

Real-Time Fake News Detection: An AI-Powered Web Application with Trust Scores

Abhishek Kumar, Chandra Shekhar, Rohit, Nitesh

Department of Computer Science , Vivekananda Global University,

Jaipur

23TEC2CS291@vgu.ac.in

ABSTRACT

In order to slow down the rapid spread of false information that erodes public confidence and decision-making in digital media contexts, this work proposes a full-stack AI system for real-time categorization of news headlines and articles as authentic or fraudulent. A Python inference service containing transformer-based models (e.g., BERT) and classical (TF IDF + logistic regression) models with optional ensembles, a Node.js/Express REST API, a React-based web interface, and a MongoDB datastore for predictions and user feedback are all integrated into the platform.

The system aims for at least 90% accuracy, provides immediate predictions with an interpretable trust score, and includes a feedback loop to constantly enhance detection performance through recurring dataset refreshes and retraining. In addition to operational metrics like API latency to guarantee interactive responsiveness and user engagement indicators to measure impact, evaluation makes use of Accuracy, Precision, Recall, F1, and AUC ROC. Data minimization, anonymization, and open communication of model constraints to minimize damage from misclassifications are examples of security, privacy, and ethical measures. A production-ready, cloud-deployed web application for credibility assessment is the result of a staged roadmap that addresses requirements, dataset curation and preprocessing, model development, full-stack implementation, deployment, and continuous monitoring for drift and bias.

Keywords: Misinformation, Text Classification, TF IDF, Logistic Regression, BERT, Ensemble Models, Trust Score, Real-Time Inference, AUC ROC, F1 Score, Node.js/Express, React, MongoDB, REST API, Feedback Loop, Model Retraining, Model Drift, Privacy, and Fake News Detection.

I. INTRODUCTION

As audiences depend more and more on digital news sources, misinformation spreads swiftly across online platforms, impacting democratic processes, public health, and social decision-making. The speed and volume of false information are too great for manual fact-checking to handle, necessitating automated, real-time detection that delivers actionable signals at the point of consumption. In order to lessen detrimental propagation, this study operationalizes a workable solution that combines a feedback-driven learning loop, confidence-based trust ratings, and robust NLP categorization with an approachable web interface.

II. Problem statement and goals

The main objective of the system is to provide interpretable trust scores to facilitate decision-making while classifying input text as true or fraudulent with a target accuracy of at least 90%. Incorporating user feedback for ongoing enhancement is a secondary objective, guaranteeing that the model gradually adjusts to new subjects, wording conventions, and hostile strategies. For interactive use, the solution must have minimal latency, allowing for immediate verification prior to content being shared or quoted.

III. Related Context

There is a need for automated screening that can quickly triage content because existing solutions frequently rely on manual verification, which is unable to keep up with digital virality. By incorporating detection at the time of reading and sharing, a web-first strategy lowers friction and generates immediate authenticity signals to enhance media literacy results. In accordance with production constraints, a phased roadmap from requirements to deployment facilitates disciplined delivery and post-deployment monitoring for bias and drift.

IV. System overview

A Python-based model inference service, a Node.js/Express REST API, a React frontend, and a MongoDB database for the persistence of predictions, feedback, and metadata make up the solution. The model service receives frontend requests, validates them, and returns a label and trust score that are shown to the user and saved for analytics and retraining. Curating datasets, preprocessing text, training models, and creating versioned artifacts with controlled rollout and rollback options are all done by an offline training pipeline.

V. Scope Overview

Multimedia analysis, cross-lingual support, and sophisticated explainability are left for further development in version 1, which concentrates on text-only detection, interactive predictions, and a feedback loop. The system needs autoscaling and caching techniques to stay responsive because it assumes normal web-scale usage patterns with sporadic spikes during significant events. Anonymization and access controls are used to protect data, and privacy-sensitive design reduces the amount of user-submitted text that is retained.

VI. Dataset and preprocessing

With deduplication, normalization, and stratified splits by label and subject, the dataset is made up of tagged news items that have been carefully selected from public sources and quality-controlled pipelines. Tokenization and normalization appropriate for both transformer tokenization and classical vectorization (TF IDF) stages are part of text preparation. Future domain adaption tactics are supported by the preservation of metadata like source and timestamp for topic-aware evaluation and drift tracking.

