CMS-R-74 (OMB 0938-0467) — Strengthening IEVS Quality and Burden Reduction

Submitted on behalf of Promise as an organizational comment to CMS-10410

Promise.

Comment

Introduction & Scope
We submit this comment to support continuation of the Income & Eligibility Verification System (IEVS) information collection while improving **data quality, auditability, accessibility, and burden reduction**. My remarks focus on (1) match quality controls, (2) human/applicant-in-the-loop (HITL/AITL) corrections, (3) accessible channels (voice/IVR), (4) privacy-by-design, and (5) machine-readable interoperability across programs.

1) Match Quality Controls (false-positive/negative + drift)
CMS should specify minimum QA for IEVS matching:
a) Track false-positive (FP) and false-negative (FN) rates by source (e.g., wage data, PARIS) and by channel of intake; publish thresholds and confidence bands used in matching.
b) Require drift monitoring of matching rules/scores with documented re-calibration cadence (e.g., quarterly) and change logs.
c) Define a minimum match log per record: algorithm/rule ID, model version, features used, raw score, threshold, outcome, and final adjudication (accepted/overturned).
d) Encourage a risk-based review: only matches outside tolerance or with low confidence require staff attention.

2) HITL/AITL Corrections (burden-reducing)
Permit and encourage applicant-in-the-loop corrections for mismatches without forcing full resubmission. Required elements: field-level provenance (source, method, timestamp, confidence), attestation capture, and audit trail linking the correction to the final decision. This reduces staff burden and shortens time-to-resolution while improving accuracy.

3) Accessible Channels (voice/IVR equivalence)
Authorize IVR/virtual agent collection of verification clarifications with recorded consent and channel-parity validations (same data elements, same disclosures, same business rules as web). Log channel, language, consent method, and escalation outcomes to the same audit trail as web submissions. Require multilingual prompts and disability accommodations (pause/repeat, barge-in, TTY/relay).

4) Privacy & Security (least-privilege and auditability)
Adopt least-privilege access, retention safe-harbors, and cryptographic or tamper-evident logging for match/audit records. Emphasize PHI minimization and

role-based access controls. Clarify expectations for vendors (e.g., SOC 2/StateRAMP-equivalent) without prescribing a single cloud.

5) Interoperability (machine-readable standards)
Recommend machine-readable field schemas and enumerations for IEVS match status, confidence, and adjudication so states can reuse verified facts across Medicaid/CHIP/SNAP/TANF with consent. Encourage exportable, field-level provenance (source system, document ID, extraction/match method, confidence, timestamps) to support appeals and QC.

6) Transparency on PARIS Use
Request a simple, annual attestation describing PARIS participation, match logic at a high level, QA metrics (FP/FN), and corrective actions taken—lightweight for states but valuable for consistency and oversight.

7) Burden Estimates & Automation
Allow states to credit automation that reduces manual verification (e.g., AITL corrections, standardized schemas, API-based exchanges). Encourage states to report deltas in staff time and rework avoided, so PRA burden reflects modernization gains over time.

Conclusion
These recommendations maintain program integrity and reduce administrative burden by improving match quality, enabling applicant corrections, ensuring accessible channels, and standardizing interoperable, auditable data. Thank you for the opportunity to comment.

Attachment 1 — Accessible Channel Parity Checklist

Use this as a one-pager you can paste/upload. Check what you already meet; circle planned items + target dates.

Program / State: _____ Submitter: _____ Date: _____

A) Scope & Channel Parity

- ☐ Intake/renewal data elements are identical across Web / SMS / Voice (IVR or virtual agent) / Assisted.
- ☐ Same validations (required fields, formats, eligibility rules) across channels.
- ☐ Same consent text and attestation captured across channels (language-appropriate).
- ☐ Channel, language, and session identifiers are logged with each submission.

B) Consent & Provenance Logging

- ☐ Per field, we log: source (applicant, doc, authoritative dataset, staff), extraction method, confidence score, timestamp.
- ☐ Consent object includes method (web/sms/voice/paper), version of disclosure, and user attestation.
- ☐ Voice channel stores audio hash / transcript segment IDs tied to fields.

