# **360-Degree Feedback Software for Government of India News Monitoring Using AI/ML**

**Introduction**

Innovative AI/ML technologies are increasingly recognised as enablers of sustainable development. In this project, we propose an AI-driven 360° feedback system to help the Press Information Bureau (PIB) of India monitor public sentiment and media coverage of government initiatives. The PIB is the Government of India’s nodal agency for disseminating policy information to print and electronic media, issuing press releases in English, Hindi, Urdu, and other Indian languages, and relaying media feedback back to the government. Given India’s linguistic diversity, manually tracking regional news (including code-mixed content) is difficult. An automated system will improve the timeliness and coverage of feedback. In particular, this project aligns with SDG 9 (Industry, Innovation, and Infrastructure) by harnessing advanced AI/ML infrastructures, SDG 10 (Reduced Inequalities) by providing inclusive coverage of regional languages, and SDG 16 (Peace, Justice, and Strong Institutions) by enhancing transparent and accountable governance.

## **Problem Statement**

The PIB currently disseminates information (via press releases) to over 8,400 newspapers and media outlets nationwide, translating releases into 13 languages (e.g. Punjabi, Gujarati, Kannada, Odia, Assamese, Manipuri). It needs effective, real-time feedback from regional media reactions. **A major challenge** is that relevant news appears in hundreds of regional news sites, e‑papers, and broadcast videos, often in local scripts or mixed languages. To address this, the system should automatically monitor and analyse media content end-to-end. Specifically, the system must:

* **Crawl and ingest news:** Automatically crawl ~200 selected regional news websites (print and online) to collect published news stories in multiple Indian languages, including code-mixed and mixed-script content.
* **Departmental categorisation:** Use multi-label classification to tag each story with the relevant government department or ministry, based on predefined tags or learned categories.
* **Tonality classification:** Classify each story’s tone toward the Government of India as favourable (positive), neutral, or unfavourable (negative).
* **Real-time alerts:** Immediately notify the concerned PIB officer (e.g. via SMS or Android push notification) whenever a negative or critical news story is detected for their department.
* **E-paper OCR and clipping:** Scan and OCR regional newspaper e-paper editions, identify and extract news clippings relevant to government policies, and generate standardised digital clippings that include newspaper name, edition, page number, and date.
* **E-paper analysis:** Classify the extracted clippings by department and sentiment, and present them in a sortable/filterable dashboard (filter by tonality, media source, etc.).
* **Video monitoring:** Crawl YouTube channels of major news broadcasters, parse news bulletin transcripts using closed captions or audio-to-text, identify segments about government policies, and categorise those video segments by department and tone. Provide immediate alerts for any negative coverage.
* **Operational alignment:** Integrate with existing PIB workflows. The system must have role-based dashboards, audit trails, and explainable AI outputs (key phrases or saliency to justify decisions). It must meet public-sector requirements for privacy, security, and ethics (SDG 16) and satisfy performance SLAs for latency, throughput, availability, and reliability.

Implementing this will support equitable information access (addressing SDG 10) and build smart infrastructure (SDG 9). For example, robust code-mixed processing ensures that feedback includes underrepresented language communities (SDG 10), and real-time alerts help institutions respond quickly (SDG 16).

## **Objective**

The primary objective is to **build an AI/ML-driven, real-time 360° media and public sentiment feedback system** for the Government of India (PIB). This system must process multilingual content (including code-mixed and regional scripts), perform department-wise news categorization, detect sentiment (positive/negative/neutral), and generate actionable alerts and reports. In doing so, it fosters technological innovation (SDG 9) and enhances government transparency (SDG 16) while reducing information disparities among regions (SDG 10). The system will also include social media sentiment as needed, but with a focus on formal news channels and e-papers. The goal is to integrate this feedback loop into official PIB workflows so that negative coverage can be addressed promptly and policies can be refined based on measured public reaction.

## 

## **Related Work and Literature Review**

This section surveys existing systems and research relevant to automated media monitoring, multilingual NLP, and feedback-driven reporting.

