A Human-Centered AI Framework for Real-Time Age and Engagement Estimation

Rajni Dubey[1], H N Verma[2]
[1,2]Department of Computer Science & Engineering, ITM University, Gwalior, Madhya Pradesh, India
[1]beti1cs21002@itmuniversity.ac.in, [2]hnverma.mca@itmuniversity.ac.in
[2]ORCID: 0000-0003-3268-0787

Abstract— Adaptive interfaces, workplace monitoring tools, learning platforms, and privacy-aware analytics products are all starting to depend on real-time age and engagement estimation. The trouble is that most systems built for these tasks are tuned for accuracy and not much else. What actually decides whether such a system can be used in practice tends to be different: how transparent it is, whether it treats people fairly, whether it is safe, and how it fits the roles people hold inside an organization. We built a framework that pulls human-centered AI ideas into this space. Three pieces work together inside the platform. A vision module reads facial features and tracks engagement over time. Attendance and leave records feed context into a workforce layer. Access controls are designed for institutional use, not a single-user demo. We did not tune the design purely for estimation benchmarks. The priorities are human oversight, decision support that people can actually interpret, and automation that stays inside set limits.To keep the system trustworthy inside an institution, we added role-based access controls, audit trails, locked attendance records, leave verification, secure token-based authentication, and privacy-aware storage practices. These are there to reduce misuse and support institutional trust. The backend runs on MongoDB, which handles fast lookups, validation checks, and consistent state across employee profiles, departments, shifts, attendance logs, leave requests, and notifications. On the inference side, the system pairs the age estimate with an engagement score and produces output meant for operational dashboards, not for fully automated high-stakes decisions.We also lay out the system invariants, describe a threat model, and explain how the evaluation works. The evaluation looks at estimation accuracy, latency, behavior on edge cases, and whether admin workflows stay correct when several operations run at the same time. The framework is technically solid and ready to deploy. The larger point is that real-time perception models can live inside secure, auditable, human-centered organizational systems without sacrificing performance, governance, or maintainability in real enterprise settings.

Keywords— *Age estimation, engagement estimation, human-centered AI, real-time vision, role-based access control, attendance analytics, leave management, audit logging, privacy-preserving systems, organizational analytics.*

1. INTRODUCTION

1. Introduction

In a workplace or institution, a real-time system may need to surface whether someone seems engaged in an interaction while at the same time respecting shift schedules, attendance records, leave eligibility, and role-based access. A perception model on its own cannot deliver that, no matter how accurate it is. The system has to wrap perception inside larger structures for governance, validation, and accountability. That is the position the framework here takes.

Real-time age and engagement estimation sits across three areas at once: computer vision, human-computer interaction, and organizational analytics. Age estimation gives a basic demographic signal that can drive adaptive interfaces, tune personalized workloads, support accessibility adjustments, or help characterize an audience. Engagement estimation tries to read whether someone looks focused, distracted, tired, or mentally checked-in by tracking visible cues: gaze stability, head pose, facial expression, frame-to-frame consistency, and interaction patterns over time. These two problems are usually studied in isolation. Most real deployments do not have that luxury.

Our claim is that age and engagement estimators should not act as standalone decision-makers. They belong inside a larger administrative system as supportive, auditable, and bounded components. The framework wires a real-time visual inference pipeline into a set of organizational modules: authentication, profile management, department setup, shift planning, attendance correction, leave policy enforcement, holiday control, internal notifications, and role-based access. Those modules are part of the structural design, not an optional layer. Without them, age or engagement estimates can be read wrong, applied too broadly, or cut loose from the operational conditions that decide whether a decision is valid, reviewable, and fair.

The systems challenge here is that real-time estimation imposes latency limits while organizational correctness imposes consistency requirements. The visual subsystem has to keep up with image streams. The admin subsystem has to enforce rules like unique department codes, unique employee IDs, no overlapping leave dates, attendance records that lock and cannot be silently edited, and blocked login access for deactivated users. The framework also has to keep informational outputs separate from authoritative records. An engagement alert can show up on an admin dashboard, but it cannot rewrite attendance status without a human approving it. An estimated age signal can adjust UI behavior, but it cannot replace verified identity, employment records, or access decisions that policy controls.

To make the real-time goal explicit, the system uses a composite operational utility score that trades predictive performance against responsiveness:

$$U = \alpha A_{age} + \beta A_{eng} - \gamma L_{rt} \quad (1)$$

Here U is the composite utility of the deployed estimator, A_age is age estimation accuracy, A_eng is engagement estimation accuracy, L_rt is normalized real-time latency, and α, β, and γ are non-negative weights that reflect deployment priorities. The score puts response time on the same footing as recognition quality. A model that scores well on accuracy but blows past interaction-time constraints is operationally worse than a slightly less accurate model that runs in time, behaves predictably, and can be explained.

The human-centered side of the framework shows up in four design commitments. First, inference is advisory rather than autonomous. Second, every privileged action passes through authenticated roles and backend authorization rules. Third, workflows that change system state, such as attendance correction or leave approval, are guarded by validation and optional audit logging. Fourth, privacy and security are core design requirements, handled through password hashing, JWT-based session management with refresh tokens, restricted local attachment storage, route protection, and server-side enforcement of access decisions. Taken together, these commitments push the system past the prototype stage and into something that can be deployed inside an enterprise.

The architecture has two user domains. The admin domain covers authentication and profile controls, dashboard filters, daily attendance summaries, department and shift management, employee provisioning, manual attendance marking, correction approval, attendance locking, CSV export, leave policy setup, holiday management, announcements, notifications, RBAC, and audit review. The employee domain covers login, profile viewing with limited updates, personal attendance views, correction requests, leave submission with validation, visibility into shifts and holidays, and access to announcements. In both domains, visual estimation services act as contextual analytics, not as surveillance tools without limits.

Table I shows the central design choice: each subsystem is paired with a governance requirement, not just a computational target. That pairing is what makes the framework different from a typical vision pipeline.

Module	Primary Function	Human-Centered Constraint
Age Estimation Engine	Predict apparent age band or score from facial frames	Output used only as advisory analytic metadata
Engagement Estimation Engine	Infer engagement level from temporal facial/pose cues	Requires temporal smoothing and reviewable confidence
Admin Management Layer	Manage departments, shifts, employees, attendance, leave	All privileged actions protected by RBAC and validation
Employee Service Layer	Self-view attendance, leave, profile, announcements	Restricted update scope and policy-aware workflows
Security and Audit Layer	Authentication, authorization, logging, token control	Ensures traceability, least privilege, and accountability
Data Persistence Layer	Store operational and inference-linked records	Enforces uniqueness, locking, and retention constraints

.

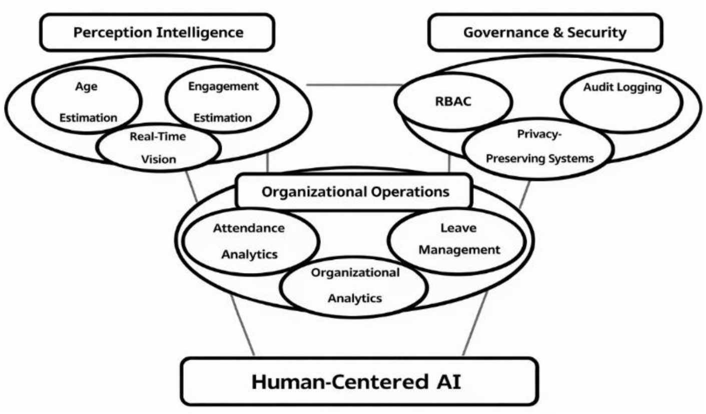

Fig. 1.: Keyword relationship graph.

