

Talk naturally. Leave with better English.

A conversation-first English coach powered by Tavus CVI.

Fluent Me closes the gap between knowing a phrase and retrieving it in a real conversation. The learner talks face to face, asks for help when useful, rehearses one exact phrase, and leaves with one grounded next step.

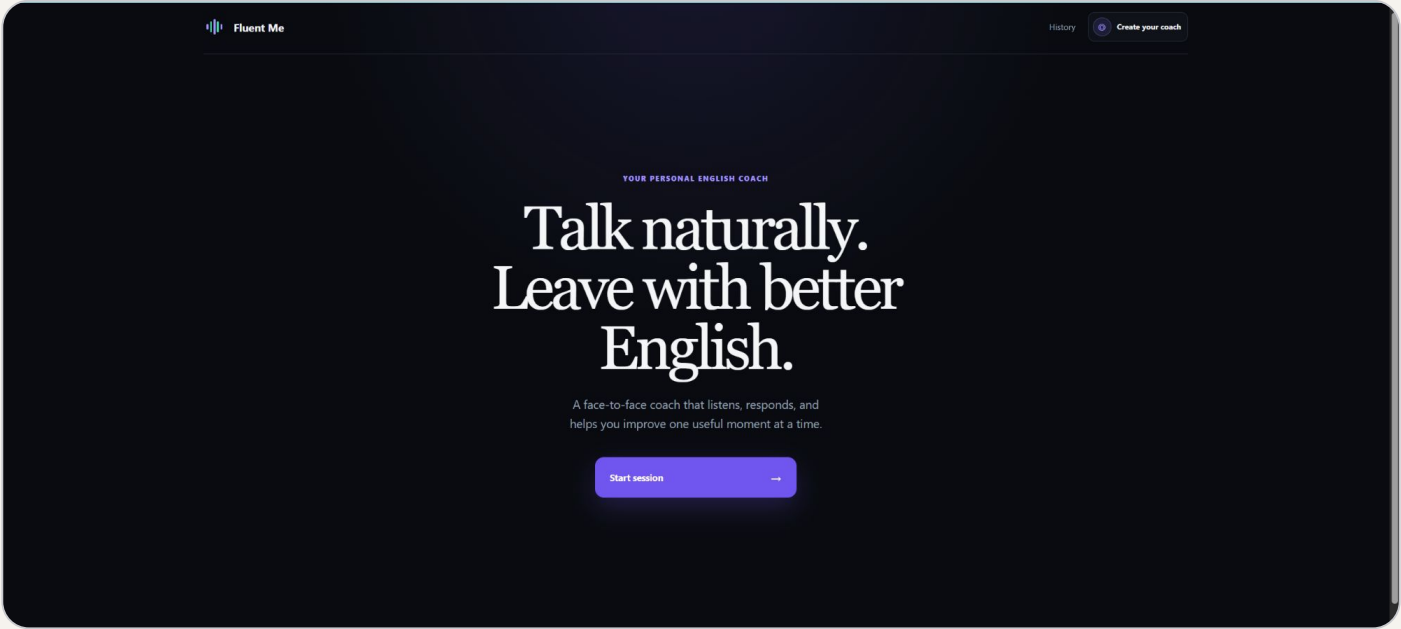
VIDEO-FIRST A person to address, listen to, and imitate.	EVIDENCE-BOUNDED Available signals stay separate from interpretation.	LEARNER-CONTROLLED Feedback, camera, history, and memory remain optional.
--	---	---



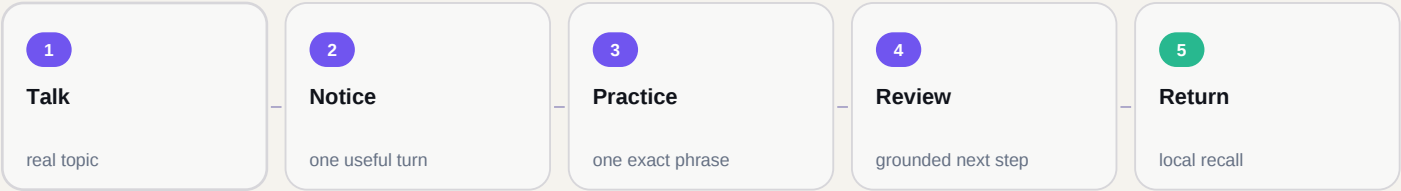
01 / PRODUCT THESIS

The face is part of the learning loop.