* **Practical media-monitoring systems:** Several government and private prototypes demonstrate the feasibility of automated news crawling, e-paper OCR, and video transcript analysis. For example, some media intelligence platforms already crawl news websites and e-papers for keywords and perform basic sentiment tagging. However, many of these systems rely on **rule-based methods or classical ML models** (e.g. regex filtering, simple lexicons) and are limited in language scope. They often lack deep benchmarking and struggle with Indian regional languages and code-mixed text. Few commercial tools fully integrate web feeds, OCR-ed newspapers, and broadcast videos into a single pipeline. Moreover, most do not emphasize AI explainability or fairness metrics needed for public-sector accountability.
* **Transformer-based multilingual NLP:** Recent advances in deep learning have produced powerful multilingual language models. Models like mBERT, IndicBERT, and XLM-RoBERTa have been shown to outperform classical approaches on Indian-language sentiment and classification tasks. For instance, IndicBERT (a BERT variant trained on 12 Indian languages) and XLM-R achieve higher accuracy on noisy, short texts (like headlines or snippets) than TF-IDF/SVM pipelines. Researchers have found that **multilingual transformers excel** at code-switching content. Variants incorporating lexicons or n-grams (e.g. hybrid *LeBERT*-style models) further boost performance for brief, mixed-language text. These transformer models can serve as powerful baselines for categorising and toning stories across India’s diverse media landscape.
* **Classical ML baselines:** Even with transformers available, classical ML pipelines remain useful benchmarks. Techniques like TF-IDF combined with feature selection (e.g. ant-colony optimization for picking n-grams) and classifiers (SVM, Naive Bayes) often yield decent performance, especially when data is scarce. Such methods are more transparent but typically do not generalize as well to multiple languages or domains. They also lack built-in mechanisms for real-time integration or model explainability. In practice, a hybrid approach (using classical models for fast initial filtering and transformers for final decisions) could be considered.
* **360° feedback and continuous monitoring:** The concept of 360-degree feedback comes from organizational performance systems, emphasizing continuous, holistic input and bias reduction. In the context of media monitoring, this means not only pulling data from web, print, and broadcast sources, but also providing constant, actionable insights to stakeholders. Key principles include regular feedback loops, fairness calibration, and dashboard-driven reporting. Some HR and management studies suggest designing such systems with role-based interfaces and audit trails, which parallels our need for secure, user-specific dashboards.

The literature shows that while partial solutions exist (news crawlers, OCR tools, sentiment analyzers), **few integrated systems** handle all source types (web, e-paper, video) in multiple Indian languages with explainable AI. This gap motivates our end-to-end approach.

## **Key Findings from Literature**

* **Multilingual & code-mixed processing is essential:** Studies consistently show that transformer models (mBERT, IndicBERT, XLM-R) outperform traditional methods on Indian language sentiment and classification tasks. Robust preprocessing (language identification, transliteration, normalisation) further improves results. Handling code-switching is critical for real-world media text (headlines often mix English with local script). This aligns with SDG 10 (Reduced Inequalities), since effective processing of regional languages ensures marginalised voices are not overlooked.
* **Short-text optimisation improves accuracy:** News headlines and social snippets are very short and noisy. Research indicates that lexicon-augmented transformers or n-gram aware models yield higher F1 scores than plain transformers or bag-of-words on such texts. Incorporating domain-specific sentiment lexicons or using models pre-trained on news corpora can boost tonality classification performance.
* **Classical baselines provide a strong reference:** TF-IDF with feature selection and SVM can be tuned to produce reasonable baselines for news classification. These classical models are fast and interpretable, offering transparency for key features. However, they generally underperform on highly multilingual/mixed data and are not end-to-end in real time. They remain useful for ablation studies and low-resource settings.
* **Integration and fairness remain under-addressed:** Few existing systems unify web crawling, OCR'd e-papers, and video ASR/CC parsing into a single pipeline with real-time alerts. Moreover, most research lacks emphasis on model calibration and subgroup fairness. A government-grade solution must monitor performance across languages, regions, and publishers to avoid biases. For instance, underperformance on a particular language could inadvertently skew feedback (counter to SDG 10). Ensuring fairness and transparency is as crucial as raw accuracy.
* **Evaluation and compliance gaps:** There is a shortage of publicly available, multilingual datasets in the government news domain, making benchmarking difficult. Reports often omit metrics like calibration error or per-language confusion matrices. Additionally, few papers discuss operational KPIs such as ingestion-to-alert latency or system uptime. Privacy, security, and ethical considerations are also rarely addressed in the literature, but are mandatory for public-sector deployment (supporting SDG 16 on accountable institutions).