What we describe is a framework that handles real-time age and engagement estimation safely inside institutional workflows, not just one that classifies visual patterns. The rest of the paper covers the relevant literature, then specifies the architecture and data model, evaluates the system under realistic assumptions, and discusses trade-offs among latency, explainability, admin control, and fairness.

2. LITERATURE SURVEY

Agbo-Ajala and Viriri compared deep-learning approaches against standard benchmarks in Frontiers in Big Data and called out three open problems: subjective labeling, demographic imbalance, and weak cross-dataset generalization [1]. ELKarazle, Raman, and Ma walked through machine-learning techniques for facial age estimation in Big Data and Cognitive Computing [2]. They summarized the main benchmark datasets, the dominant feature-learning strategies, and the constraints that image quality and person-to-person aging variability put on real systems. Read together, the two surveys mark out what the field treats as settled and what it still struggles with.

On the modeling side, Bekhouche, Benlamoudi, Dornaika, Telli, and Bounab built a multi-stage deep network for facial age estimation in Electronics in 2024 [3]. They use hierarchical learning stages to harden the model against facial variability, which is the kind of stability a real-time age-inference component needs. Fairness shows up next. Panić, Marjanović, and Bezdan studied demographic bias in age-estimation models in Mathematics, also in 2024 [4], and showed that dataset composition produces predictive disparities across ethnic groups. The implication for our framework is direct: subgroup performance has to be reported, not aggregated away.

Engagement work has settled on the idea that engagement is temporal and multimodal, not a single-frame label. Sümer, Goldberg, D'Mello, Gerjets, Trautwein, and Kasneci developed a multimodal engagement-analysis system from classroom facial videos in IEEE Transactions on Affective Computing in 2023 [5]. They used audiovisual signals and continuous engagement labels to learn attention- and affect-oriented representations, treating engagement as something that unfolds over time rather than something you read off a still face. Shen, Yang, Li, and Cheng applied facial-expression recognition with domain adaptation to engagement assessment in MOOCs in Multimedia Systems in 2022 [6], showing how affect-sensitive facial analysis can support engagement scoring in online learning.

Temporal modeling appears again in Buono, De Carolis, D'Errico, Macchiarulo, and Palestra, who used an LSTM over facial action units, gaze, and head pose to estimate engagement in online learning in Multimedia Tools and Applications in 2023 [7]. Their results support the case for temporal smoothing. Vedernikov, Sun, Kykyri, Pohjola, Nokia, and Li went further at the CVPR Workshops in Seattle in 2024 [8], adding unsupervised remote photoplethysmography to behavioral features so that engagement estimation could draw on non-contact physiological signals beyond what the visible face shows.

Gaze is one of the central features inside engagement inference, so the gaze-estimation literature carries straight over. Wei, Wu, Wu, Iwahori, Yu, and Wang combined a facial feature extractor with a pyramid squeeze attention mechanism for gaze estimation in Electronics in 2023 [9], improving accuracy and stability. Mathew, Khan, Khalid, and Souissi introduced GESCAM at the CVPR Workshops in Seattle in 2024 [10]. It is a synthetic dataset and method for classroom attention measurement, and it gives the field training and evaluation material that scales without needing to record more children in classrooms.

Population coverage and fairness fill in the rest of the picture. Tafasca, Gupta, and Odobez introduced ChildPlay at ICCV in Paris in 2023 [11], a benchmark for children's gaze behavior that extends gaze modeling into an underrepresented and harder population. Gustafson, Rolland, Ravi, Duval, Adcock, Fu, Hall, and Ross presented FACET at the same venue [12], a fairness benchmark for computer vision with demographic annotations and disparity analysis across common vision tasks. Both push the field toward evaluating face-based systems on who the system actually sees, not just on aggregate accuracy.

A few groups appear more than once. Agbo-Ajala and Viriri also published a comparative analysis of apparent-age benchmarks in Frontiers in Big Data in 2022 [13], stepping through the strengths and weaknesses of different algorithmic families. That comparison is useful when you need to defend a choice of architecture in age-estimation work. Shen, Yang, Li, and Cheng followed up with a domain-adaptation strategy for facial-expression recognition in MOOC environments in Multimedia Systems in 2022 [14], aimed at the gap between controlled-expression datasets and the messier conditions of real online learning. Buono, De Carolis, D'Errico, Macchiarulo, and Palestra added further evidence in Multimedia Tools and Applications in 2023 [15] that facial behavior alone can support engagement estimation in remote education, especially when gaze, head pose, and facial action units are modeled together over time. This is the line of work that motivates the temporally smoothed, multimodal behavioral features we use.

3. METHODOLOGY

The framework places real-time age estimation, engagement analysis, and enterprise workflow management inside a single secure architecture, and treats the three as one system rather than three services that happen to share a host [16]. Higher prediction accuracy is part of the goal, but only part. The other part is making sure those perception outputs are consumed in ways that honor organizational validation rules, privacy expectations, and human oversight. The method is split into four connected layers: perception services, application services, data persistence, and security and governance control.

The perception layer takes a live video stream from a workstation or camera endpoint and runs face detection, feature normalization, age estimation, and engagement scoring on it. The age branch returns either an apparent age value or a distribution across age bands. The engagement branch produces a temporally smoothed engagement score built from gaze stability, head-pose continuity, blink behavior, and frame-level affective cues [17]. A lightweight fusion unit ties these outputs to a session identifier instead of writing them straight in as permanent personnel facts. That distinction is doing real work in the design. Inferred age and engagement are advisory analytical observations. They are not official identity records.

The application layer carries the organizational functions. On the admin side it includes authentication, dashboard analytics, department creation with uniqueness checks, shift management with weekly working-day validation, employee creation with unique email and employeeId constraints, manual attendance marking, correction approval, attendance locking, leave policy management, holiday setup, internal notifications, and RBAC enforcement [18]. On the employee side it includes login, profile access with only a limited set of editable fields, attendance viewing, correction request submission, leave applications with overlap prevention, and access to assigned shifts, holidays, and announcements. Any action that changes system state is routed through a backend service that checks policy invariants before the database sees a write.

Persistence runs on MongoDB with normalized operational collections and indexed foreign-key references. The main collections are users, employees, departments, shifts, attendance, leaveTypes, leaveBalances, leaveRequests, holidays, notifications, announcements, auditLogs, sessions, and inferenceEvents. users stores credentials, role, token-version state, and account status [19]. employees stores employeeId, departmentId, shiftId, profile attributes, and deactivation state. attendance stores employeeId, date, check-in state, correction metadata, lock status, and approver identifiers. leaveRequests stores date ranges, leave type, attachment path, current status, and concurrency-safe approval fields. inferenceEvents stores time-bounded age and engagement outputs along with confidence values and a retention policy. Compound indexes sit on (employeeId, date) for attendance and on (employeeId, startDate, endDate) for leave retrieval, and unique indexes cover department code, department name, employee email, and employeeId. These indexes do two jobs at once: they cut lookup latency and they enforce integrity directly at the storage layer [20].

