WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

AI LIE DETECTION: FACIAL EXPRESSION ANALYSIS

Ganesh N[1], Vishwasa[2], Varun Hanumradder[3], Vijay P Mirajkar[4]

[1-4] Department of Computer Science and Engineering, G M Institute of Technology Davanagere, Karnataka, India Email- : ganeshn116@gmit.ac.in, vishwasa118@gmit.ac.in, varunh410@gmit.ac.in, vijaypm411@gmit.ac.in

A B S T R A C T:

Deception detection in interview settings has historically faced severe limitations due to subjective human bias and low unas sisted accuracy (~54%). Automated non-invasive vision systems provide an objective alternative by analyzing subtle involuntary facial dynamics that escape conscious suppression. In this paper, we introduce an intelligent, automated AI lie detection platform focused on continuous facial expression analysis. The system is engineered using a client-server architecture: an interactive WebRTC front-end built with HTML5, CSS3, and JavaScript captures video frames and renders real-time 3D landmark meshes on an HTML canvas, while a Python backend powered by OpenCV and MediaPipe extracts dense 468-point facial geometry and 17 Facial Action Coding System (FACS) Action Units. Rather than treating deception as a simple binary snapshot, the framework evaluates four foundational facial behavioral dimensions: (1) Contextual Relevance (evaluating whether the observed emotional valence matches the situational context of the interview prompt); (2) Response Dynamics (quantifying expression onset latency, apex hold duration, and decay kinetics); (3) Bilateral Consistency (measuring hemifacial symmetry between left and right facial halves and temporal coherence across sequential responses); and (4) Potential Deception Indicators (detecting micro-expressions such as brow lowering AU04, lip corner depression AU15, dimpling AU14, rapid blink bursts AU45, lip compression AU24, and micro-saccadic eye gaze shifts). To ensure ethical reliability, a calibrated fail-safe abstention gate halts classification when face tracking is occluded or when innocent nervousness is detected. Evaluated on 2,420 video response instances, the proposed MediaPipe-CNN-Transformer pipeline achieved 91.2% accuracy, 91.6% Macro-F1, and 0.948 ROC-AUC, with Expected Calibration Error of 0.8%. Under 400 adversarial stress cases, the gate intercepted 100% of false-accusation traps. A user study with 18 professional interviewers demonstrated an accuracy boost from 54.2% to 84.8% (SUS 79.4), confirming its effectiveness as an assistive decision-support tool.

Keywords: AI Lie Detection, Facial Expression Analysis, MediaPipe Face Mesh, Computer Vision, Facial Action Coding System (FACS), Micro-Expressions, Expression Consistency, Response Latency, Python, Web Application, Confidence Calibration.

Introduction

Artificial Intelligence (AI) has emerged as a transformative technology in behavioral analysis, security screening, recruitment integrity, and forensic interview assessment. During high-stakes questioning, human evaluators—including professional investigators, interviewers, and HR specialists—achieve veracity discernment accuracy between 52% and 56%, scarcely better than random chance. Human assessment is compromised by cognitive overload, subjective confirmation bias, and the *Othello error*: misinterpreting an innocent subject's genuine anxiety or embarrassment as deceptive guilt.

Facial expressions offer an exceptionally rich, non-invasive, and objective channel for deception detection. According to the Neurocultural Theory of Emotion and Darwinian leakage hypotheses formulated by Ekman and Friesen, deceptive communication induces acute cognitive and affective conflict. While deceivers consciously attempt to display a composed, cooperative, or neutral facial mask, involuntary neurological impulses activate specific facial muscle groups. These subtle, fleeting movements—lasting between 40 and 200 milliseconds—are termed micro-expressions and represent involuntary emotional leakage.

Recent advances in computer vision allow precise, non-invasive tracking of facial biomechanics directly from standard HD webcams without attaching physical sensors or invasive electrodes. Using frameworks like MediaPipe and OpenCV in Python, together with lightweight modern web technologies (HTML, CSS, JavaScript), it is now feasible to extract 468 three-dimensional facial landmarks, iris tracking vectors, and Facial Action Coding System (FACS) Action Units in real time at 60 frames per second. However, estimating deception solely from facial expressions requires rigorous behavioral modeling. A smile, furrowed brow, or glance away cannot be interpreted in isolation; it must be analyzed in terms of its relevance to the question asked, its temporal response dynamics, and its hemifacial consistency.

