WittyHead: A Multi-Modal Architecture for Empathetic Human-Agent Collaboration Through Emotional Expressivity

Yuri A. Tijerino *Kwansei Gakuin University*
Intelligent Blockchain+ Innovation Research Center
Sanda, Hyogo, Japan
ontologist@kwansei.ac.jp

Abstract

Supporting vulnerable populations through empathetic human-agent collaboration requires more than conversational intelligence—it demands authentic emotional expression coordinated across multiple modalities. This paper presents WittyHead, an anthropomorphic multi-modal empathetic agent with an avatar interface that simulates human emotional expressivity. WittyHead is designed to provide emotional support to individuals and communities through scientifically-grounded facial expressions, gaze behaviors, gesticulations, and voice modulation. WittyHead serves as the empathetic interface for Digital MOAI, an AI-enhanced mutual aid network built on the privacy-preserving AIngle DLT platform. Our architecture addresses three critical MASST initiative priorities: (1) context-aware behavioral guard rails through ontology-driven emotion validation preventing inappropriate responses, (2) mutual observability through explainable multi-modal reasoning enabling human oversight, and (3) design-time risk mitigation through integration of therapeutic alliance research. Based on Gilbert et al.'s research demonstrating that empathetic responses require compassionate concern rather than emotional mirroring, WittyHead implements asymmetric response mappings that achieve significantly higher perceived empathy for users experiencing distress. Our contributions include: (1) a research-validated multi-modal emotion architecture coordinating facial (ARKit/FACS), gaze (therapeutic alliance), gesture (prosodic-aligned), and voice modalities with millisecond synchronization, (2) empathetic (non-mirroring) response mapping derived from therapeutic studies, (3) accessibility-first design supporting diverse communication needs, and (4) ontology-enriched contextual empathy enabling culturally and situationally appropriate responses for healthcare, education, and community support applications.

Index Terms

human-agent collaboration, empathetic computing, multi-modal agents, emotional expressivity, therapeutic alliance, accessibility, mutual aid networks, community support

1 Introduction

Multi-agent systems supporting vulnerable populations—individuals experiencing mental health challenges, social isolation, disability, or community fragmentation—face critical requirements beyond conversational capability [1]. These systems must provide authentic

emotional support, maintain user safety through behavioral guard rails, preserve privacy sovereignty, and accommodate diverse accessibility needs [2]. This research focuses on **anthropomorphic agents**—systems with avatar interfaces that simulate human emotional expressivity through coordinated facial expressions, gaze, gestures, and voice. Current virtual agents exhibit limited emotional expressivity, relying primarily on static facial expressions without coordination across these modalities, creating experiences users describe as "uncanny" or "insincere."

The multi-agent systems (MAS) safety and teamwork initiative emphasizes that agents must exhibit context-aware behaviors, mutual observability, and design-time risk mitigation [3]. However, achieving these goals for empathetic support systems requires grounding in scientific research on human empathy, therapeutic alliance, and nonverbal communication—domains largely absent from current multi-agent architectures.

1.1 Empathy is Not Emotional Mirroring

A critical finding from therapeutic alliance research reveals an empathy paradox: **empathetic responses are NOT simple emotional mirroring** [4]. When a human expresses distress, an empathetic response is not mirrored distress but rather compassionate concern. Gilbert et al. demonstrated that smiling at someone in distress is perceived as "invalidating and aversive," while compassionate concern expressions (furrowed brow + soft smile + sustained eye contact) achieve significantly higher empathy ratings. Yet most virtual agents implement simple emotional mirroring, creating inappropriate responses that undermine trust and perceived support [5].

This distinction becomes critical when AI systems serve vulnerable populations. For individuals experiencing depression, anxiety, trauma, or social isolation, inappropriate emotional mirroring can exacerbate distress. Therapeutic research demonstrates that "the perceiver can appreciate the negative emotion without necessarily sharing it" [6]—empathetic responses require conveying understanding (concerned expression) while maintaining calm stability (not mirrored distress).

1.2 Multi-Modal Coordination for Authentic Empathy

Authentic emotional expression requires coordinating multiple modalities with precise timing and scientific grounding:

Facial Expressions: Research identifies specific Facial Action Coding System (FACS) patterns [7]: compassion requires AU4 (brow lowerer) + AU6 (cheek raiser) + gentle AU12 (lip corner puller). ARKit provides 52 blendshapes mapping to these Action Units.