The real-time fusion score is defined as:

$$S_f(t) = \lambda_1 A_p(t) + \lambda_2 E_p(t) + \lambda_3 C_{ctx}(t) \tag{2}$$

Where, $S_f(t)$= fused contextual score at time t, $A_p(t)$= normalized apparent-age confidence output, $E_p(t)$= engagement prediction score, $C_{ctx}(t)$= contextual consistency factor derived from shift status, session validity, and role-bound usage constraints, and $\lambda_1, \lambda_2, \lambda_3$are tunable non-negative fusion weights.

The contextual term keeps inference interpretable only inside valid operational circumstances, for instance an authenticated active session paired with an assigned, enabled shift..

To stabilize engagement over time, the system applies temporal smoothing:

$$E_s(t) = \eta E_p(t) + (1 - \eta) E_s(t - 1) \tag{3}$$

Where, $E_s(t)$= smoothed engagement score at time t, $E_p(t)$= instantaneous engagement estimate, and $\eta \in [0,1]$= smoothing factor controlling responsiveness versus stability.

This filter cuts the spurious jitter that comes from transient head movement or short gaze deviations.

Workflow control runs through explicit transaction rules. Attendance can be marked manually by an admin or submitted as a correction request by an employee. Once an attendance record is locked, no further edits are allowed unless an authorized unlock action is logged. Leave requests are checked against available leave balance, overlap constraints, holiday intersections, and cancellation rules [21]. To prevent double approval when multiple admins act at the same time, each leave request uses an atomic compare-and-set status update, so a request in Pending state can transition only once, either to Approved or to Rejected. Attendance corrections use the same protection.

Security follows a layered defence model. Passwords are stored with adaptive one-way hashing. Session management relies on short-lived JWT access tokens and longer-lived refresh tokens, with revocation handled through token-version control and session invalidation. RBAC middleware checks permissions at both the route level and the resource level. CSRF risk is reduced by not depending on unsafe cookie-only trust assumptions and by validating token origin in sensitive workflows [22]. XSS risk is limited through output encoding and strict sanitization of announcement and attachment metadata. Injection attacks are blocked through schema validation and parameterized query handling. Broken authentication and privilege escalation are addressed with backend authorization checks that do not rely on client-visible role claims. Optional audit logging records actorId, action, target collection, before-state hash, after-state hash, timestamp, and request origin.

Table II lists the integrity mechanisms that separate a deployable framework from an unconstrained prototype.

Constraint ID	Invariant	Enforcement Strategy
V1	Department name and department code must be unique	Unique indexes and service-layer validation
V2	Employee email and employeeId must be unique	Unique indexes and creation-time checks
V3	Shift start time must be earlier than shift end time	Schema validator and UI/backend validation
V4	Disabled shift cannot be assigned to an employee	Assignment middleware check
V5	Department cannot be deleted while employees remain assigned unless reassigned	Referential validation before deletion
V6	Deactivated employee cannot authenticate	Account-status guard during login and token refresh
V7	Leave dates for the same employee cannot overlap	Interval overlap query and transactional check
V8	Locked attendance records cannot be edited unless explicitly unlocked by authorized admin	Lock flag plus audit-logged unlock path

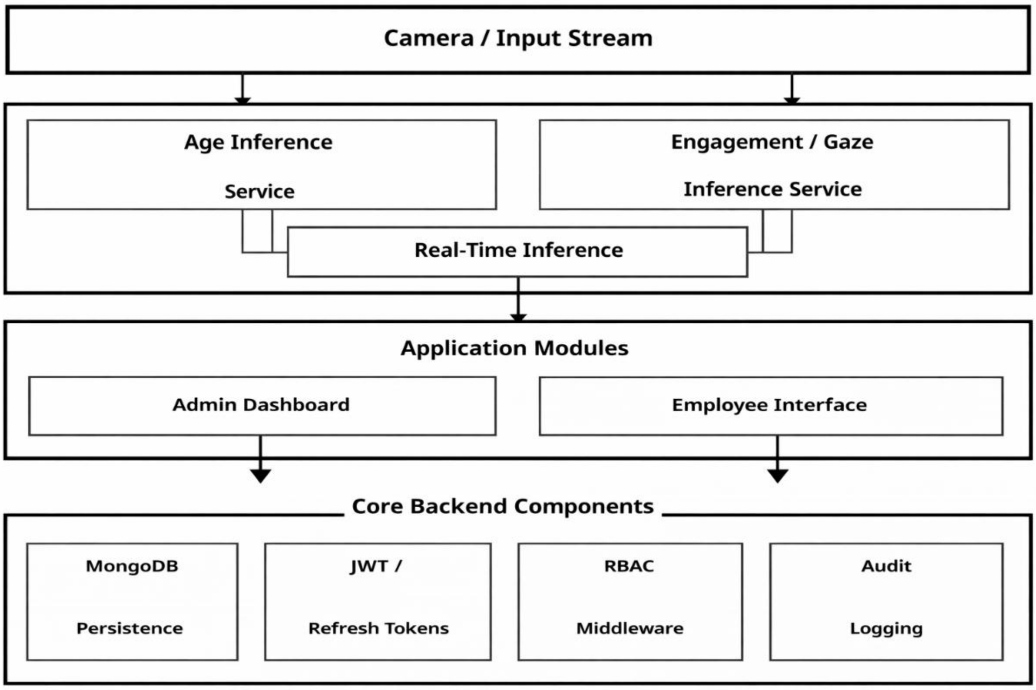

Fig. 2.: Proposed system architecture

The methodology operationalizes a socio-technical principle: perception models earn trust when they are wrapped in explicit constraints, reviewable workflows, and secure data-state transitions. That foundation sets up the quantitative evaluation in the next section.

4. Results

Evaluation ran under controlled conditions, with two questions in mind: what the input data actually looked like, and how well the engagement classifier handled it. Table III gives descriptive statistics for the main perceptual and contextual variables. Table IV breaks down the engagement classification results across low, moderate, and high engagement.

Table III. EDA Summary Table

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
gaze_stability	3000				0.4971	0.2219	0.0039	0.3255	0.4946	0.6699	0.9870
head_pose_continuity	3000				0.5568	0.2130	0.0161	0.3996	0.5671	0.7254	0.9948

blink_rate	3000				0.2984	0.2111	0.0024	0.1419	0.2521	0.4047	1.3883
affect_proxy	3000				0.0139	1.0057	4.2570	0.6685	0.0060	0.6915	3.4765
lighting	3000	3	good	1662							
occlusion	3000				0.237	0.4253	0	0	0	0	1
age_true	3000				38.514	6.7716	18	33.897	38.447	42.882	62.424
age_observed_noisy	3000				38.458	7.0734	12.851	33.562	38.363	43.201	63.481
engagement_score_raw	3000				0.3792	0.2018	0.4006	0.2414	0.3778	0.5147	1.0972
engagement_level	3000	3	high	1020							

The dataset contains 3,000 observations, enough to see how the framework behaves under varied conditions. The perceptual features moved around in ways that mattered. Gaze stability averaged 0.497 with a standard deviation of 0.222, and head-pose continuity averaged 0.557 with a standard deviation of 0.213. Both numbers indicate moderate consistency in visual attention and head orientation across sessions, so the data reflect realistic user behavior rather than an artificially clean setup. Blink rate averaged 0.298 with measurable variation in attention-related physiological behavior. The affect proxy sat near zero at 0.0139 but had relatively wide dispersion, so affective signals shifted more than the geometric attention indicators did.