These findings indicate that while AI can greatly enhance media monitoring, careful engineering and governance are needed to meet real-world requirements.

## **Implications for the Proposed System**

Based on the above review and the project goals, the following design principles and actions are recommended:

* **Bias-aware modeling:** Adopt techniques such as balanced training sampling, adversarial training, or post-hoc calibration to mitigate biases across languages and regions. Regularly audit model performance on underrepresented languages. This supports SDG 10 (Reduced Inequalities) by striving for equitable accuracy across groups.
* **End-to-end real-time pipeline:** Integrate web crawling, API feeds, OCR for e-papers, and ASR/closed-caption parsing into one data pipeline. Use streaming processing frameworks (e.g. Apache Kafka/Flink) to meet latency and throughput SLAs. Ensure each processed piece of content (story, clipping, video segment) is tagged and stored in a database. Enable monitoring (logs, metrics) and alerting systems (e.g. Prometheus, Grafana) to maintain high availability. This advanced infrastructure supports SDG 9 by building resilient, innovative IT systems.
* **Explainability and human oversight:** Provide interpretable outputs for each AI decision (e.g. highlighting key phrases or sentiment-bearing words). Develop an interface for human reviewers to override or flag system decisions. Use active learning pipelines so the system learns from corrections over time. These measures increase trust and accountability, in line with SDG 16 (building strong institutions).
* **Rigorous evaluation:** Track comprehensive metrics including macro F1 score, AUROC, and expected calibration error for each language and department. Maintain confusion matrices to identify common misclassifications. Set acceptable disparity thresholds (e.g. ≤5% difference in error rates) between language subgroups. Also monitor operational KPIs: alert latency (ingestion→notification), processing throughput, system uptime. Report these metrics in dashboards for stakeholders.
* **Compliance by design:** Incorporate privacy and security safeguards from the start. For example, encrypt stored data, follow government IT security policies, and ensure no sensitive personal data is collected. Perform ethical reviews to ensure the system does not inadvertently censor or misrepresent any group. These steps reinforce both SDG 16 (just, accountable institutions) and SDG 10 (by protecting vulnerable populations).

By following these implications, the project will not only achieve its technical goals but also adhere to the high standards required for government projects, contributing to sustainable development objectives.

## **Sustainable Development Goals (SDG) Alignment**

This project contributes to several UN Sustainable Development Goals:

* **SDG 9 – Industry, Innovation and Infrastructure:** Building a novel AI/ML-driven media monitoring system fosters technological innovation and resilient infrastructuresdgs.un.org. The project develops advanced tools (multilingual NLP, real-time pipelines) that exemplify innovative data infrastructure.
* **SDG 10 – Reduced Inequalities:** By processing news in diverse regional languages and scripts, the system promotes inclusive access to information. It ensures that government feedback includes underrepresented linguistic communities, reducing information inequality across India’s regions.
* **SDG 16 – Peace, Justice and Strong Institutions:** The system enhances government transparency and accountability by systematically tracking media sentiment and revealing public reactionssdgs.un.org. Explainable AI and audit trails strengthen trust in institutions. The automated feedback loop helps build an inclusive, responsive information ecosystem.

By aligning with these goals, the capstone project not only solves a practical problem but also advances the university’s emphasis on sustainable, socially beneficial innovation.