VII. Model design

Through hyperparameter optimization of regularization, learning rates, and class weighting to address any label imbalance, initial training optimizes for Formula One. Ablation studies are used to examine feature and component contributions, while cross-validation and topic-aware splits are used to evaluate generalization and lessen domain bias.

VIII. Training Methodology

Through hyperparameter optimization of regularization, learning rates, and class weighting to address any label imbalance, training optimizes for Formula One. Ablation studies are used to examine feature and component contributions, while cross-validation and topic-aware splits are used to evaluate generalization and lessen domain bias. To facilitate safe canary deployments, auditing, and reproducibility, artifacts are versioned with dataset snapshots and evaluation metrics.

IX. Evaluation Metrics

Accuracy, Precision, Recall, F1, and AUC ROC are used to quantify performance; when there is a significant class imbalance, AUC PR is taken into account. Throughput under load, API and model delay distributions, and user engagement as measured by feedback volumes and correction rates are examples of operational metrics. In order to diagnose failure mechanisms and adjust decision thresholds for desirable trade-offs between false positives and false negatives, confusion matrices are utilized.

X. Metric Formulas

The formula for precision is $\text{Precision} = \frac{TP}{TP+FP}$. To measure the accuracy of positive predictions, use the formula $\text{precision} = \frac{TP}{TP+FP}$. The formula for recall is $\text{Recall} = \frac{TP}{TP+FN}$. To measure the coverage of real positives, use $\text{Recall} = \frac{TP}{TP+FN}$. The formula $F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$ is used to define F1. To balance precision and recall, use the formula $F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$.

XI. Architecture and component

Input text is gathered by the frontend (React), which also shows predictions and trust scores and offers a clear feedback option for incorrect classifications. Validation, rate limitations if necessary, request tracing, and message delivery to the Python model service are all enforced by the backend (Node.js/Express). The model service logs inference metadata for tracking and retraining triggers, allows model versioning, and hosts one or more models behind a reliable API.

XII. API design

GET /api/v1/health for service metadata, POST /api/v1/predict for text categorization, and POST /api/v1/feedback for corrective labels and remarks are examples of core endpoints. Label, confidence, and latency are among the responses that reveal prediction outcome and service performance in a single call. The main inference path can remain accessible while authentication and rate limiting are introduced for protected endpoints or to prevent abuse.

XIII. Database Schema

To facilitate analytics and audits, the predictions collection keeps track of text, predicted label, confidence, model version, and timestamps. To facilitate human-in-the-loop data quality operations, the feedback collection keeps track of text, predicted label, user-provided adjustments, and processing status. Dataset metadata facilitates governance, refresh cycles, and compliance adherence by tracking sources, licenses, and label quality indicators.

XIV. Implementation details

With its obvious input affordances, loading states, and error messaging, the frontend places a strong emphasis on accessibility and responsive design to lower friction. The backend maintains observability across services and streamlines debugging by using standardized error semantics and structured logging. To adjust performance and latency trade-offs, the model service offers batching, warm-start techniques, and thresholds that may be changed.

XV. MLOps and deployment

Before staged rollouts, CI/CD pipelines do unit, integration, and end-to-end testing while creating container images for frontend, backend, and model services. With the use of golden datasets and automated checks, model versioning with canary releases enables safe experimentation and rapid regression rollback. System SLOs and model health are tracked via monitoring dashboards, which initiate retraining or rollback when drift or performance decline is identified.

XVI. Security, privacy, and ethics

By retaining only the information required for analytics and improvement and encrypting both in transit and at rest, data reduction lowers risk. User-submitted text is protected by anonymization and access constraints, and consent notices outline how input helps update the model. To reduce the harm caused by incorrect classifications and to encourage careful, informed use, limits should be communicated clearly.

XVII. Risks and Mitigations

Topic-aware evaluation, periodic dataset refreshes, and a variety of sources are used to address bias and data quality. Cross-source testing and threshold tuning to balance error types in line with use-case priorities help reduce the risks of generalization. Autoscaling, health checks, and incident runbooks including model and infrastructure rollback procedures all help to lower operational risks.