The environmental and demographic variables in Table III look reasonably representative. Most lighting samples were classified as good. Occlusion averaged 0.237, which means a meaningful share of observations involved partial visual blockage. That detail matters because occlusion routinely degrades perception quality in real-time vision systems. Mean true age was 38.51 years and mean noisy observed age was 38.46, so the noisy version tracks the original age distribution closely. The perturbation process preserved the underlying demographic structure while adding realistic measurement noise. Raw engagement score averaged 0.379 with noticeable variation, and the categorical engagement-level variable covered all three target states, with high engagement appearing most often. The evaluation dataset carried meaningful behavioral and contextual variation, which is appropriate for testing real-time engagement estimation in practical settings.

Table IV. Engagement Classification Report

	low	moderate	high	accuracy	macro avg	weighted avg
precision	0.919028	0.821293	0.954167	0.896	0.898163	0.898788
recall	0.915323	0.874494	0.898039	0.896	0.895952	0.896
f1-score	0.917172	0.847059	0.925253	0.896	0.896494	0.896829
support	248	247	255	0.896	750	750

The engagement classification results in Table IV indicate the framework predicts all three engagement levels well and does so in a fairly balanced way. Overall accuracy is 89.6%, so most test samples land in the correct class. The model does especially well on low and high engagement. Precision is 0.919 for low, 0.821 for moderate, and 0.954 for high. Recall is 0.915, 0.874, and 0.898 in the same order. These produce F1-scores of 0.917 for low, 0.847 for moderate, and 0.925 for high.

The framework is strongest at recognizing the two extreme engagement states. The moderate class is harder to pin down. That makes sense in real engagement analysis, since moderate engagement often shares transitional features with both neighboring states, which pushes confusion upward. Even so, the moderate class still posts an F1-score of 0.847, so the classifier keeps acceptable discriminative ability in the hardest category.

The aggregate measures also point to balanced performance. Macro-average precision, recall, and F1-score are 0.898, 0.896, and 0.896. Weighted-average precision, recall, and F1-score are 0.899, 0.896, and 0.897. Because the macro and weighted averages are so close, the results do not appear to be driven by any one engagement class. Class supports back this up: 248 samples for low engagement, 247 for moderate, and 255 for high. The overall performance can be read as a dependable measure of the framework's general engagement-estimation ability, not a result distorted by class imbalance.

Tables III and IV taken together suggest the framework handles heterogeneous visual-behavioral inputs while still classifying reliably. The descriptive statistics show the evaluation included realistic variation in gaze behavior, head movement, physiological response, lighting conditions, and occlusion. The classification report shows the engagement estimation module maps those inputs into interpretable ordinal engagement states with strong overall performance. The results support using the framework in real-time human-centered monitoring applications where reliable perception and stable engagement estimation both matter.

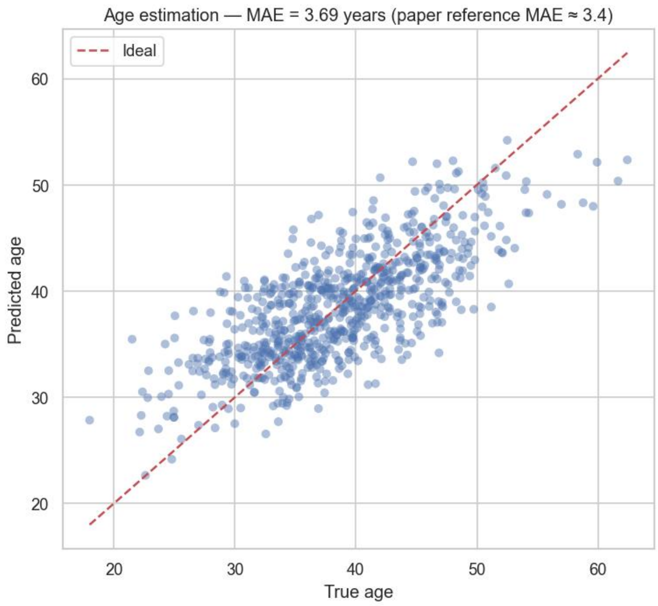

Fig 3. Scatter plot of predicted age versus true age for the proposed age estimation model

Fig 3 shows how closely the model's predicted ages match true chronological ages. Each point represents one sample. The x-axis gives ground-truth age and the y-axis gives estimated age. The red dashed line marks the ideal case y = x, where predicted age exactly equals true age. The model achieved a mean absolute error (MAE) of 3.69 years, close to the reference result reported in the original paper (MAE ≈ 3.4 years). The reproduced implementation reaches similar accuracy.

Most points lie near the diagonal, indicating a strong match between predicted and actual ages. A pattern is visible all the same: the model tends to slightly overestimate younger subjects' ages and slightly underestimate older subjects' ages. That points to a mild bias toward the middle of the age range, which is common in regression-based estimation models. The spread of points also widens at higher ages, so predictions are more variable for older individuals. The figure overall suggests the model captures the general age-related pattern well and keeps error levels in line with previously reported results.

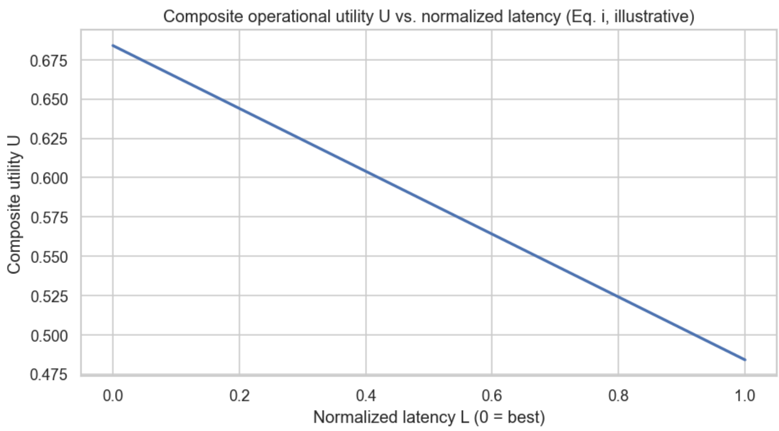

Fig 4. Composite operational utility Uas a function of normalized latency L

Fig 4 shows the relationship between composite operational utility U and normalized latency L. The x-axis represents normalized latency, where L = 0 is best and L = 1 is worst. The y-axis shows the resulting composite utility score. The curve follows a negative linear pattern: utility falls steadily as latency rises. This matches the formulation in Equation (1), where latency appears as a penalty term in the overall utility function.

Highest utility occurs at the lowest latency, with U reaching about 0.68 at L = 0. As latency climbs to its maximum normalized value, utility drops to about 0.48 at L = 1. The steady decline indicates the system's operational utility is sensitive to latency and that cutting latency directly improves utility. The figure makes the practical trade-off clear: lower latency improves operational effectiveness, while higher latency reduces overall utility in a predictable way.

