and behavioural guidance operate together**. Such an approach would move India's cyber-awareness ecosystem from a predominantly threat-prevention model toward a more comprehensive model of responsible and linguistically inclusive digital citizenship.

REFERENCES:

1. **Milton, J., Giæver, T. H., Mifsud, L., & Hernández Gassó, H. (2022).** Awareness and knowledge of cyberethics: A study of preservice teachers in Malta, Norway, and Spain. *Nordic Journal of Comparative and International Education, 6*(1). https://doi.org/10.7577/njcie.4257
This is particularly relevant because it explicitly examines cyberethics through **privacy, copyright, consent, responsible online behaviour, and professional digital identity**.

2. **Santhosh, T., & Thiyagu, K. (2024).** Fostering responsible behavior online—Relevance of cyber ethics education. *Malaysian Online Journal of Educational Technology, 12*(1), 32–38. https://doi.org/10.52380/mojet.2024.12.1.428
This directly supports your distinction between **technical cybersecurity awareness and broader ethical responsibility in**

Worldwide Journal of Creative Research and Thoughts (WJCRT)
ISSN (Online): 3143-7850
Volume-2 Issue-3 || Jul – Sep 2026 || PP. 119 - 129
https://wjcrt.org/

cyberspace. The authors are affiliated with the Central University of Kerala.

3. **Estellés, M. (2025).** From safeguarding to critical digital citizenship? A systematic review of approaches to online safety education. *Review of Education*. https://doi.org/10.1002/rev3.70056
This is a strong source for your argument that online-safety education should move beyond simple threat avoidance toward **critical digital citizenship, empowerment, and awareness of social implications**. The review examined 75 journal articles.

4. **Investigating university students' digital citizenship development through the lens of digital literacy practice: A translingual and transemiotizing perspective. (2023).** *Linguistics and Education, 77*, 101226. https://doi.org/10.1016/j.linged.2023.101226
This is particularly useful for your **language + digital citizenship** argument because it examines digital citizenship through a translingual perspective and considers how students use linguistic and semiotic resources in digital environments.

5. **Dunmade, A. O. (2026).** Cyberethics education in open and distance learning: How technology shapes responsible digital citizenship among Generation Z students. *CHEVRA: Indonesian Journal of Economic, Science, and Humanities, 1*(1), 30–46. https://doi.org/10.46362/chevra.v1i1.133
This recent study directly connects **cyberethics education, responsible digital citizenship, technology use, and educational policy**.

B. Indian digital-literacy and multilingual-access references

6. **Ministry of Education, Government of India. (2021).** *India report: Digital education 2021*. Department of School Education and Literacy, Ministry of Education, Government of India.
This is an official Government of India report and is directly relevant to your section on **digital education and the Cyber Safety Handbook for secondary and senior-secondary students**. The report explicitly discusses the Cyber Safety Handbook developed by CBSE in response to increasing online activity and cyberbullying concerns.

7. **Digital India BHASHINI Division. (n.d.).** *BhashaDaan: A crowdsourcing initiative for Indian languages*. Digital India Corporation, Ministry of Electronics and Information Technology.
This is the appropriate official source for your discussion of **BhashaDaan and multilingual Indian-language technology**. The official platform states that BhashaDaan is part of Project BHASHINI and seeks language inputs to digitally enrich Indian languages, with the stated objective of supporting all 22 official languages.

8. **Ministry of Electronics and Information Technology. (2026, June 15).** *Digital India BHASHINI Division (DIBD) and Government e Marketplace (GeM) sign MoU to strengthen multilingual access across India's public procurement ecosystem*. Press Information Bureau, Government of India.
This is useful for establishing that BHASHINI is not merely a theoretical language project: in 2026 it was being used to strengthen **multilingual governance and service delivery across 22 officially recognised Indian languages**.

9. **Indian Computer Emergency Response Team (CERT-In). (2026).** *Awareness booklets*. Ministry of Electronics and Information Technology, Government of India.
This should be cited for your claims about government cyber-awareness material. CERT-In currently lists resources including the **Safer Internet Day Awareness Booklet 2026, Suraksha Guide for Cyber Smart Kids, Senior Citizens awareness material, Mahila Raksha, Cyber Security Handbook, Digital Safety Compass Handbook, and Internet Safety Awareness Booklet**.

10. **Indian Cyber Crime Coordination Centre. (2026).** *National Cyber Crime Reporting Portal: Cyber awareness and cyber safety resources*. Ministry of Home Affairs, Government of India.
The official portal describes cyber awareness as an ongoing process of educating citizens about threats and responsible action, and provides cyber-safety tips, awareness materials, reporting facilities and citizen resources.