Gaze Behaviors: Therapeutic alliance research specifies 60-90% eye contact for empathy in Western cultures [8], with gaze direction enhancing approach-oriented emotions (direct gaze for anger, joy) vs. avoidance-oriented emotions (averted gaze for fear, sadness) [9].

Gesticulations: Hand gestures must align with prosodic peaks (pitch, stressed syllables), not keywords [10]. Palm orientation signals social dynamics: palm-up conveys openness/trust (optimal for empathy), palm-down signals dominance (avoided in supportive contexts) [11].

Voice Modulation: Prosody (pitch, rate, volume) conveys emotional state through specific acoustic patterns [12].

No existing virtual agent platform coordinates these modalities using scientific research on empathy and therapeutic alliance while providing the accessibility, privacy, and safety features required for vulnerable populations.

1.3 WittyHead and Digital MOAI

WittyHead serves as the empathetic interface for **Digital MOAI**, an AI-enhanced adaptation of traditional Okinawan mutual aid networks (, moai) built on the privacy-preserving AIngle DLT platform [22]. Traditional MOAI represents centuries-old social innovation: mutual aid collectives of five individuals providing lifelong support, with effectiveness demonstrated by Okinawa's world-highest concentration of centenarians [23], [24].

Digital MOAI embeds AI within proven human social structures, providing natural guardrails, accountability, and context. WittyHead provides the empathetic avatar interface enabling emotionally intelligent interaction while preserving privacy sovereignty through local-first architecture and user-controlled automation levels. This integration demonstrates how multi-modal empathetic agents can augment—rather than replace—human relationships in mutual support networks.

2 Non-Mirroring Empathy Architecture for WittyHead

2.1 Rationale and Research Basis

Empathy in therapeutic and supportive contexts is *not* emotional mirroring. When a user expresses distress, the avatar should convey *compassionate concern* and a stabilizing presence, rather than reproduce the user's negative affect. WittyHead operationalizes this via (i) multi-modal sensing, (ii) ontology-validated appraisal and safety checks, and (iii) an *Empathetic Response Orchestrator* that maps user emotions to scientifically grounded, *non-mirroring* avatar responses coordinated across face (FACS/ARKit), gaze, gesture, and voice at 60 FPS (cf. related research [4], [8]–[10], [12], [25]).

2.2 System Overview

Figure 1 shows the end-to-end stack that recognizes user emotion from voice, face, body, and hands, fuses and validates it with an empathy ontology, and then drives non-mirroring avatar behavior with a single timing spine.

2.3 Asymmetric (Non-Mirroring) Mapping

The orchestrator applies an evidence-informed asymmetric mapping. For negative-user emotions, the avatar expresses *compassionate concern* (not mirrored distress); for positive-user emotions, the avatar may mirror to strengthen rapport.