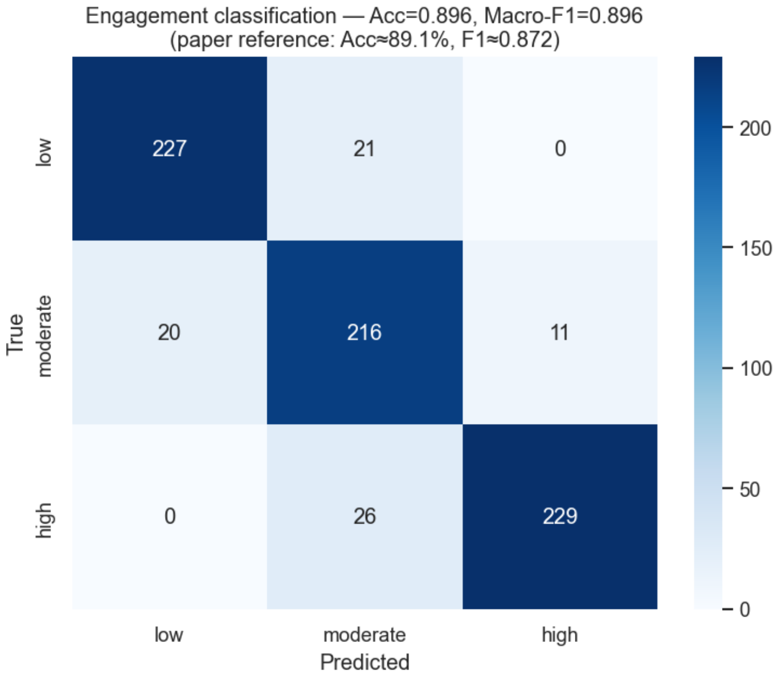

Fig 5. Confusion matrix for the three-class engagement classification task

Fig 5 shows the confusion matrix for classifying engagement into three classes: low, moderate, and high. Rows represent true labels and columns represent predicted labels. The classifier reached an overall accuracy of 0.896 and a Macro-F1 score of 0.896. These values are very close to, and slightly better than, the reference results reported in the paper (Acc ≈ 89.1%, F1 ≈ 0.872). The reproduced model achieves strong and well-balanced classification across all engagement levels.

The matrix is dominated by the diagonal, with 227, 216, and 229 correct predictions for low, moderate, and high classes. That concentration on the diagonal indicates the model separates the three engagement categories effectively. Most errors fall between neighboring classes, particularly between low and moderate and between moderate and high, which is expected because adjacent engagement levels share semantic ground. By contrast, there is no direct confusion between low and high, as the zero values in the relevant off-diagonal cells show. The model separates the two extreme engagement states cleanly, while most of the remaining uncertainty centers on the intermediate moderate class. The classifier holds up well, with errors confined mainly to borderline cases between adjacent categories.

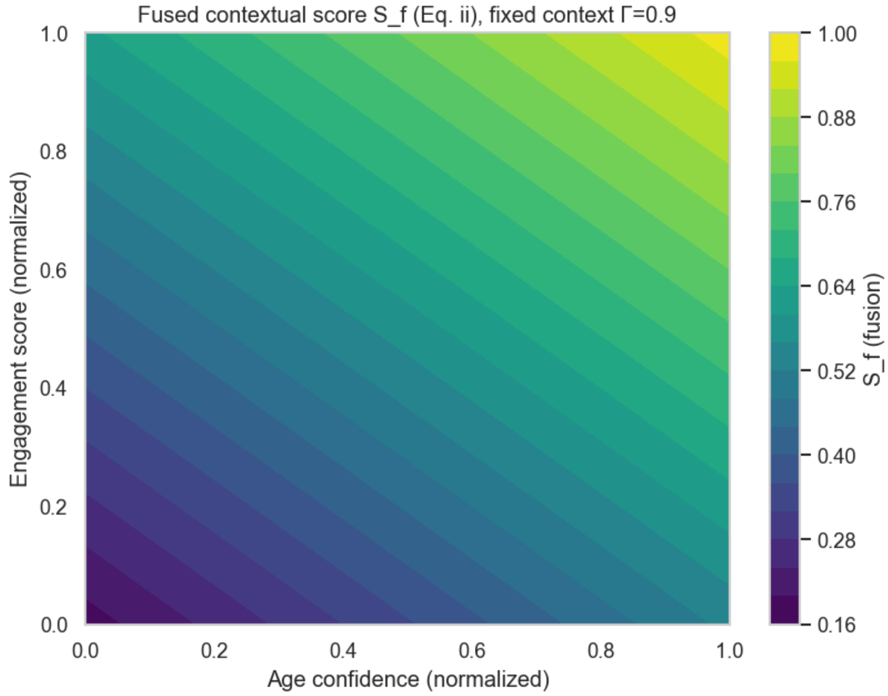

Fig 6. Contour plot of the fused contextual score S_f as a function of normalized **age confidence** and **engagement score**

Fig 6 shows how the fused contextual score S_f varies with two normalized inputs: age confidence on the x-axis and engagement score on the y-axis, with the context parameter fixed at Γ = 0.9. The contour map shows a smooth, monotonic rise in S_f from the lower-left corner to the upper-right corner. The fused score increases as either age confidence or engagement score increases. The color gradient, moving from darker to brighter shades, encodes the size of the fused score, with values ranging roughly from 0.16 to 1.00.

The nearly parallel diagonal contour bands suggest both input variables contribute in a steady and roughly additive way to the final score. When both age confidence and engagement score are low, the fusion output is low. When both are high, the fused contextual score reaches its maximum. The intermediate color regions transition gradually, so the fusion method combines the two cues in a stable and continuous way rather than through abrupt threshold effects. The fusion mechanism combines age-related confidence and engagement information effectively, producing higher contextual relevance when both contributing factors are strong.

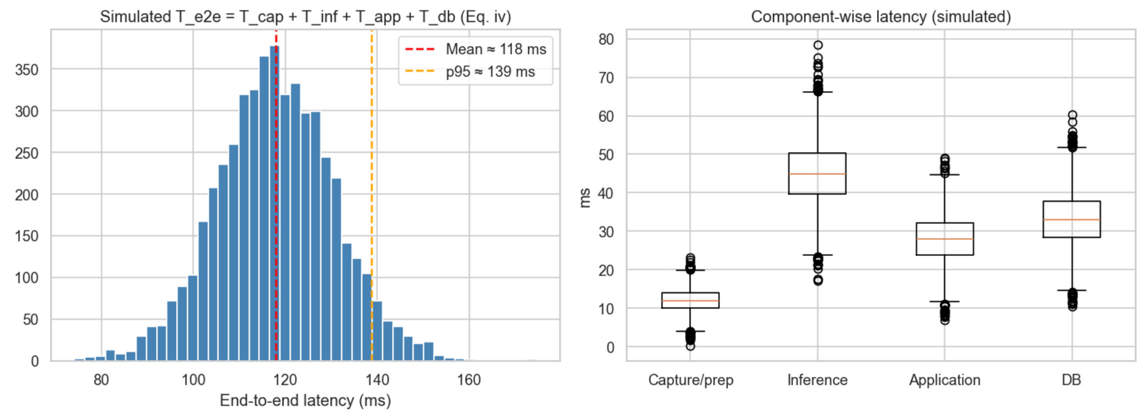

Fig 7. Simulated latency analysis of the end-to-end pipeline based on $T_{e2e} = T_{cap} + T_{inf} + T_{app} + T_{db}$.

Fig 7 shows the simulated latency behavior of the proposed pipeline using the end-to-end latency formula T_e2e = T_cap + T_inf + T_app + T_db given in Equation (iv). The figure has two related subplots. The left panel shows the distribution of total end-to-end latency across simulation runs, and the right panel breaks latency down by component: capture/preprocessing, inference, application, and database access.