XVIII. Limitations

In addition to being restricted to monolingual text, Version 1 lacks sophisticated interpretability beyond trust scores and is unable to evaluate audio, video, or graphics. The intricacy of ensembles and transformer variations is limited by latency limitations, necessitating careful balances between responsiveness and precision. Without regular feedback ingestion, validation, and retraining schedules, label noise and changing narratives might impair performance.

XIX. Future work

Richer explainability interfaces, interaction with mobile apps or browser extensions, and multilingual and cross-lingual models are among the planned features. Expanding multimodality to handle music and graphics would improve practical utility and cover a wider range of disinformation types. Collaborations with community dashboards and fact-checkers may enhance the measurement of societal impact and the quality of labels.

XX. Conclusion

This study describes a production-oriented method for detecting fake news in real time that incorporates trust scores, a feedback loop for ongoing development, a straightforward online interface, and strong natural language processing models. Delivering high accuracy and low latency while preserving privacy, openness, and operational resilience is the goal of the architecture, assessment strategy, and MLOps procedures. The system is positioned to lessen the propagation of false information and to change as narratives do by coordinating scope, deployment, and governance with practical limitations.

XXI.**References**

- Diwate, A., Aditeya, N., “Food Waste Management in Restaurants,” IOSR-JHSS, Bengaluru case study, 2023.
- Gurjar, P., “Study of Food Waste from Restaurant and Wedding Ceremonies of Indore City and Its Management,” IRJMETS, 2020

XXII.**Literature review: global state of food waste**

According to the 2024 Food Waste Index, 1.05 billion tonnes of food were wasted in 2022 at retail, food service, and household levels, or around 19% of the food that was available to customers. This shows how waste persists in all economies and how urgent prevention is. Assumptions that household waste is concentrated only in wealthier nations are challenged by the fact that the majority of waste (631Mt, 60%) originates in households, followed by food service (290 Mt) and retail (131 Mt). Average household generation is between 79 and 88 kg per capita annually across income groups, with little difference between high-and middle-income settings. Together, food loss and waste account for 8–10% of worldwide greenhouse gas emissions. The 2024 data emphasizes the importance of urban systems and warm climate contexts as correlations, with rural areas showing greater circularity as a result of home composting and animal feeding Together, food loss and waste account for 8–10% of worldwide greenhouse gas emissions. The 2024 data emphasizes the importance of urban systems and warm climate contexts as correlations, with rural areas showing greater circularity as a result of home composting and animal feeding.

VI. Methods: estimation framework

Boundaries and Scope: Determine the bulk quantities of food waste from retail, food service, and residential sources, separating edible and inedible components, and documenting final destinations non accordance with the FLW Standard and UNEP sector recommendations. Utilizing activity data such as population, meals served, and retail sales volumes, the study design combines direct measurement (waste composition analysis and direct weighing), diaries and surveys, business records, and mass balance/material flow analysis. Methods are chosen by sector, normalized, and scaled to national totals. **Uncertainty:** Use qualitative and quantitative uncertainty assessment in accordance with FLW Standard Chapter 9, taking into account propagation when aggregating across sectors; report method selection, sampling frames, scaling factors, and confidence ratings.

V. Core equations

Total number of households: $F W$

$H H = w H H \cdot P$ $F W H H = w H H \cdot P$, where $w H H$ $w H H$ is the amount of household food waste per capita (kg/capita/year) and P is the population. The values for $w H H$ $w H H$ are obtained from diary-based measurements or representative waste audits and normalized to yearly rates.

Total for food service: $F W F S = M \cdot w m$ $F W F S = M \cdot w m$, scaled by subsector shares and service formats in accordance with UNEP's sampling and normalization guidelines, where M M is the total number of meals served and $w m$ $w m$ is waste per meal (kg/meal).

Total retail: $F W$

$R T$ Equals $S \cdot r$ $R T F W$

$R T = S \cdot r$ $R T$, where S S is the mass of food handled or sold and r r $R T$ is the measured loss/waste rate, using normalization factors from corporate data and national statistics as well as sector-specific measuring techniques.

System total and Food Waste Index: $F W T O T = F W H H + F W F S + F W R T$, $F W I = F W T O T P$ $F W T O T = F W H H + F W F S + F W R T$, $F W I = P$ $F W T O T$ presented in accordance with UNEP reporting rules as kg/capita/year for cross-

country comparison.