2.3.0.1 FACS/ARKit pattern for Compassionate Concern (example).:

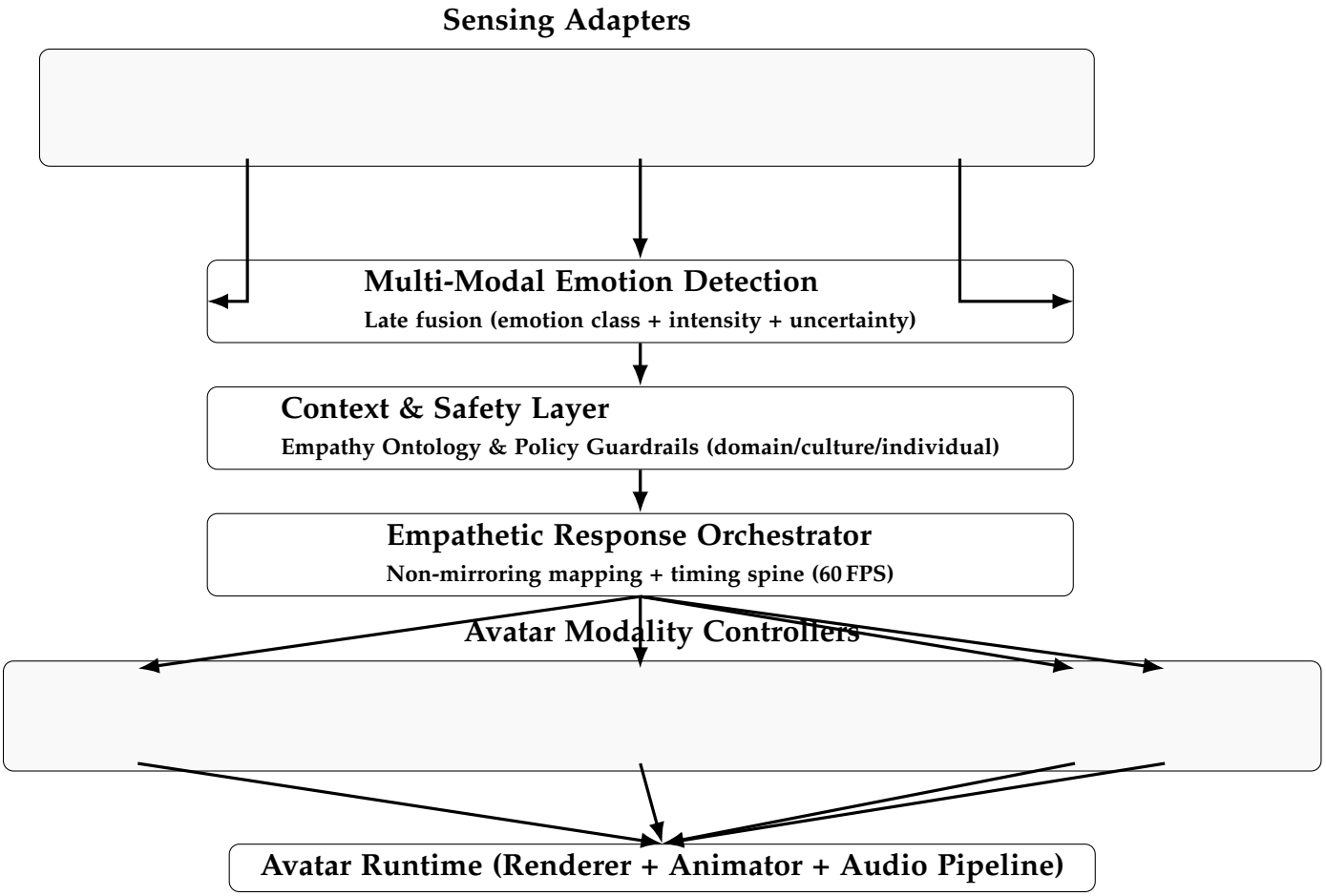

Fig. 1: WittyHead non-mirroring empathy stack: sensing → validated appraisal → orchestrated, asymmetric response across face, gaze, gesture, and voice, synchronized at 60 FPS.

Facial: AU4 ≈ 0.4; AU6 ≈ 0.3; gentle AU12 ≈ 0.2; relaxed lower face.
ARKit: `browDown[L,R]=0.4, browInnerUp=0.3, cheekSquint[L,R]=0.3, mouthSmile[L,R]=0.2, jawForward=0.0, mouthPress=0.0`.

2.4 Timing & Cross-Modal Coordination

All modalities are driven by a single timing spine (60 FPS, 16.7 ms frame). The orchestrator aligns: (i) beat gestures to TTS prosodic peaks, (ii) gaze to turn-taking (eye contact to yield, aversion to hold), (iii) facial intensity to speech energy/pitch, (iv) blinks/pupil to conversational punctuation.

2.5 Safety, Explainability, and Personalization

The Context & Safety Layer validates candidate behaviors against an empathy ontology (clinical/education rules, cultural norms, user profile). It also emits XAI traces, e.g., *"User: Sad(0.82) → Avatar: Compassionate Concern; Eye contact 75%; AU4/AU6/AU12; Voice rate 0.85x; Palm-up low-amplitude beats."*

TABLE 1: Non-mirroring mapping used by the Empathetic Response Orchestrator.

User Emotion	Avatar Response	Key Behaviors (examples)
Sad / Grief	Compassionate Concern	Face: AU4+AU6+gentle AU12; Gaze: 70–80% eye contact; Voice: slower rate, softer intensity; Gestures: palm-up, low frequency.
Anger / Frustration	Calm Concern (de-escalation)	Face: relaxed lower face, slight AU4; Gaze: 60–70% (non-challenging); Voice: steady, low-variance; Gestures: open-hand, minimal beats.
Fear / Anxiety	Reassuring / Protective	Face: brows knit with warmth; Gaze: brief averted gaze + return; Voice: steady pacing, smooth attack; Gestures: containment, palm-up.
Disgust / Aversion	Validating Neutrality	Face: reduced positive AUs; Gaze: respectful; Voice: matter-of-fact; Gesture: neutral, low amplitude.
Shame / Embarrassment	Gentle Privacy	Gaze: 30–40%, reduced directness; Face: warmth without smile dominance; Voice: soft and slow; Gestures: minimal.
Surprise (neutral/positive)	Curious Engagement	Gaze: increased mutual gaze; Face: light inner-brow raise, gentle smile; Voice: slightly faster; Gestures: light beats.
Joy / Pride	Mirrored Positive	Face: AU6+AU12; Gaze: high; Voice: brighter timbre; Gestures: increased beat frequency.