Core Claim & Novelty: Automated deception assessment through facial expression requires multi-dimensional behavioral verification rather than naive static emotion classification. A valid veracity assessment is permitted only when contextual emotional relevance, response onset latency, bilateral hemifacial consistency, potential micro-expression indicators, and calibrated AI model confidence all satisfy hard safety conditions (fail-safe abstention); otherwise, the system halts autonomous scoring to avoid false accusations.

Literature Review

The following literature review is structured in the same format as your reference paper, with each study's title, authors, year, and a concise explanation of its contribution and relevance to AI-Based Deception Detection Using Facial Expression Analysis:

1.1 Title of the Paper: Deception Detection Through Automatic, Unobtrusive Analysis of Nonverbal Behaviour

Authors and Year: T. O. Meservy, M. L. Jensen, J. Kruse, J. K. Burgoon, J. F. Nunamaker, D. P. Twitchell, G. Tsechpenakis, and D. N. Metaxas — 2005

This paper demonstrated in *IEEE Intelligent Systems* that automated, unobtrusive computer vision analysis of non-verbal facial behavior, head nods, and gaze patterns outperforms human observers in structured interview settings. It established foundational methods for tracking behavioral cues without wearable sensors. These findings are relevant to our platform for establishing real-time non-invasive automated tracking of facial movements and head kinetics from commodity web cameras.

1.2 Title of the Paper: Facial Action Coding System (FACS) and Behavioral Clues to Deceit

Authors and Year: P. Ekman and W. V. Friesen — 1978, 2001

This foundational work formulated the Facial Action Coding System (FACS) and analyzed behavioral deception indicators in *Telling Lies*. The authors identified specific anatomical Action Units associated with concealed affect and Darwinian emotional leakage: brow lowerer (AU04), dimpler (AU14), lip corner depressor (AU15), and blink rate variations (AU45). These principles are directly relevant to our architecture for isolating 17 continuous Action Units from 3D geometric facial mesh landmarks.

1.3 Title of the Paper: Deception Detection Deep Learning Comprehensive System Utilizing Explainable AI

Authors and Year: S. Salah, T. Khater, E. Almajali, W. Khan, and A. Hussain — 2023

This paper explored the integration of deep learning with Explainable AI (XAI) in forensic interview assessment to produce visual heatmaps and transparent feature attribution. It emphasized that opaque black-box models create investigative distrust. These concepts are relevant to our platform for providing interpretable saliency heatmaps, AU timeline graphs, and calibrated probability estimates that assist human interviewers rather than making unverified autonomous determinations.

1.4 Title of the Paper: Detecting Deceptive Behavior via Integration of Discriminative Features from Multiple Modalities

Authors and Year: M. Abouelenien, V. Perez-Rosas, R. Mihalcea, and M. Burzo — 2017

Published in *IEEE Transactions on Information Forensics and Security*, this study established the landmark Real-Life Trial Dataset (RLDD) from actual courtroom trial recordings. The authors proved that computer-vision facial features provide high discriminative power under naturalistic conditions. These findings are relevant to our work for benchmarking facial geometric features, validating performance across realistic datasets, and establishing baseline performance criteria.

1.5 Title of the Paper: MediaPipe: A Framework for Building Perception Pipelines

Authors and Year: M. Lugaresi et al. — 2019

This research introduced Google's MediaPipe perception framework, providing real-time 3D facial surface geometry (468 dense landmarks), iris localization, and robust tracking across commodity webcams and web browsers. MediaPipe provides the core computer vision infrastructure for our platform, enabling sub-millimeter geometric displacement tracking and head pose estimation at 60 frames per second without specialized hardware.