A text model can rewrite a sentence. Speaking practice also requires turn-taking, retrieval pressure, modelling, and continuity. Fluent Me uses the live video coach as the primary interaction surface and keeps analysis secondary until the learner asks for it.



Simplified entry surface: one product promise and one action. Topic, duration, and local-save choices are optional rather than prerequisites.



Design decision

The first prototype was a rigid lesson sequence. The redesign keeps one continuous conversation, with Feedback, Practice, and Review as callable tools instead of locked stages.

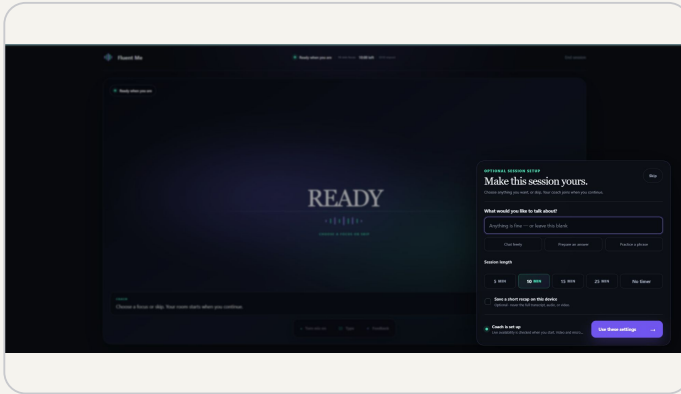
Product boundary

Camera is off by default. Detailed analysis is hidden behind a drawer or disclosure so a first-time learner can speak before learning the interface.

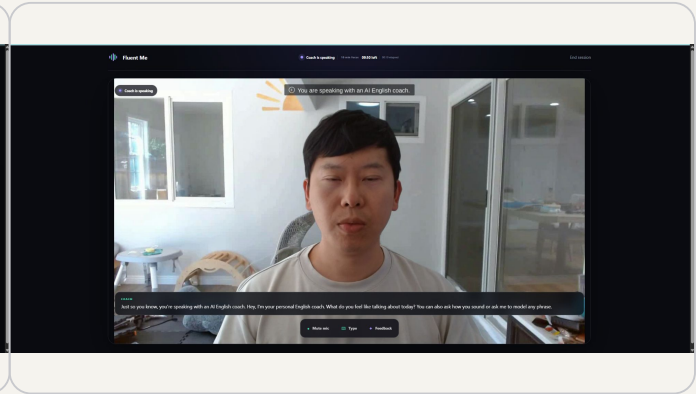
SUPPLEMENT

Product surfaces

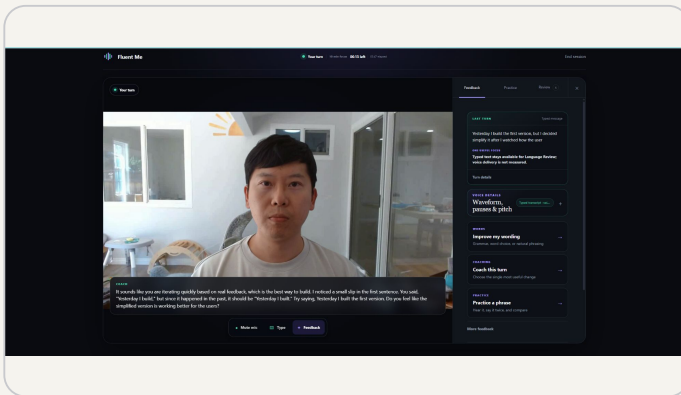
Additional screenshots are included only when corresponding files exist in submission/media.