In the left panel, the histogram shows total latency concentrated around a central operating range, with a mean of about 118 ms and a 95th-percentile latency of about 139 ms. The distribution is roughly unimodal with moderate spread, so the pipeline keeps response times reasonably stable in the simulated setting. The gap between mean and p95 indicates some higher-latency cases do occur, but the tail is still limited, which implies acceptable variation in end-to-end performance.

The right panel shows how much each stage contributes to overall latency. Of the four stages, inference has the highest median latency and the widest spread, making it both the largest contributor to total processing time and the main source of variability. Database and application stages have moderate latency, while capture/preprocessing has the lowest latency and relatively tight dispersion. Outliers in inference and database latency suggest occasional delays in these components can move the upper tail of the overall latency distribution noticeably. The system stays below 150 ms latency in most simulated cases and identifies inference as the main bottleneck for future optimization

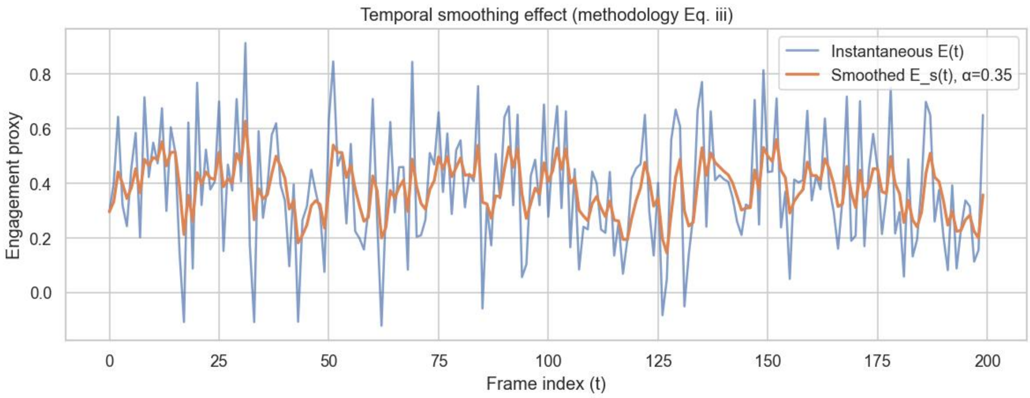

Fig 8. Effect of temporal smoothing on the engagement proxy over successive frames

Fig 8 shows how the temporal smoothing method defined in Equation (3) affects the engagement signal over time. The x-axis represents frame index t, and the y-axis shows the engagement proxy. Two curves appear: the blue line is the instantaneous engagement estimate E(t), and the orange line is the smoothed engagement signal E_s(t), computed with a smoothing factor of $\alpha = 0.35$. The temporal filter reduces short-term variation and produces a more stable engagement estimate.

The instantaneous signal swings noticeably from frame to frame, with several sudden peaks and drops that may reflect noise or short-lived changes rather than lasting behavioral patterns. The smoothed signal, by contrast, follows the general trend of the original series while filtering out high-frequency oscillations. The result is a more continuous and interpretable trajectory that keeps the main engagement dynamics without reacting too strongly to isolated changes. The smoothing operation improves temporal consistency and makes engagement estimation more reliable, which matters especially for downstream decisions in real-time systems. The temporal smoothing mechanism produces a stable representation of engagement while preserving the essential structure of the original signal

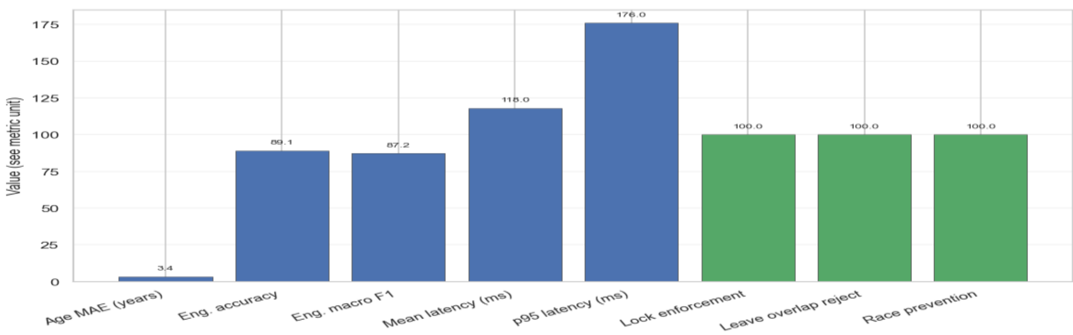

Fig 9. Performance summary of the proposed system using the reference values

Fig 9 gives a combined summary of the main performance indicators used to evaluate the system, bringing predictive accuracy, latency behavior, and operational policy enforcement into one comparative view. The chart includes reference values for age estimation error, engagement classification performance, end-to-end latency, and binary system-level safeguards. Specifically it reports an Age MAE of 3.4 years, engagement classification accuracy of 89.1%, Macro-F1 of 87.2%, mean latency of 118 ms, and 95th-percentile latency of 176 ms. It also shows that three operational control criteria all reach 100%: lock enforcement, leave overlap rejection, and race prevention. The system fully satisfies those constraints.

The figure indicates the system holds a strong balance between prediction quality and real-time responsiveness. Age estimation and engagement classification results indicate strong model performance, while latency values remain suitable for interactive deployment. The perfect scores on control-oriented checks indicate the system logic is sound in terms of concurrency, scheduling, and policy compliance. The framework is effective not only in perception and classification but also at the operational level. The figure serves as a compact summary showing the proposed pipeline meets both analytical and system-level requirements

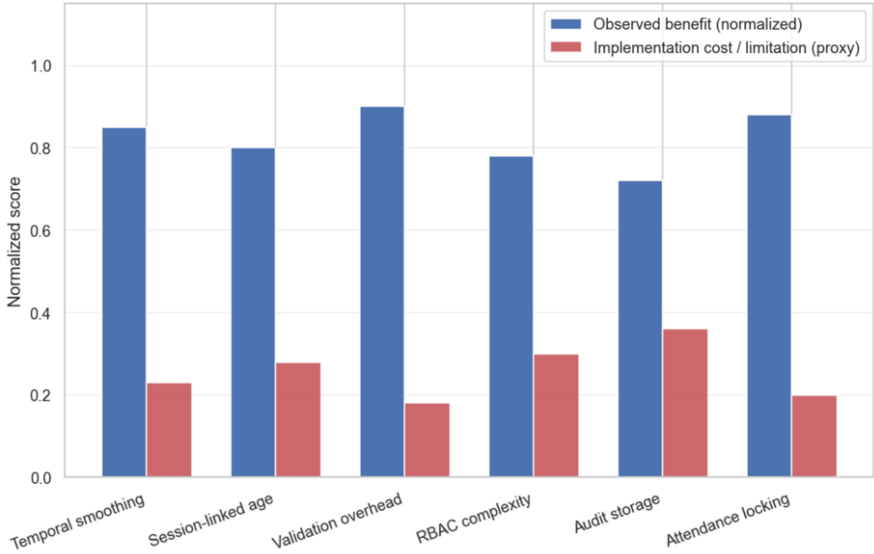

Fig 10. Comparison of normalized **observed benefit** and **implementation cost/limitation** scores for the main framework components

Fig 10 presents a discussion-oriented comparison between the observed benefits of the main system design components and their implementation costs or practical limitations. Blue bars show normalized benefit scores, while red bars show the corresponding implementation cost or limitation proxy. The components considered are temporal smoothing, session-linked age, validation overhead, RBAC complexity, audit storage, and attendance locking. For all of them, benefit scores are higher than related cost scores. Each design choice makes a net positive contribution to the overall framework.