## **Conclusion**

This report has outlined a comprehensive plan for an AI/ML-driven, 360° feedback system to monitor Indian media coverage of government initiatives. We have defined the system requirements, reviewed relevant literature and existing tools, and distilled key findings. The proposed solution will integrate cutting-edge multilingual NLP models, real-time data engineering, and fairness-aware design to meet PIB’s needs. Throughout, we have emphasized alignment with SDGs – fostering innovation (SDG 9), reducing inequalities (SDG 10), and strengthening institutions (SDG 16). By adhering to these principles and monitoring a robust set of metrics, the project aims to deliver a transparent, effective media monitoring platform that improves policy feedback loops in India.

## 

## 

**360-Degree Feedback Software for the Government of India Related News Stories in Regional Media using Artificial Intelligence / Machine Learning**

## **Abstract**

The Press Information Bureau (PIB) serves as the Government of India’s interface with print and electronic media, disseminating information and gathering feedback on government initiatives. Given India’s linguistic diversity and the vast volume of regional media content, manual tracking is challenging and inefficient. This project proposes an AI/ML-driven, real-time, 360° media monitoring and sentiment feedback system capable of crawling regional news websites, processing e-papers via OCR, analyzing YouTube broadcasts through ASR, and classifying content by department and sentiment (positive, neutral, negative). The system will issue real-time alerts for negative coverage, integrate into PIB workflows, and provide dashboards for analysis, ensuring transparency, explainability, and bias-awareness. The solution aligns with UN Sustainable Development Goals (SDGs) by promoting innovation (SDG 9), reducing inequalities through multilingual inclusion (SDG 10), and strengthening institutions (SDG 16).

## **1. Introduction**

The PIB disseminates official information in English, Hindi, and Urdu, later translated into multiple regional languages to reach ~8,400 newspapers and media outlets nationwide. Effective feedback from regional media is vital for governance, but currently requires manual, fragmented processes. Advances in AI/ML enable automated systems capable of multilingual text analysis, sentiment detection, and real-time monitoring, offering opportunities to modernise PIB’s workflows while ensuring inclusivity and transparency.

This project addresses the challenge of creating an end-to-end automated feedback system tailored for India’s diverse linguistic and media environment.

## 

## **2. Problem Statement**

The current feedback system at PIB is limited in scope and scalability due to:

* **Multilingual and code-mixed content** across ~200 media websites, e-papers, and YouTube channels.
* Lack of **real-time detection** of unfavourable coverage.
* **Manual effort** is required for monitoring, classification, and reporting.
* Absence of **auditability, explainability, and fairness calibration** in existing tools.

**Need:** An AI/ML-driven, bias-aware, real-time 360° system to automate crawling, classification, tonality detection, alerts, and reporting while aligning with PIB workflows and ensuring data privacy, security, and ethical compliance.

**SDG Alignment:**

* **SDG 9 (Industry, Innovation, and Infrastructure):** Develops innovative infrastructure for media analysis.
* **SDG 10 (Reduced Inequalities):** Ensures regional languages and underrepresented voices are included.
* **SDG 16 (Peace, Justice, and Strong Institutions):** Strengthens transparency and accountability through explainable AI.

## **3. Objective**

To build an AI/ML-driven, real-time **360° media and sentiment feedback system** for PIB that:

* Crawls ~200 regional websites, e-papers, and YouTube channels.
* Processes multilingual and code-mixed content using OCR/ASR/NLP.
* Classifies stories into government departments.
* Analyses sentiment as positive, neutral, or negative.
* Sends **real-time alerts** for negative/critical coverage.
* Provides dashboards for filtering, visualisation, and reporting.
* Ensures fairness, explainability, security, and operational reliability.

## **4. Survey on Related Works**

### **4.1 Practical Monitoring Systems**

Existing systems show feasibility of automated crawling, OCR, and alerts but are often rule-based, limited in multilingual scope, and lack explainability.