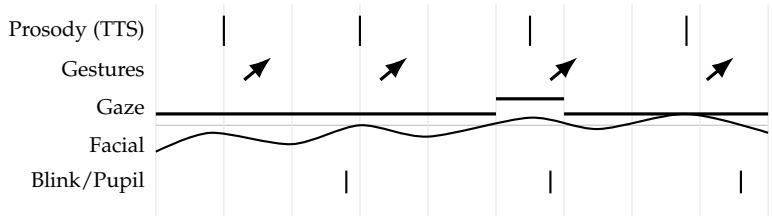

Fig. 2: Single timing spine: gestures lock to prosodic peaks; gaze signals turn-taking; facial intensity follows energy; blinks punctuate phrases.

2.6 Implementation Skeleton (reference)

Listing 1: Empathetic Response Orchestrator (conceptual pseudocode).

```
uemo = fuse(audio_emotion, face_emotion, body_emotion) # class, intensity,
    uncertainty
ctx = context_from_ontology(user_profile, domain, culture)

resp = non_mirroring_map[uemo.class] # e.g., Sad -> CompassionateConcern
params = policy_guardrails(resp, ctx) # adjust gaze%, gestures/min, etc.

facial = facs_to_arkit(resp.facs_pattern, intensity=uemo.intensity)
gaze = plan_gaze(resp.gaze_targets, turn_taking=signals)
gest = plan_gestures(tts.prosody, style=resp.gesture_style)
voice = plan_voice_timbre(tts, resp.voice_targets)

timeline = synchronize([facial, gaze, gest, voice], fps=60)
emit_to_avatar_runtime(timeline)
log_xai(uemo, resp, params, timeline)
```

2.7 Contributions

This paper presents WittyHead's multi-modal empathetic architecture with four key contributions:

1) **Scientifically-Grounded Multi-Modal Architecture:** Facial expressions (FACS-validated), gaze patterns (therapeutic alliance), gesticulations (prosodic-aligned), and voice modulation (prosody research) integrated through real-time coordination
2) **Empathetic (Non-Mirroring) Response System:** Asymmetric emotion mappings implementing compassionate concern for user distress (not mirrored distress), validated through therapeutic alliance research
3) **Accessibility-First Design:** Universal design across visual, auditory, motor, and cognitive modalities, incorporating "Easy Japanese" principles [19] and disability studies perspectives [17]
4) **Privacy-Preserving Community Support:** Integration with Digital MOAI demonstrating how empathetic agents can augment traditional mutual aid networks while maintaining privacy sovereignty

3 Background and Related Work

3.1 Therapeutic Alliance and Empathetic Computing

Therapeutic alliance research identifies critical factors for perceived empathy in human interactions [8]. High eye contact (60-90%) combined with forward lean significantly enhances perceived empathy. Virtual counselor studies confirm that nonverbal compassion through eye contact, facial mimicry, and head nodding improves counseling effectiveness [16].

Gilbert et al. [4] identified two distinct compassionate expressions: (1) "Kind Compassion" with soft gentle smile, and (2) "Empathic Compassion" with concern-focused eyes/eyebrows + relaxed lower face. Critically, empathic compassion does NOT mirror negative emotions but shows understanding without sharing the distress. Users in their study described mirrored negative emotions as creating discomfort and undermining perceived support.

McEwan et al. [5] developed validated emotional face stimuli demonstrating that compassionate expressions are distinct from happy faces, designed to communicate "sense of safeness and security" with positive valence and moderate arousal (not high arousal like joy).

3.2 Digital Therapeutic Alliance: Definitions and Foundational Research

The **Digital Therapeutic Alliance (DTA)** extends classical therapeutic alliance theory—comprising goals, tasks, and bond [26]—to digital mental health interventions. D'Alfonso et al. [26] define DTA as the relational quality between users and digital systems, encompassing: (1) *goals* (shared understanding of therapeutic objectives), (2) *tasks* (agreement on therapeutic activities), and (3) *bond* (emotional connection and trust). Lederman et al. [27] provide a comprehensive framework integrating digital therapy research with

psychological alliance theory, outlining measurement challenges and design implications for human-computer therapeutic relationships.