C) Identity, Privacy & Security

- ☐ Minimum necessary PHI collected; PII class tagged per field.
- ☐ Encryption in transit/at rest; role-based access; retention schedule applied automatically.
- ☐ Vendor controls: SOC 2 (or StateRAMP/FedRAMP-equivalent) documented.
- ☐ IP/device/session hashes logged for audit; no raw device fingerprints retained beyond policy.

D) LEP & Accessibility

- ☐ Multilingual flows for top LEP languages; plain-language prompts.
- ☐ Voice flows meet TTY/relay accommodations; pause/repeat; barge-in; escalation to human.
- ☐ Low-bandwidth SMS fallbacks for web/voice failures.

E) Model-Risk & Quality (for automated matching/extraction)

- ☐ False-positive / false-negative rates tracked by data source and channel.
- ☐ Drift monitoring with thresholds that trigger HITL review.
- ☐ Versioned matching/extraction algorithms with change logs.

F) Performance & Equity Metrics (disaggregated where feasible)

- ☐ Completion rate by channel & language
- ☐ Time-to-decision (median, p90)
- ☐ Abandonment rate by step
- ☐ IVR/Virtual agent containment target: ____% (current: ____%)
- ☐ Human handoff target: ____% (current: ____%)
- ☐ Renewal success rate by channel & LEP status

G) Governance & Evidence

- ☐ Annual accessibility/LEP review and stakeholder testing plan.
- ☐ Appeals/QC workflows can export full provenance (field-level) on request.
- ☐ Data-sharing agreements reflect **verify-once, reuse-everywhere** with consent controls.

Attachment 2 — Model-Risk & Quality Metrics

Metric	Definition	Unit	Target	Calculation	Notes
IVR Containment Rate	% of calls resolved without human	%	≥ 40%	resolved_without_handoff / total_calls	Track by language & intent
Human Handoff Rate	% routed to human	%	≤ 35%	handoffs / total_calls	Includes warm transfers
Time-to-Decision (p50/p90)	From submission to adjudication	minutes	p50 ≤ 60	median/ percentile	Split initial vs renewal
FP Match Rate	False positives in data matching	%	≤ 2%	fp / total_matches	By source (wage, PARIS, etc.)
FN Match Rate	False negatives	%	≤ 2%	fn / total_matches	Requires adjudicated ground truth
Extraction Confidence	Mean confidence for AI-extracted fields	0–1	≥ 0.90	avg(confidence)	Flag < threshold to AITL
AITL Resolution Rate	% low-confidence fields corrected/confirmed by applicant	%	≥ 80%	resolved / flagged	By channel
Appeal Overturn Rate	% decisions reversed on appeal	%	≤ 3%	overturned / total_decisions	Monitor drift/bias
Equity Gap	Completion rate gap (English vs LEP)	p.p.	≤ 5 p.p.	completion_en – completion_lep	By channel

Attachment 3 — Human-in-the-Loop Governance Example Index

Figure Ref	Simple Example	Governance Principle
Figure 1	*AI highlights "pay now" on a utility bill and suggests Crisis award.* Caseworker must still confirm the document's full context and check if any conflicting policy applies.	NLP recommendations must be confirmed by a human before award.
Figure 2	*AI pre-fills gross income from two paystubs.* Caseworker must confirm dates span a full 30-day window before proceeding.	Field extraction must be human-verified before used in eligibility logic.
Figure 3	*A SNAP award letter is detected by AI, but the expiration date is unclear.* Caseworker reviews the letter and manually enters eligibility end date.	Ambiguous extraction must defer to human confirmation.
Figure 4	*An ID image is flagged by AI as expired due to OCR.* Caseworker sees the date is actually MM/DD/YYYY format (not DD/MM/YYYY) and overrides the AI suggestion.	Overrides must be logged and permitted when confidence is low.
Figure 5	*AI generates a case note: "Application missing ID."* Staff edits the message to clarify which household member is missing ID before submission.	AI-generated messages must be editable and attributed.