### **4.2 Transformer-Based Multilingual NLP**

Models like **mBERT, IndicBERT, XLM-R** outperform classical ML for multilingual, code-mixed sentiment tasks, especially for short/noisy texts such as headlines.

### **4.3 Classical ML Baselines**

TF-IDF + SVM + metaheuristic feature selection (e.g., Ant Colony Optimization) provide transparent baselines but lack robustness for multilingual, real-time integration.

### **4.4 360° Feedback Principles**

From HR literature: feedback loops, bias reduction, explainability, and actionable dashboards are directly relevant to media monitoring workflows.

## **5. Key Findings**

* **Multilingual handling is essential:** Transformers with preprocessing outperform classical pipelines.
* **Short-text optimization improves accuracy:** Lexicon-augmented transformers excel on headlines and snippets.
* **Classical baselines are useful references:** Provide benchmarks but not scalable for this problem.
* **Integration gaps remain:** Few systems unify web, e-papers, and video.
* **Compliance and fairness are under-addressed:** Critical for government-grade solutions.

## **6. Proposed Methodology**

### **6.1 Bias Handling in Sentiment Analysis**

* **Balanced & Diverse Training Data:** Regularly updated datasets covering all languages/regions.
* **Fairness Audits:** Metrics to detect systemic misclassification across languages/regions.
* **Debiasing Techniques:** Fairness constraints, re-sampling, post-hoc thresholding.
* **Human-in-the-Loop Validation:** Ambiguous/negative stories reviewed by PIB officers.
* **Transparency:** Explainable AI outputs (keywords, saliency maps).
* **Continuous Monitoring:** Detect drift and emerging bias patterns.

### **6.2 System Development Workflow**

**Step 1: Requirements Gathering & Design**

* Stakeholders: PIB officers, news editors, IT admins.
* Scope: ~200 websites, multiple e-papers, and YouTube news channels.
* Architecture: Crawlers → OCR/ASR → NLP → Classification → Alerts → Dashboard.

**Step 2: Data Collection**

* Web Crawling: Automated scripts/APIs for regional media.
* E-paper OCR: Tesseract/Google Vision for scanned text.
* YouTube ASR: Closed-captioning or Whisper/Google STT.

**Step 3: Preprocessing**

* Language detection, transliteration, and normalisation.
* Metadata tagging (newspaper name, date, edition, etc.).

**Step 4: Classification & Sentiment Analysis**

* Department mapping via classifiers.
* Tonality analysis using IndicBERT/XLM-R fine-tuned on Indian news.

**Step 5: Real-Time Notification**

* Negative/urgent stories trigger SMS, email, or app alerts.
* Priority ranking for viral/high-impact coverage.

**Step 6: Dashboard & Visualisation**

* Filter by department, sentiment, time, and edition.
* Charts showing sentiment trends and coverage.
* Access to clippings, scans, and transcripts.

**Step 7: Audit Trail & Reporting**

* Logs for transparency and accountability.
* Reports by department, sentiment, or region.

**Step 8: Deployment & Scaling**

* Cloud infrastructure (AWS/GCP/Azure).
* Multi-language UI for PIB users.
* Secure data pipelines compliant with government IT standards.

**Step 9: Testing & Maintenance**

* Pilot with select departments.
* Iterative model retraining on fresh data.
* Expansion of monitored sources.

### **6.3 End Product (User Experience Flow)**

1. A regional newspaper publishes a critical article.
2. System crawls, OCRs/transcribes, classifies as negative → Dept. X.
3. Alert sent instantly to PIB officer via SMS/app.
4. Officer reviews the story on the dashboard, takes corrective steps.
5. System tracks follow-up stories and trends.

## **7. Implications for the Proposed System**

* **Bias-Aware Modelling** ensures fairness across languages (SDG 10).
* **End-to-End Real-Time Engineering** builds innovative digital infrastructure (SDG 9).
* **Explainability & Human Oversight** strengthen transparency (SDG 16).
* **Rigorous Evaluation Metrics** ensure reliability and accountability.
* **Compliance by Design** upholds ethics, privacy, and security.