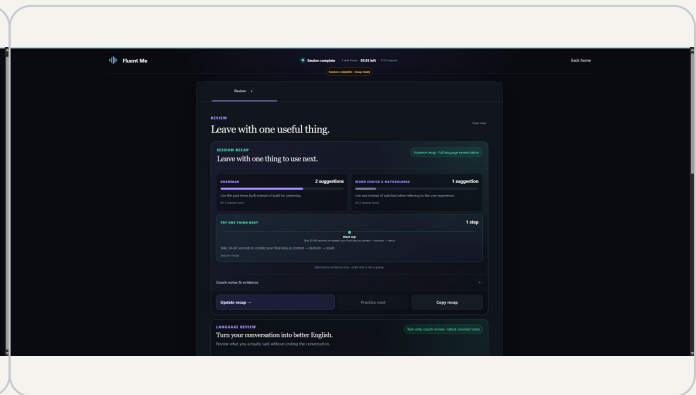
Setup



Coach Live



Feedback



Recap

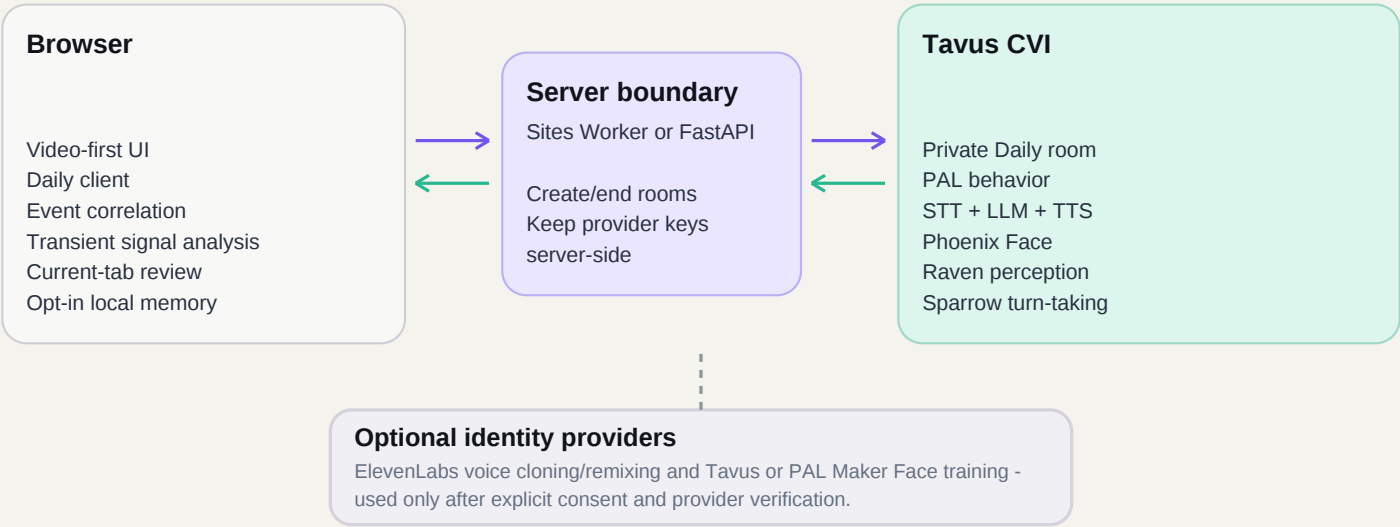
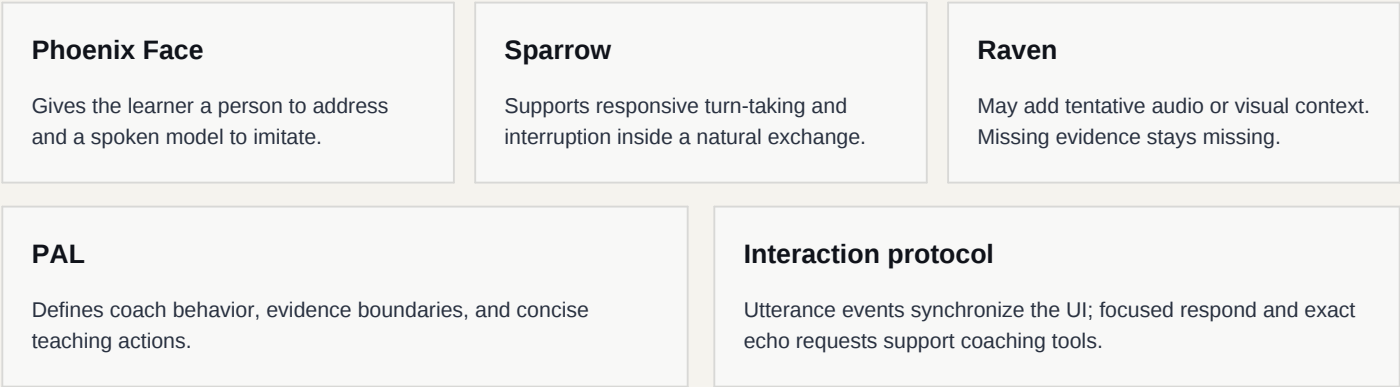
Evidence note