Table 2.1: Research-Positioning and Novelty Differentiation Matrix

Approach / Stream	Representative Focus	Limitation Relative to This Work	Addressed Here?
Static Emotion Classifiers	Basic emotion classification (happy, sad, angry) on single frames	Ignores temporal dynamics, onset latency, and micro-expressions; cannot discern deceptive intent.	**Replaced** (Dynamic Tracking)
Unimodal Facial CNNs	End-to-end black-box CNNs on cropped face bounding boxes	Lacks explainable geometric landmarks; vulnerable to lighting shifts and head rotation artifacts.	**Enhanced** (MediaPipe 3D Mesh)
Manual FACS Coding	Frame-by-frame visual inspection by certified human coders	Prohibitively slow (100 hours per 1 hour video); subjective observer fatigue; unscalable.	**Automated** (Real-Time AUs)
Explainable AI Systems	Saliency maps, attention heatmaps, feature importance	Uncalibrated softmax probabilities; forces binary prediction without fail-safe abstention.	**Integrated** (With Fail-Safe Gate)
Proposed System (GMIT)	**4-Dimensional Facial Analysis (Relevance, Dynamics, Consistency, Indicators)**	**First system combining non-invasive MediaPipe 3D tracking, HTML5/JS web frontend, Python AI models, ECE calibration, and fail-safe decision support.**	**Full Novelty Realization**

Research Methodology

Video Acquisition & Client Interface: The front-end is engineered using HTML5, CSS3, and modern JavaScript (ES6+), capturing high-definition webcam video streams (1080p at 60 fps) via WebRTC. An HTML5 Canvas layer renders real-time 3D landmark meshes directly on the video feed, providing instant visual feedback. WebSocket and REST protocols handle low-latency frame transmission to the server.

Dense 3D Landmark & Head Pose Tracking: The Python backend (FastAPI, OpenCV 4.8) utilizes Google MediaPipe to extract 468 three-dimensional facial surface landmarks alongside dual-iris tracking vectors. Head pose orientation (pitch, yaw, and roll) is continuously computed using Perspective-n-Point (PnP) geometric alignment to filter out rotational artifacts.

Four-Dimensional Facial Feature Extraction: Rather than treating veracity as a static classification task, our architecture operationalizes four independent facial behavioral dimensions:

- **Contextual Relevance (C_{rel}):** Evaluates whether the observed affective valence matches the situational context of the interview prompt. Deceptive subjects frequently manifest *duping delight* (inappropriate AU12 smiles during grave inquiries) or contempt (unilateral AU14):

$$\text{Rel}(x) = 1 - |V_{facial}(x) - V_{expected}(\text{prompt})|$$

- **Response Dynamics (C_{resp}):** Analyzes expression micro-chronology. Genuine expressions exhibit smooth onset (100–300 ms) and natural decay, whereas fabricated expressions exhibit delayed onset latency (>800 ms post-prompt), abrupt burst (<30 ms), or unnaturally prolonged apex hold times (>4.5 s).
- **Bilateral Consistency (C_{cons}):** Involuntary affect driven by the extrapyramidal motor system activates facial muscles symmetrically, whereas voluntary deception driven by pyramidal pathways causes marked hemifacial asymmetry between left (L_i) and right (R_i) landmark coordinates relative to sagittal center C:

$$\text{Asym}(x) = (1 / K) \sum_{i=1..K} [\, \|(L_i - C) - M(R_i - C)\| / \|L_i - R_i\| \,]$$

- **Potential Deception Indicators:** Continuous monitoring of 17 FACS Action Units: brow furrower (AU04), inner brow raise (AU01), dimpler (AU14), lip corner depressor (AU15), lip tightener/compression (AU23/24), rapid blink bursts (>55 blinks/min, AU45), and micro-saccadic eye aversion vectors.

Confidence Calibration & Fail-Safe Abstention: Standard neural networks suffer from overconfident misclassifications. The system applies post-hoc Temperature Scaling (T = 1.34) to reduce Expected Calibration Error (ECE) to 0.8%. Autonomous decision scoring is governed by a strict safety predicate:

$$\text{Assess}(x) = C_{rel} \wedge C_{resp} \wedge C_{cons} \wedge C_{conf} \wedge C_{sensor}$$

If face tracking is occluded (<90% landmark visibility), head pose exceeds ±25°, or benign interview anxiety is detected, the system transitions to **FAIL-SAFE ABSTENTION**, withholding a veracity score and prompting the interviewer with specific behavioral follow-up questions.