Fraction of food: F_W

$e_d = f_{ed} \cdot F_W T O T F W_{ed} = f_{ed} \cdot F_W T O T$, employing sensitivity analysis where there are gaps in the literature and tenable edible/inedible category partitioning un accordance with categorization guidelines.

Emissions from management pathways are expressed as follows:

$E = \sum p F W p \cdot E F p$ $E = \sum p F W p \cdot E F p$. In accordance with FLW Standard Appendix D, which expresses mass in environmental terms, and EPA's Wasted Food Scale, which uses pathway rankings to prioritize mitigation, $E F p$ $E F p$ is an emission factor.

The total uncertainty is $u(F_W T O T) = u(F_W H H)$

Two plus $u(F_W F S)$

Two plus $u(F_W R T)$

$2 u(F_W T O T) = u(F_W H H)^2 + u(F_W F S)^2$

$2 + u(F_W R T)^2$.

According to FLW Standard guidelines, with sectoral uncertainty calculated from sample, technique accuracy, and scaling error.

VI. Data and indicators

National teams may combine data to create a clear Food Waste Index that complies with SDG 12.3 reporting criteria thanks to UNEP's sector- appropriate measuring techniques, normalization factors, and scaling procedures with worked examples for households, food service, and retail. In addition to recommending modular definitions and variable quantification techniques to accommodate varying data capacities while preserving comparability and credibility, the FLW Standard specifies the minimal reporting elements: timeline, scope, destinations, methodologies, assumptions, and uncertainty. With UNEP offering capacity building and PPP models to support implementation, these tools work together to help nations establish baselines, monitor progress, and include food waste into climate (NDC) and circular economy initiatives.

VII. Results: global baseline and sectoral distribution

The problem's universality is further supported by the 2024 global estimate of 1.052 billion tonnes of food waste at the retail and consumer levels in 2022, which translates to an average of 132 kg/capita/year across all sectors combined, with households consuming an average of 79 kg/capita/year and a slight variation across income groups. Using cautious edibility assumptions, families alone wasted the equivalent of nearly one billion meals every day, with household waste totaling 631 Mt, food service 290 Mt, and retail 131 Mt. This shows significant preventative potential in domestic settings. Although national and subnational data coverage has increased significantly since 2021, according to UNEP data, more national baselines—particularly outside of households—are desperately needed for reliable trend tracking beyond 2030.

VIII. Table: sectoral waste in 2022 Environmental and social impacts

Prevention directly alleviates upstream resource constraints and downstream methane hazards, particularly from landfills and sewers, as food loss and waste collectively contribute to around 8–10% of global greenhouse gas emissions and occupy a significant portion of agricultural area, according to UNEP estimates. According to EPA's revised hierarchy, the best ways to protect the environment and promote circularity are through prevention, donation, and upcycling; landfilling, incineration, and disposing of food down the drain are the least desirable and disproportionately methane-intensive methods for waste streams that are high in food. The humanitarian gap is still very much present, with 783 million people experiencing acute food insecurity, according to UNEP. For this reason, prevention-first programs should prioritize strategies for fair access and donations.

IX. Management pathways and prioritization

Anaerobic digestion with beneficial use of digestate, compost, anaerobic digestion with disposal of digestate, apply to land, landfill, incinerate, and send down the drain are the pathways that are ranked from most to least preferred by the EPA's Wasted Food Scale. Each tier represents lifecycle environmental outcomes and circular potential. In order to provide practitioners with a clearer ordering when developing integrated systems that align with local infrastructure, materials, and policy restrictions, the scale specifically updates the 1990s Food Recovery Hierarchy utilizing 2023 pathway assessments. The greatest climatic and resource benefits per unit of reduced waste are obtained by avoiding disposal and giving priority to human use cascades (prevention, donation, and upcycling) because food decomposes quickly and contributes to methane emissions in landfills and sewers.