A screenshot demonstrates interface state only. A live recording is still required to prove microphone publication, Tavus Face playback, model response, and remote-room cleanup.

02 / WHY TAVUS

Tavus is structural, not decorative.

The product works because a real-time video conversation, turn timing, bounded perception, coach behavior, and interaction events operate together. The browser supplies the learning experience; the server protects credentials and owns room lifecycle.



Trust boundary: the browser receives a private room URL and short-lived meeting token. Provider API keys remain server-side. Optional identity providers are separate from the core conversation flow.

03 / EVIDENCE INTEGRITY

Useful feedback without a mystery score.

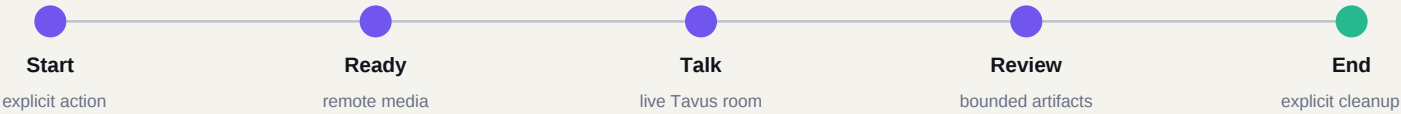
Fluent Me keeps measured evidence, model perception, and teaching guidance separate. It does not collapse every available signal into an English level or confidence number.

1. Measured or counted Turn timing, transcript words, filled-pause matches, adjacent repeats, and transient browser-signal summaries.	2. Model perception Raven observations only when returned, labelled qualitative and tentative rather than ground truth.	3. Teaching guidance A recast, stress model, or next practice step that helps rehearsal but is not presented as measurement.
SUPPORTED OUTPUT - Actual learner transcript from Tavus conversation events - Duration and WPM when timing and enough words are available - Transcript counts for filled pauses and adjacent repeats - Within-turn waveform, estimated pauses, level movement, and pitch candidates when browser signal is available - Latest-12-turn Language Review with explicit coverage	EXPLICITLY NOT CLAIMED - No pronunciation, accent, fluency, or emotion score - No phoneme or syllable accuracy without a specialist provider - No claim that waveform height compares across turns - No inference of inner emotion, personality, intelligence, or ability - No percentage-improved or population comparison	
Bounded language review Grammar, word choice, natural expression, and a fact-preserving polished version use at most the latest 12 non-empty learner turns. Names, numbers, links, and meaning must be preserved; the full learner transcript remains a current-tab artifact.		