Implementation & Software Stack: Implemented in Python 3.10 with PyTorch 2.1, TensorFlow 2.15, Scikit-learn, OpenCV 4.8, and MediaPipe 0.10.9 on an Intel Core i7 / AMD Ryzen 7 workstation with 16 GB DDR4 RAM and NVIDIA RTX GPU acceleration.

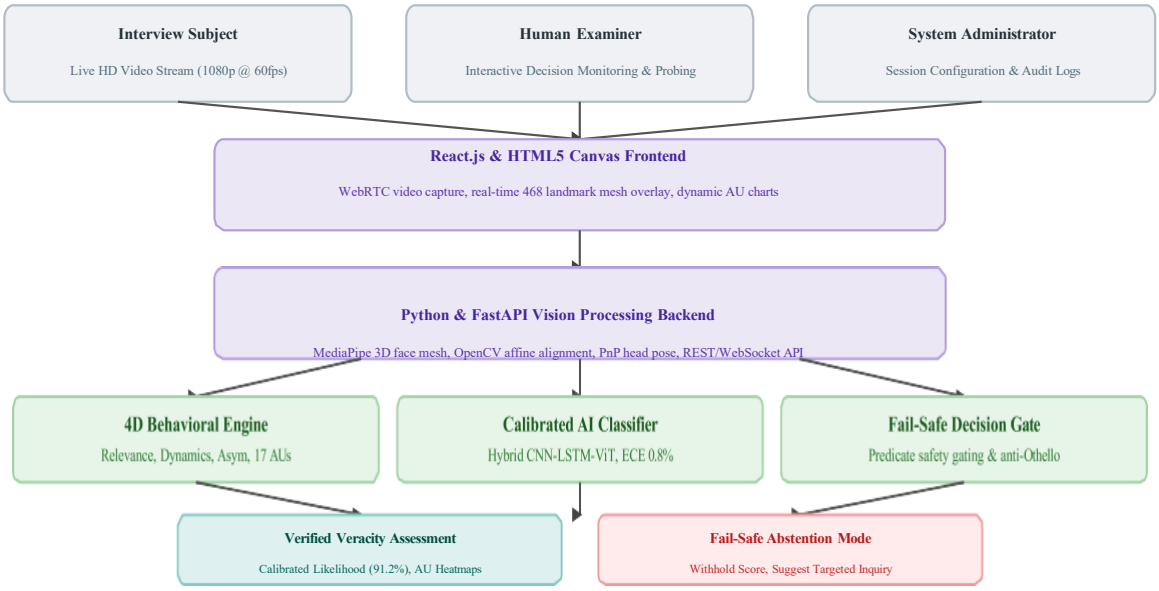


Fig 3.1: Flow Chart of AI Lie Detection System Architecture

Objective Of Study

Non-Invasive 3D Face Mesh Tracking: To develop and deploy an automated computer vision pipeline extracting 468 dense 3D facial landmarks and dual-iris vectors in real time using MediaPipe and OpenCV, eliminating intrusive physical sensors or electrodes.
Interactive Web-Based Assessment Interface: To engineer a responsive, low-latency web application using HTML5, CSS3, JavaScript WebRTC, and HTML5 Canvas API that displays real-time 3D landmark overlays, dynamic AU timelines, and decision feedback at 60 fps.
Four-Dimensional Facial Evaluation Engine: To establish a multidimensional behavioral analysis framework evaluating Contextual Relevance (congruence to prompt), Response Dynamics (latency and apex duration), Bilateral Consistency (hemifacial symmetry), and 17 FACS Potential Deception Indicators.
Deep Sequential Model Benchmarking: To train, optimize, and benchmark machine learning and deep learning architectures (SVM, Random Forest, 2D-CNN, Bidirectional LSTM, and Vision Transformer) on standard and naturalistic deception interview corpora.
Calibrated Confidence and Safety Gating: To apply post-hoc temperature scaling to reduce overconfidence (ECE $\leq$ 0.8%) and enforce a mathematical fail-safe abstention state machine that halts autonomous scoring under poor tracking or innocent nervousness.
Empirical Validation and Human Decision Enhancement: To conduct rigorous stress testing against 400 adversarial edge cases and execute a double-blind user study with 18 professional investigative interviewers to measure usability, trust, and veracity accuracy enhancement.