## 

## **Proposed Tech Stack**

| **Layer** | **Tools / Frameworks** | **Purpose** |
| --- | --- | --- |
| **Data Collection** | **- Python Web Scrapers (BeautifulSoup, Scrapy) - YouTube Data API** | **Crawl ~200 regional media sites and YouTube channels** |
| **OCR (E-papers)** | **- Tesseract OCR - Google Cloud Vision API** | **Convert scanned e-papers and images into text** |
| **ASR (Videos)** | **- OpenAI Whisper** **- Google Speech-to-Text** | **Speech-to-text for regional language video bulletins** |
| **Preprocessing** | **- Python (NLTK, Indic NLP Library) - FastText embeddings** | **Language detection, transliteration, normalization, tokenization** |
| **NLP Models** | **- IndicBERT - XLM-RoBERTa - mBERT** | **Multilingual department classification and sentiment (positive/neutral/negative)** |
| **Bias & Fairness** | **- IBM AI Fairness 360 Toolkit - Fairlearn** | **Detect and mitigate systemic bias across languages/regions** |
| **Backend Services** | **- FastAPI** | **Serve AI models, APIs, and handle system logic** |
| **Database** | **- MongoDB (for unstructured logs)** | **Store news articles, classifications, metadata, audit trails** |
| **Dashboard & Alerts** | **- ReactJS** **- PowerBI** **- Firebase** | **Visualization dashboards, real-time push notifications, SMS/email alerts** |
| **Deployment & Infra** | **- Docker, Kubernetes - AWS / GCP / Azure Cloud** | **Containerized deployment, scalability, monitoring, high availability** |
| **Monitoring & Logging** | **- Prometheus - Grafana - ELK Stack (Elasticsearch, Logstash, Kibana)** | **System health, latency tracking, and operational monitoring** |
| **Security & Privacy** | **- SSL/TLS - Role-Based Access Control (RBAC) - Audit Logging** | **Ensure compliance with government data and IT security standards** |

## **8. SDG Alignment**

* **SDG 9:** Fosters innovation in multilingual AI/ML infrastructure.
* **SDG 10:** Reduce inequalities by including underrepresented languages.
* **SDG 16:** Builds transparent and accountable institutions.

## **9. Conclusion**

This project proposes an innovative, bias-aware, multilingual, real-time 360° media monitoring system for PIB. By integrating crawling, OCR, ASR, sentiment analysis, notifications, and dashboards, it addresses the challenges of scale, diversity, and transparency in government media monitoring. Its alignment with multiple SDGs ensures societal relevance while providing technical novelty and innovation.

### **📅 Capstone Project Timeline (Aug – Nov 2025)**

**Phase 1: Foundations (Problem Statement & Objectives)**

* **Duration:** Aug 8 – Aug 20
* **Milestone:** Review-1 (CA-01) – Aug 13 (CSE), Aug 20 (Allied CSE)

**Phase 2: Literature & Model Design**

* **Duration:** Aug 21 – Sep 2
* **Milestone:** Review-2 (CA-02) – Sep 3 (CSE), Sep 10 (Allied CSE)

**Phase 3: Core Development & Initial Progress**

* **Duration:** Sep 3 – Sep 22
* **Milestone:** Review-3 (CA-03) – Sep 23 to Sep 26

**Phase 4: Integration & Dashboard Development**

* **Duration:** Sep 27 – Oct 20
* **Milestone:** Review-4 (CA-04) – Oct 28 to Oct 31

**Phase 5: Testing, Refinement & Finalization**

* **Duration:** Oct 21 – Oct 27
* **Milestone:** Final Internal Submission (System + Demo Ready)

**Phase 6: Documentation & Research Paper Preparation**

* **Duration:** Oct 28 – Nov 16
* **Milestone:** Final Report & Research Paper Draft Ready

**Phase 7: Final Viva (Presentation Only)**

* **Duration:** Nov 17 – Nov 21
* **Milestone:** Review-5 (CA-05) – Final Viva & External Evaluation