Human Behavioral Simulation & Analysis Framework

**Use Case:** Real-time simulation & reaction analysis of a subject in isolation under multimedia stimulus.

1. Introduction

This framework integrates computer vision, audio cues, physiological signals, and large language models (LLMs) to analyze a human subject’s emotional, cognitive, and behavioral responses to screen content in isolated environments. Applications include: - Military training evaluations - Executive stress testing - Psychological resilience research - Critical job screening

This simulation aims to model and analyze human behavior by showing targeted visual/audio stimuli (e.g., emotional films, VR content, emergency training footage), while continuously recording multimodal responses such as facial expressions, speech, micromovements, physiological reactions (optional), and mental state transitions over time.

2. Modules Overview

|  |  |
| --- | --- |
| Module | Key Features |
| Visual Analysis | Facial expressions, eye gaze, body posture, micro-movements |
| Audio Emotion & Reaction | Breathing patterns, speech analysis, sighs, murmurs |
| Physiological Signals | (Optional) HRV, GSR, EEG via wearables |
| Context-Aware LLM Analysis | Screen content understanding + subject response analysis |
| Behavior Classification | Mental fatigue, boredom, stress, attention, confusion |
| Simulation Adaptation Engine | Adjusts difficulty/content based on real-time analysis |

3. Visual Stream Processing

**3.1 Preprocessing** - CLAHE: Enhance facial contrast - Denoising: Median filtering - Lighting normalization: Local gamma correction

**3.2 Detection & Tracking** - YOLOv8: Full body, hand, and face detection - ByteTrack: Identity persistence across frames - OpenFace + OSNet: Gaze + appearance tracking

**3.3 Facial & Body Cues** - FER+: Emotion classification - OpenFace: Eye gaze direction, head pose - PoseNet or MediaPipe: Full-body pose estimation - CBAM + ResNet50: Feature enhancement

4. Audio Stream Processing

**4.1 Raw Audio Inference** - Whisper: Speech-to-text (mumbling, self-talk, mutters) - wav2vec 2.0: Emotion-rich voice embeddings - Librosa: Prosody (tone, pitch, tremors)

**4.2 Non-Speech Audio** - Breath patterns, silence gaps, sighs detected via amplitude and spectral features

5. LLM-Based Contextual Analysis

**5.1 Input** - Screen content (script, video, game logs) - Subject response (verbal or behavioral markers)

**5.2 Implementation** - LLaMA 3 + spaCy: Content abstraction → knowledge graph - LLM comparison: Check alignment of content vs subject reaction - Confusion heatmap: Marks areas of disengagement or cognitive overload

**5.3 Real-world Alignment** Inspired by psychological studies analyzing viewer empathy, confusion, and emotional engagement from passive video watching, this module brings a deeper semantic analysis layer that other sensor-only systems lack.

6. Temporal Analysis

**SlowFast (Video) + BiLSTM (Multimodal)** - Slow path (4 fps): Long-term behavior - Fast path (16 fps): Short-term spikes (stress, blinking, fidgeting) - BiLSTM + Attention: Joint modeling of vision + audio features - Output: Mental state timeline (e.g., Calm → Confused → Alert)

7. Physiological Data Integration (Optional)

* Heart Rate Variability (HRV): Stress and engagement
* Galvanic Skin Response (GSR): Arousal levels
* EEG: Attention and cognitive load tracking
* Fused via Transformer-based temporal encoder
* Matches methodologies from studies using non-contact HRV, GSR under emotional film stimuli to assess internal arousal and stress buildup

8. Evaluation Output

|  |  |
| --- | --- |
| Output | Description |
| Real-time reaction map | Heatmap of attention, stress, confusion |
| Event-based trigger detection | Detects moments of disengagement or extreme emotion |
| Session report | Emotional timeline, attention drops, peak triggers |
| Subject profile scoring | Stress resilience, confusion tolerance, focus score |
| Simulation feedback | Suggests better content delivery pace/density |

9. Accuracy Table (Sample)

|  |  |  |  |
| --- | --- | --- | --- |
| Module | Model Used | Task / Dataset | Accuracy / Metric |
| YOLOv8 | Object Detection | COCO | AP: 53.9% |
| FER+ | Emotion Detection | FER+ | ~85% |
| SlowFast | Action Recognition | Kinetics-400 | Top-1: 79.8% |
| Whisper | Speech Recognition | LibriSpeech | WER: 2.5–4.9% |
| wav2vec 2.0 | Emotion via audio | LibriSpeech | WER: 1.8–3.3% |
| BiLSTM + Attention | Multimodal fusion | Custom human sim dataset | ~90% (reported in similar works) |
| LLaMA 3 | Semantic alignment | MMLU | ~86% |

10. Applications Beyond Simulation

* High-stakes job screening
* Training under pressure (e.g., pilots, air traffic control)
* Therapy progress evaluation
* Deepfake exposure to analyze authenticity reactions
* PTSD simulation response studies

11. Workflow Diagram Guide (To Draw Yourself)