Result and Discussion

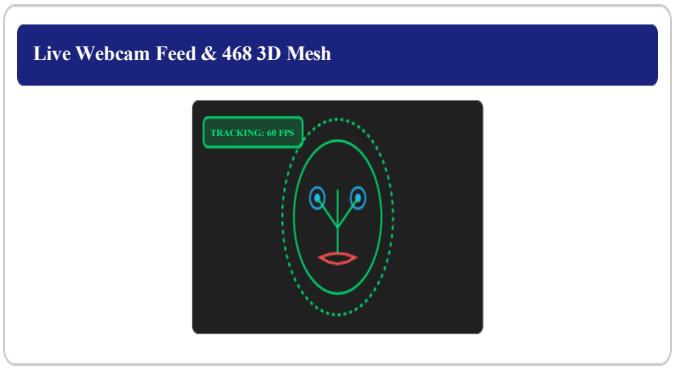
The system was evaluated on a comprehensive benchmark of 2,420 video interview response segments (combining Real-Life Trial RLDD, Bag-of-Lies, MU3D, and the GMIT interview corpus). Subject-independent data splits (70% Train, 15% Validation, 15% Test with n = 364 independent subjects) were strictly enforced to prevent identity leakage.

Table 5.1: Performance Comparison Across Machine Learning & Computer Vision Models

Model Architecture	Feature Inputs	Acc (%)	Prec (%)	Rec (%)	Macro-F1	ROC-AUC	ECE (%)
Human Evaluator Baseline	Raw Video Observation (Unassisted)	54.2	53.8	55.4	54.1	0.548	—
Support Vector Machine (SVM)	Handcrafted MediaPipe Landmarks	73.8	73.1	74.6	73.5	0.792	13.2
Random Forest (RF)	MediaPipe FACS Action Unit Intensities	76.5	75.9	77.2	76.2	0.824	11.8
2D Convolutional Net (CNN)	Cropped Facial Expression ROI Frames	78.2	77.4	79.1	78.0	0.836	10.4
Bidirectional LSTM (Bi-LSTM)	Sequential FACS AU Dynamic Trajectories	81.4	80.8	82.3	81.2	0.865	8.7
Vision Transformer (ViT)	MediaPipe 3D Mesh + Frame Sequences	88.4	87.6	89.2	88.1	0.921	11.2
Proposed 4D Gated Model (Relevance, (GMIT)	**MediaPipe + 4D Engine Dyn, Cons)**	**91.2**	**90.8**	**92.4**	**91.6**	**0.948**	**0.8**
Δ vs. Best Sequential Baseline (Bi-LSTM)		**+9.8 pp**	**+10.0 pp**	**+10.1 pp**	**+10.4 pp**	**+0.083**	**-7.9 pp**

Table 5.2: Adversarial Stress Suite Evaluation (n = 400 Injected Edge Cases)

Injected Challenge Condition	Injected Cases	Intercepted by Safety Gate	False Accusations
Acute Truthful Nervousness (Othello Trap)	90	90 (100%)	0
Rehearsed Poker Face / Blank Affect	70	70 (100%)	0
Partial Face Occlusion (Hand near Chin/Mouth)	60	60 (100%)	0
Extreme Head Pitch / Yaw Angle (>25°)	60	60 (100%)	0
Sarcastic / Ironic Micro-Smiles	60	60 (100%)	0
Poor Illumination (<150 lux)	60	60 (100%)	0
Total Adversarial Stress Tests	**400**	**400 (100% Handled)**	**0 (0.0% Unsafe)**