04 / LIFECYCLE + PRIVACY

Design the room lifecycle, not only the happy path.

Real-time AI has operational states that a form demo can ignore. The timer starts only after remote coach readiness, and manual end, timed end, connection failure, and page exit converge on remote cleanup attempts.



Current tab
Full transcript view, waveform, pitch contour, attempt evidence, recap, and Language Review.

Optional on device
At most 20 compact finalized recaps and learner-approved phrases with scheduling metadata.

Never in local history
Raw audio/video, full transcript, waveform, pitch contour, Raven observations, or provider credentials.

Learning History	Opt-in before the session; max 20 compact finalized entries; individual delete and clear-all controls.
Learning Memory	Only an explicitly saved phrase plus fixed review metadata. Reveal or rehearsal does not advance recall.
Review rhythm	A visible 1/3/7/21/60-day product rule. Not a measured personal forgetting curve; Not quite returns in 10 minutes.

Optional identity is a separate consent path
Voice cloning and Face training require explicit ownership/authorization, local review before submission, provider entitlement, and asynchronous verification. The stock coach remains the fallback. A clone or Face is not described as complete merely because the setup UI exists.

05 / DELIVERY TRUTH

What the code supports - and what still needs proof.

This report does not substitute interface state, mocked provider responses, or automated contracts for a continuous live evaluator run against the exact submitted build.

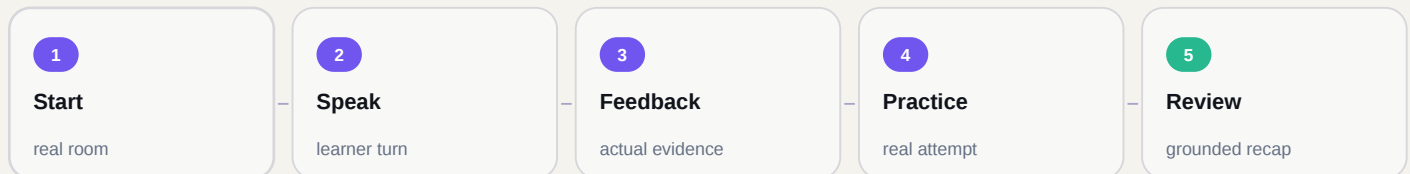
IMPLEMENTED IN THE PRODUCT

- Conversation create/join/end paths in hosted Worker and FastAPI implementations
- Video-first conversation with on-demand Feedback and Practice surfaces
- Timed sessions, bounded recap, and latest-turn Language Review contracts
- Transient signal visualization and transcript-derived evidence
- Opt-in compact History, learner-approved Memory, and fixed review schedule
- Consent-gated optional personalization integration paths

REQUIRES DIRECT VERIFICATION

- Signed-out reviewer access to the exact deployment
- Live microphone reaching Tavus and a relevant spoken response
- Moving and audible Tavus Face in the final continuous recording
- Raven availability for the exact demonstrated turn, if claimed
- ElevenLabs entitlement and a real accepted voice-clone/remix request
- Completed Face training and personal PAL playback, if demonstrated

REVIEWER WALKTHROUGH



Next product work

Maintain controlled reviewer access and spend controls; instrument latency without retaining learner media or full transcript history; validate specialist pronunciation scoring before adding it.

Assessment boundary

Before adding phoneme, syllable, stress, or prosody scores, evaluate a specialist pronunciation provider and validate the learner benefit.

Source materials: [submission/README.md](#), [submission/ARCHITECTURE.md](#), [submission/DEMO_SCRIPT.md](#), and [submission/DELIVERY_CHECKLIST.md](#). Public links and the exact code revision are supplied in the submission handoff.