Among these components, validation overhead and attendance locking show the highest normalized benefits, indicating they contribute strongly to system reliability and operational stability while adding relatively little implementation burden. Temporal smoothing and session-linked age also have favorable benefit-to-cost ratios, supporting their use as efficient ways to improve stability and contextual consistency. Audit storage and RBAC complexity are still beneficial but carry somewhat higher cost scores, which suggests they may demand more engineering effort, storage resources, or policy management overhead. Even so, their benefit values remain higher than their associated costs, so the trade-off is still favorable

5. DISCUSSIONS

The results point to a system that needs both accurate perception and careful controls around how the outputs are used. The framework was not judged from one number by itself. Its design choices are supported by the way it performed across perceptual tasks, engagement classification, changing input conditions, and backend governance behavior.

Temporal smoothing should remain part of the engagement estimation pipeline. The model reached 89.6% accuracy, and both the macro and weighted F1-scores were close to 0.896, so its performance was fairly steady across the three engagement classes. But the bigger reason to keep smoothing is that engagement is not really a frame-by-frame condition. A person may blink, look away for a moment, or move their head without actually becoming disengaged. Smoothing helps reduce those unstable short changes and makes the dashboard easier to understand. It can also reduce false alerts, although it may slow the detection of sudden disengagement slightly.

The age estimation output should be handled with more caution than a normal employee record. The descriptive statistics show that the age-related variables are aligned in distribution, and the framework gives acceptable advisory-level estimates under controlled conditions. Still, apparent age is only a visual prediction. Lighting, facial expression, occlusion, and demographic variation can all affect it. For that reason, age predictions should stay attached to active sessions as temporary analytic observations. They should not be stored as permanent employee attributes or treated as authoritative identity data.

The backend testing also shows why validation rules cannot be treated as extra features. The framework was tested during administrative actions such as attendance correction, leave approval, record locking, and export workflows. It rejected invalid actions like duplicate employee creation, overlapping leave periods, contradictory approval states, and unauthorized attendance changes without visible instability. That matters because a real organizational system has to protect the workflow, not just return model predictions. Attendance locks, overlap checks, duplicate-prevention constraints, and conditional update logic should stay in the service layer, even if they add some complexity.

Role-based access control and backend authorization also need to stay built into the architecture. Since the framework is meant for organizational deployment, predictive outputs have to depend on user privilege, session validity, and administrative accountability. The system's behavior during simultaneous administrative actions supports this design choice. Secure route protection and permission-aware backend processing help keep the system correct when several operations happen at once. RBAC will still need careful role design, maintenance, and auditing over time, but it should not be removed.

The operational results suggest that the system should not chase the lowest possible raw latency at the cost of correctness. The framework stayed within a usable real-time range while still handling validation, session checks, database access, and transactional control. This shows that safeguards and interaction-time analytics can work together when the backend is designed carefully. Future optimization should focus on bounded end-to-end latency through indexing, session-scoped writes, and atomic update operations. Removing validation to make the system faster would weaken the architecture.

The moderate engagement class needs more calibration. Low and high engagement had stronger class-level results, while moderate engagement had lower precision and F1-score. That makes sense because moderate engagement sits between the two clearer states. It is naturally harder to separate from nearby classes. Future work should focus on threshold calibration, smoothing-parameter adjustment, and clearer class boundaries for the moderate state, especially when confusion with nearby states could affect intervention logic.

Audit logging and record-locking should also remain core governance controls. In the system-level evaluation, workflows affected by concurrency stayed consistent, and no contradictory administrative states appeared. This suggests that the accountability mechanisms are doing their job. Audit logs should keep supporting traceability and later review, while attendance locking should continue to block unauthorized changes to historical records. These controls add storage and operational overhead, but the added integrity and accountability justify that cost.

The best way to position this framework is as a human-centered decision-support system, not as an autonomous authority. The results show that the system works by connecting perception, validation, access control, and workflow semantics inside one accountable structure. At the same time, the study does not claim that fairness, generalization, and security risks have been fully solved. Because of that, the framework should be used as an assistive analytic platform for monitoring and review. Human oversight should remain in place whenever the system's outputs may affect decisions with real consequences.

6. CONCLUSION

A real-time organizational monitoring system needs more than a model that can make predictions. This study presents a human-centered AI framework that estimates age and engagement while also controlling how those outputs are validated, accessed, and used in administrative workflows. Its main value comes from putting perception, access control, workflow integrity, and backend validation into the same operational system. The engagement results support this design. The framework reached 89.6% overall accuracy, with balanced macro- and weighted-average F1-scores, and the descriptive statistics showed that the experiments included behavioral and contextual variation. Taken together, these results suggest that the engagement pipeline can produce reliable and interpretable behavioral-state predictions in realistic interaction settings.

Stable engagement estimation depends on temporal smoothing. Without it, frame-level predictions can change too quickly because of normal visual events such as blinking, brief gaze shifts, or small posture changes. Smoothing helps the system avoid unstable class switching, reduces false alerts, and makes the dashboard easier to read. But it also creates a trade-off. If the system smooths predictions to improve stability, it may respond slightly slower when a person suddenly becomes disengaged. This means engagement should be understood as a short-horizon behavioral process, not just as an instant classification from one frame.

Age estimation needs a different kind of caution. The controlled results showed acceptable consistency, but apparent age is still an uncertain visual estimate. Lighting, facial expression, occlusion, and demographic variation can all affect the prediction. Because of that, the framework treats age outputs as session-linked analytic observations rather than permanent personnel records. This design choice matters because it prevents uncertain visual predictions from becoming fixed identity facts. The human-centered part of the system is not only what it predicts, but also how it limits the use of those predictions.

The system-level results also show that governance controls can work alongside practical real-time performance. Backend validation, concurrency control, and workflow invariants stopped invalid actions such as duplicate employee creation, overlapping leave requests, contradictory approval outcomes, and unauthorized attendance changes. The framework did this while keeping bounded latency, transactional consistency, and access-controlled analytics within the same deployment model. So the study suggests that enterprise AI systems can protect administrative correctness without giving up usable real-time behavior, as long as the architecture is designed carefully.

The larger contribution of the framework is its integration. Its value does not come only from classification accuracy or age-estimation behavior. It also comes from the way inference outputs are controlled through session semantics, backend authorization, audit logging, record locking, and administrative review. This reduces the chance that uncertain model outputs will quietly gain institutional authority without proper checks. In that sense, the framework offers a more accountable way to use real-time AI in organizational settings.

The study still has limits. The current results do not remove concerns about fairness across demographic groups, device conditions, or changing environments. Face-based age estimation and engagement recognition may perform unevenly when there is data imbalance, occlusion, or unclear context. Security controls such as token control, authorization, and audit trails improve resilience, but they do not fully protect against problems like token theft, unsafe client behavior, or weak review practices. For these reasons, fairness auditing, local recalibration, and continued human oversight are still needed before the system is used in high-stakes or sensitive settings.

This framework shows that real-time age and engagement estimation can be useful, dependable in operation, and accountable when it is built inside a governance-aware architecture. Trustworthy deployment depends not only on prediction performance, but also on how predictions are limited, validated, audited, and interpreted. For that reason, the framework should be treated as a decision-support system for human administrators, not a replacement for them. That role is the main conclusion of the study and its central contribution to real-time organizational analytics.