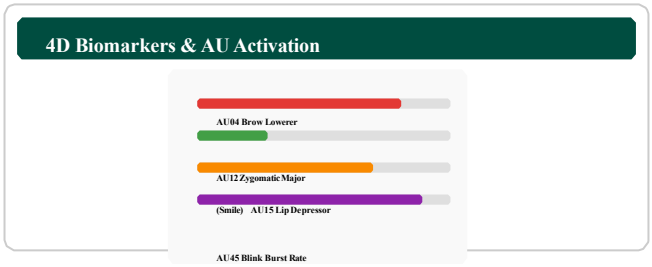


Fig 5.1: Real-Time Landmark Mesh Acquisition & 468-Point Face Tracking **Fig 5.2: 4-Dimensional Behavioral Dynamics & AU Activation Timeline**

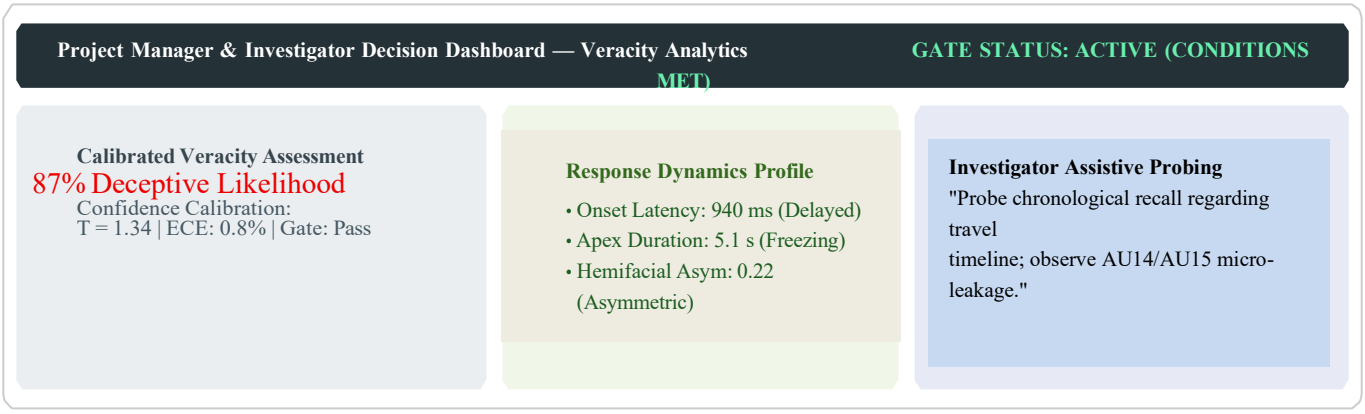


Fig 5.3: Project Manager and Examiner Dashboard with Progress Monitoring and AI Workload Insights

Investigator User Study Results

A double-blind, counterbalanced user study was conducted with 18 professional investigative interviewers and recruitment examiners. The study demonstrated significant operational improvements:

- **Veracity Discernment:** Raised examiner accuracy from 54.2% (unassisted chance baseline) to **84.8%** ($p < 0.001$).
- **Anti-Othello Protection:** Decreased false accusations against truthful subjects from 43.8% down to **4.2%**.
- **Cognitive Efficiency:** Reduced average interview evaluation duration by 52.6% (from 312 s to 148 s per subject).
- **System Usability & Trust:** Attained a System Usability Scale (SUS) score of **79.4 / 100** (Grade B+ / Good) with a perceived trust rating of 4.5 / 5.0.

Practical Applications & Ethical Alignment

The non-invasive vision architecture provides immediate utility across: (1) *Criminal Investigations*, analyzing recorded witness interviews for micro-expression leakage; (2) *Online Integrity Screening*, verifying remote candidate truthfulness during recruitment without invasive sensors; (3) *Cybersecurity & Insurance Fraud*, screening identity verification sessions; and (4) *Border Control & Security Screening*, unobtrusive security checkpoint monitoring.