Figure 6	*AI recommends "Home Energy award only" due to missing disconnection language.* Caseworker spots "service interruption" in the document and upgrades to Crisis.	AI cannot be the final authority on policy classification.
Figure 7	*AI chatbot answers: "You likely qualify if your income is below $2,000."* System includes a note: "This is an estimate. Please submit your application for a full review."	AI instructional content must include disclaimers and hand-off cues.
Figure 8	*AI suggests an outreach message to clients who haven't submitted documents within 5 days.* Staff edits the message tone and adds legal disclaimer.	Communication AI must be editable and logged.
Figure 9	*Planning AI shows a ZIP code with a 300% increase in denials.* Internal review identifies a local document ID bug—not a true policy issue.	Forecasting AI outputs must be reviewed before policy is adjusted.
Figure 10	*Fine-tuned model is deployed for fraud flagging.* Performance audit shows it overflags multi-generational households. Model is rolled back and retrained.	Fine-tuned models must retain rollback, audit trail, and fairness logs.

Attachment 4 — Bibliography Appendix

Sources

- Florida CS/SB 1680 (2024) [CS/SB 1680, L83-L91]: Establishing AI ethics and provenance standards.
- Florida DCF RFI [Florida DCF RFI, L11-L15]: AI-driven fraud detection initiatives.
- OMB Memo M-24-10 [OMB Memo M-24-10, L735-L743]: Federal guidance for AI governance.
    - OMB M-24-10, Section B.3 (AI-generated public outputs must be attributed, editable, and logged)
    - OMB M-24-10, Section B.2–B.4 (Oversight, Auditability, Human Review)
    - OMB M-24-10, Section B.5 (Forecasting systems must undergo review if informing operational decisions)
    - OMB M-24-10, Section C.3: "Updates to AI systems must preserve original governance guarantees."
- NIST AI Risk Management Framework [NIST AI Risk Management Framework, L7-L9]: Traceability and risk mitigation practices.
    - NIST AI RMF, Section 4.3 (Risk Tiers for Planning and Surveillance Use Cases)
    - NIST AI RMF, Section 4.4.4: “Versioning and documentation of retraining events”
- HHS AI Strategy [HHS AI Strategy, L5-L8]: Program integrity and trustworthy AI.
- HHS "Public Benefits and AI" [HHS "Public Benefits and AI," L705-L713, L727-L735]: AI oversight and fairness.
    - HHS “Public Benefits & AI,” p. 11–13 (Prohibition on AI-only determinations)
    - HHS “Public Benefits & AI,” p. 13–15 (Transparency and explainability in communications that affect access)
    - HHS “Public Benefits & AI,” p. 17 (Clear communication obligations for AI in benefits access)
- White House AI Bill of Rights, Principle #1 (Notice and Explanation) and Principle #3 (Human Alternatives)

- California SB 1120 [California SB 1120, L253-L261]: Mandated human oversight in health and benefits eligibility.
    - California SB 1120, Sec. 2 (Human review required for algorithmically influenced eligibility decisions
    - California SB 1120, Sec. 2(d) (Instructional systems may only assist—not predict—outcomes)
    - California SB 1120, Sec. 2(f) (Public service messages generated by AI must disclose origin and allow intervention)
- Colorado SB 24-205 [Colorado SB 24-205, L232-L240]: High-risk AI obligations, bias prevention.
    - Colorado SB 24-205: “Material change in model weights must be reviewed at the same level as new model deployment.”
    - Colorado SB 24-205, Sec. 4(c) (Audit trail and impact classification of high-risk AI)
    - Colorado SB 24-205, Sec. 5 (Generative content must be auditable and subject to correction mechanisms)
    - Colorado SB 24-205, Sec. 6 (Transparency and traceability for instructional or communicative AI)
    - Colorado SB 24-205, Section 7 (Risk tiering of non-determinative systems)
- WBUR On Point AI welfare fraud analysis [WBUR On Point, L115-L123]: Risks of AI-driven benefit denials.
- Los Angeles County AI chatbot pilot [Route Fifty, L242-L250]: AI documentation best practices.
- FEMA Eligibility Decisions [FEMA Eligibility Decisions, L338-L342]: Identity verification standards for disaster aid.
    - FEMA Policy 104-009-18 (Fraud detection and disaster analytics must include human adjudication in all final actions)
- R Street Institute analysis [R Street Institute, L82-L89]: Balanced AI regulatory trends in state policies.