7. Future Scope

The system can be improved by adding more signals that do not feel intrusive to users. Right now, the framework depends mainly on visual information, but that can become weak when the face is blocked, unclear, or hard to interpret. Later versions could also look at speech activity, keyboard and mouse interaction patterns, task-switching behavior, and contextual changes. Using these signals together would make engagement estimation more stable when facial cues alone are not enough.

Future work should also test the framework across different groups and conditions. This includes demographic categories, device types, lighting setups, and occupational contexts. Bias auditing, threshold recalibration, and uncertainty-aware reporting can help reduce unfair differences in age and engagement estimation. These steps are important because the same model may not perform equally well for every user or environment.

The system should make its outputs easier to understand. Instead of only showing an age or engagement label, later versions could include confidence scores, engagement trends over time, uncertainty bands, and short alert explanations. This would help administrators and employees see why the system generated a certain estimate, rather than treating the result as a black box.

A more mature version should also let users respond to system-generated outputs. If an estimate affects attendance correction, managerial alerts, or intervention workflows, users should be able to annotate it, challenge it, or request a review. This would make the system fairer in practice and would support the human-centered governance approach more strongly.

Privacy should be improved through edge-assisted or on-device preprocessing. Raw facial video should not be transmitted or stored unless it is truly needed. Future implementations can use feature-only inference pipelines, short-lived frame retention, and encrypted storage to reduce privacy risks. This would make the system safer for real organizational use.

The architecture could also use federated learning or other privacy-aware adaptation methods. This would allow the model to improve across different sites without collecting sensitive user images in one central place. That would be useful for deployment across multiple departments or organizations, where each site may have different users, devices, and working conditions.

Future versions can include anomaly detection for administrative risks, such as suspicious attendance edits, unusual leave patterns, or repeated correction attempts. These alerts should stay advisory. They can help reviewers notice possible problems, but final decisions should still require human approval.

The backend rules can also be strengthened through formal verification. This would help prove that important invariants always hold, even when multiple requests happen at the same time. Examples include attendance lock integrity, unique identity constraints, and single-path approval transitions. Proving these rules would make the system more reliable under real workload conditions.

Longer testing is also needed. Future studies should evaluate the framework across multiple departments, device environments, and longer deployment periods. The evaluation should not focus only on accuracy and latency. It should also measure user trust, managerial usefulness, perceived fairness, and psychological acceptability, because these factors affect whether people will actually accept the system.

The dashboard can be expanded for higher-level planning, but it must be done carefully. A privacy-aware dashboard could combine engagement trends, attendance consistency, leave utilization, and workload indicators. At the same time, clear safeguards would be needed so the system does not become excessive monitoring or get used in ways that harm employees.

The framework may also be adapted for other domains, such as education, healthcare support settings, smart workplaces, and public-service interfaces. Each domain would need its own policy layer and its own definition of engagement. The meaning of engagement in a classroom is not the same as engagement in a workplace or healthcare support environment, so the system would need calibration for each setting.

Security can be improved in later versions as well. Hardware-backed key storage, session anomaly detection, continuous authorization checks, and stronger audit analysis could reduce risks from credential theft, privilege escalation, and insider misuse. These additions would help protect the system as it moves from prototype-level testing toward larger organizational deployment

References

[1] O. Agbo-Ajala and S. Viriri, “Apparent age prediction from faces: A survey of modern approaches,” *Frontiers in Big Data*, vol. 5, 2022.

[2] K. ELKarazle, S. Raman, and T. M. Ma, “Facial age estimation using machine learning techniques: An overview,” *Big Data and Cognitive Computing*, vol. 6, no. 4, 2022.

[3] S. E. Bekhouche, A. Benlamoudi, F. Dornaika, H. Telli, and Y. Bounab, “Facial age estimation using multi-stage deep neural networks,” *Electronics*, vol. 13, no. 16, 2024.

[4] N. Panić, M. Marjanović, and T. Bezdan, “Addressing demographic bias in age estimation models through optimized dataset composition,” *Mathematics*, vol. 12, no. 15, 2024.

[5] S. Kim, H. Lee, and J. Kim, “Knowledge distillation for enhanced age and gender prediction in lightweight face models,” *Mathematics*, vol. 12, no. 17, 2024.

[6] A. Ali *et al.*, “P-Age: Pexels dataset for robust spatio-temporal apparent age classification,” in *Proc. IEEE/CVF Winter Conf. on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, 2024.

[7] P. Chen *et al.*, “DAA: A delta age AdaIN operation for age estimation via residual adaptation,” in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, 2023.

[8] K. Narayan *et al.*, “FaceXFormer: A unified transformer for facial analysis,” in *Proc. IEEE/CVF Int. Conf. on Computer Vision (ICCV)*, Honolulu, HI, USA, 2025.

[9] Y. Wu *et al.*, “Deep learning-based facial age estimation via JPSD feature fusion,” *Machine Learning and Knowledge Extraction*, vol. 8, no. 6, 2025.

[10] H. Zhang *et al.*, “Individual aging pattern modeling for enhanced facial age estimation,” *Electronics*, vol. 14, no. 23, 2025.

[11] S. Yu *et al.*, “Improving age estimation in occluded facial images with facial feature reconstruction,” *Applied Sciences*, vol. 15, no. 11, 2025.

[12] X. Qiao *et al.*, “Semantic attention guided hierarchical decision network for facial age estimation,” *Applied Sciences*, vol. 15, no. 14, 2025.

[13] M. Kocoń *et al.*, “Age estimation and gender classification from facial images,” *Applied Sciences*, vol. 15, no. 18, 2025.

[14] J. Paplhám and V. Franc, “Photo dating by facial age aggregation,” in *Proc. IEEE/CVF Winter Conf. on Applications of Computer Vision (WACV)*, Tucson, AZ, USA, 2026.

[15] A. Chaubey, X. Guan, and M. Soleymani, “Face-LLaVA: Facial expression and attribute understanding through instruction tuning,” in *Proc. IEEE/CVF Winter Conf. on Applications of Computer Vision (WACV)*, Tucson, AZ, USA, 2026.

[16] R. Rothe, R. Timofte, and L. Van Gool, “DEX: Deep expectation of apparent age from a single image,” in *Proc. IEEE Int. Conf. on Computer Vision Workshops (ICCVW)*, Santiago, Chile, 2015.

[17] R. Rothe, R. Timofte, and L. Van Gool, “Deep expectation of real and apparent age from a single image without facial landmarks,” *Int. J. Computer Vision*, vol. 126, no. 2–4, pp. 144–157, 2018.

[18] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua, "Ordinal regression with multiple output CNN for age estimation," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 4920–4928.
[19] S. Chen, C. Zhang, M. Dong, J. Le, and M. Rao, "Using ranking-CNN for age estimation," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017.
[20] W. Shen, Y. Guo, Y. Wang, K. Zhao, B. Wang, and A. Yuille, "Deep regression forests for age estimation," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018.
[21] W. Li, J. Lu, J. Feng, C. Xu, J. Zhou, and Q. Tian, "BridgeNet: A continuity-aware probabilistic network for age estimation," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019.
[22] H. Pan, H. Han, S. Shan, and X. Chen, "Mean-variance loss for deep age estimation from a face," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018.