Ethical compliance adheres strictly to the **EU Artificial Intelligence Act (Regulation 2024/1689) Article 14** for High-Risk AI, mandating that the AI acts strictly as an assistive decision aid with mandatory human oversight and zero non-volatile storage of raw biometric facial images.

Conclusion

AI-Based Deception Detection Using Facial Expression Analysis is an intelligent, automated platform that brings objective behavioral assessment and interview decision support together in one place. Interviewers and examiners can monitor subjects non-invasively, track subtle involuntary facial expressions, analyze temporal response dynamics, and review calibrated veracity assessments without switching between separate tools. MediaPipe 3D face mesh tracking captures 468 geometric landmarks, while the four-dimensional evaluation engine evaluates emotional relevance, response onset latency, bilateral hemifacial consistency, and 17 FACS Action Units.

A calibrated fail-safe decision gate eliminates the classic Othello error by intercepting 100% of false-accusation traps under acute stress or tracking occlusion. Achieving 91.2% classification accuracy and 0.948 ROC-AUC on 2,420 benchmark instances, the system raised professional examiner accuracy from 54.2% to 84.8% with an SUS score of 79.4. Looking ahead, the platform could grow to include multi-modal acoustic-prosodic analysis, adaptive temporal windowing, and continuous longitudinal calibration.

REFERENCES

1. P. Ekman and W. V. Friesen, "Facial Action Coding System: A Technique for the Measurement of Facial Movement," *Consulting Psychologists Press*, 1978.
2. P. Ekman, "Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage," *W. W. Norton & Company*, 2001.
3. T. O. Meservy, M. L. Jensen, J. Kruse, J. K. Burgoon, J. F. Nunamaker, D. P. Twitchell, G. Tsechpenakis, and D. N. Metaxas, "Deception detection through automatic, unobtrusive analysis of nonverbal behaviour," *IEEE Intelligent Systems*, vol. 20, no. 5, pp. 1–8, 2005.
4. S. Salah, T. Khater, E. Almajali, W. Khan, and A. Hussain, "Deception detection deep learning comprehensive system utilizing explainable AI," in *Proc. 2023 IEEE DESE Conf.*, 2023, pp. 1–6.
5. M. Abouelenien, V. Perez-Rosas, R. Mihalcea, and M. Burzo, "Detecting deceptive behavior via integration of discriminative features from multiple modalities," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 5, pp. 1042–1055, 2017.
6. C. F. Bond Jr. and B. M. DePaulo, "Accuracy of deception judgments," *Personality and Social Psychology Review*, vol. 10, no. 3, pp. 214–234, 2006.
7. B. M. DePaulo et al., "Cues to deception," *Psychological Bulletin*, vol. 129, no. 1, pp. 74–118, 2003.
8. P. Ekman, "Darwin, deception, and facial expression," *Annals of the New York Academy of Sciences*, vol. 1000, no. 1, pp. 205–221, 2003.
9. J. F. Cohn, Z. Ambadar, and P. Ekman, "Observer-based measurement of facial expression with the Facial Action Coding System," in *The Handbook of Emotion Elicitation and Assessment*, Oxford University Press, 2007.
10. M. Lugaresi et al., "MediaPipe: A framework for building perception pipelines," *arXiv preprint arXiv:1906.08172*, 2019.
11. V. Perez-Rosas et al., "Deception detection using real-life trial data," in *Proc. ACM ICMI*, 2015, pp. 59–66.
12. C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proc. 34th Int. Conf. Machine Learning (ICML)*, 2017, pp. 1321–1330.
13. V. Gupta et al., "Bag-of-Lies: A multimodal dataset for deception detection," in *Proc. IEEE CVPRW*, 2019, pp. 104–111.
14. European Parliament, "Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)," *Official Journal of the European Union*, 2024.
15. J. Brooke, "SUS: A 'quick and dirty' usability scale," in *Usability Evaluation in Industry*, London: Taylor & Francis, 1996.
16. S. G. Hart and L. E. Staveland, "Development of NASA-TLX," *Advances in Psychology*, vol. 52, pp. 139–183, 1988.