[1] Sector	[2] Mass (Mt)	[3] Per-capita indicator
[4] Households	[5] 631	[6] 79 kg/capita/year
[7] Food service	[8] 290	[9] 36 kg/capita/year (global avg)
[10] Retail	[11] 131	[12] 17 kg/capita/year (global avg)
[13] Total	[14] 1,052	[15] 132 kg/capita/year

X. Policy instruments and PPPs

UNEP identifies multi-stakeholder coalitions, the Australian Food Pact, the UK Courtauld Commitment, and other public-private partnerships (PPPs) as successful examples for coordinating measurement, voluntary reduction objectives, redistribution logistics, and consumer engagement at the national level. Through PPP governance, governments can impose landfill restrictions, standardize date labeling, offer liability protections and tax incentives for donations, and require or promote transparent reporting. They can also invest in organics recycling and edible redistribution infrastructure. The need to scale citizen-facing preventative programs, dynamic pricing, and cold-chain improvements where climate and infrastructure conditions correlate with higher per-capita waste is reflected in the rising integration with NDCs and local action plans.

XI. Case examples

Through sectoral targets, animal-feed diversion, mandatory reporting for major generators, and national awareness campaigns, Japan has achieved a 31% overall decrease since 2008, proving the synergy between policy, measurement, and behavior at scale. Under the Courtauld Commitment, the UK cut manufacturing waste by 34% and retail waste by 26%. This was made possible by consumer activation campaigns like "Love Food, Hate Waste," sectoral guidelines, and redistribution working groups. The Netherlands demonstrates the reproducibility of coordinated PPPs in various governance contexts by reporting consumer and retail reductions through a nationwide campaign and voluntary agreements involving government, university, and commercial partners.

XII. Discussion: challenges and uncertainties

The FLW Standard's modular scope definitions and uncertainty reporting requirements address the main

estimation challenges, which include inconsistent edible/inedible classifications by culture and category, incomplete business-sector coverage, and the requirement to scale subnational data to national totals impartially. Although countries can go from modeled estimates to reliable national measuring systems with the aid of UNEP's Level 1–3 framework, more reliable retail and food service data continue to be a barrier to accurate national FWI tracking and cross-country comparisons. Cities become crucial areas for measuring and implementing solutions, necessitating focused investments in organics recovery, redistribution networks, and behavior modification initiatives catered to urban patterns of consumption, storage, and consumption.

XIII. Recommendations and roadmap

Create national baselines by implementing UNEP's sector-specific protocols to assess retail, food service, and households with clear normalization and scaling factors, as well as by adopting the FLW Standard for scope, method selection, and uncertainty.

Put prevention first: Utilize the Wasted Food Scale to plan activities that prioritize waste prevention, followed by donation and upcycling, animal feed, beneficial anaerobic digestion, and composting, with disposal pathways as a last resort.

Give PPPs official status: Using UNEP PPP models and exemplars for national adaptation, create multi-stakeholder agreements with common goals, standardized reporting, redistribution logistics, consumer campaigns, and ongoing learning.

Align finance and policy: Invest in cold chains and city organics systems, standardize date labeling, require or promote reporting, incorporate food waste into climate and biodiversity plans, and incentivize donation and prevention through liability protections and tax benefits.

XIV. Conclusion

Countries and towns can effectively quantify, prioritize, and control food waste toward SDG 12.3 by using the EPA's prevention-first Wasted Food Scale in conjunction with credible estimation based on the FLW Standard and UNEP's Food Waste Index advice. Due to the present global baseline of 1.05 billion tonnes wasted, mostly by homes, and the 8–10% GHG contribution from food loss and waste, prevention, donation, and circular valuation must be scaled up immediately by policy, PPPs, and infrastructure that is specifically targeted at urban systems.

XV. References

Food Waste Index Report 2024, United Nations Environment Programme: methodology, global estimates, sector guidelines, PPP models, and integration with NDCs and SDG 12.3.

Key conclusions of the UNEP Food Waste Index 2024 In brief: 631/290/131 Mt by sector; 79 kg/capita household average; 1.05 billion tonnes overall; 19% of consumer-available food lost; 8-10% GHG from loss and waste; one billion meals wasted daily in households.

The scope, definitions, destinations, quantification techniques, normalization/scaling, uncertainty, and reporting requirements of the Food Loss and Waste Accounting and Reporting Standard (FLW Standard), version 1.0.

Environmental justification for deprioritizing landfill, incinerator, and down-the-drain pathways; revised hierarchy and pathway descriptions for the US EPA Wasted Food Scale that prioritize prevention, donation, and upcycling.