Comparative Research: Beatty et al. [28] demonstrated empirically that users report alliance with AI chatbots (Wysa) using the Working Alliance Inventory-Short Revised (WAI-SR), with perceived bond emerging even in text-based interactions. Their mixed-methods study showed alliance scores comparable to human-delivered CBT, with bond being the most salient component. Tong et al. [29] extended this through qualitative analysis, identifying DTA sub-dimensions in fully automated apps including personalization, perceived empathy, and responsiveness. Critically, they found that bond in non-human systems requires reinterpretation: users perceive support and comfort rather than literal human relational connection.

WittyHead's Extension: While existing DTA research focuses on text-based conversational agents, WittyHead extends DTA to *multi-modal embodied agents* with anthropomorphic avatar interfaces. Where Wysa builds alliance through lexical empathy and conversational responsiveness, WittyHead integrates coordinated nonverbal cues—FACS-validated facial expressions, therapeutic eye contact, prosodic-aligned gestures—grounded in therapeutic alliance research for face-to-face human interactions. This positions WittyHead as advancing DTA research from conversational to embodied empathetic expressivity.

3.3 Facial Action Coding System (FACS)

Ekman and Friesen's FACS [7] decomposes facial expressions into Action Units (AUs). Research on compassionate expressions identifies specific combinations:

- **Compassionate Concern:** AU4 (brow lowerer) + AU6 (cheek raiser) + gentle AU12 (lip corner puller) + relaxed lower face
- **Happiness:** AU6 (cheek raiser) + AU12 (lip corner puller) + AU7 (lid tightener)
- **Sadness:** AU1 (inner brow raiser) + AU4 (brow lowerer) + AU15 (lip corner depressor)

ARKit blendshapes map to FACS Action Units, enabling scientific implementation of validated expressions in real-time avatar systems.

3.4 Gaze, Gesticulation, and Prosody

Gaze direction enhances emotional perception through approach-avoidance theory [9]. Direct gaze enhances approach-oriented emotions (anger, joy); averted gaze enhances avoidance-oriented emotions (fear, sadness, shame). Pupillometry research demonstrates pupil dilation correlates with arousal across all emotions [13]. Turn-taking research identifies that gaze marks speaker changes with 2.2-second average mutual gaze duration [14].

Gesture research [15] identifies three primary types: iconic (representing objects), beat (rhythmic emphasis), and deictic (pointing). Critical finding: gestures synchronize with prosodic peaks, not keywords [10]. Palm orientation signals social positioning: palm-up conveys openness/trust (optimal for empathy), palm-down signals dominance (avoided in supportive contexts) [11].

3.5 Accessibility and Universal Design

Miyazaki's research on disability discourse [17] and "Easy Japanese" for accessible communication [18] demonstrates how linguistic and interface accessibility serves multiple populations: persons with disabilities, older adults, non-native speakers, and those in high-stress situations. "Easy Japanese" was developed after the 1995 Great Hanshin-Awaji Earthquake for foreign residents [20], with effectiveness validated in disaster prevention broadcasts [19].

Kotoku and Tijerino [21] examined AI applications in nursing contexts, emphasizing the importance of empathetic interaction for vulnerable populations including elderly patients and those with cognitive impairments.

4 WittyHead Architecture

4.1 System Overview

WittyHead integrates six coordinated services for empathetic human-agent collaboration:

1) **Emotion Detection Service:** Multi-modal emotion recognition from audio prosody (librosa) and facial analysis (DeepFace) [25]
2) **Empathetic Response Orchestrator:** Maps user emotions to avatar responses using therapeutic alliance research, implementing asymmetric (non-mirroring) mappings
3) **Facial Expression Manager:** Generates ARKit blendshape weights from FACS-validated emotion patterns
4) **Gaze Manager:** Controls eye contact percentage, gaze direction, pupil dilation, and blink rate based on therapeutic alliance research
5) **Gesticulation Manager:** Produces prosodic-aligned hand gestures with palm orientation signaling
6) **Voice Modulation Manager:** Adjusts TTS prosody (pitch, rate, volume) for emotional expression

All services coordinate through a central orchestrator ensuring millisecond-precision synchronization at 60 FPS (16.7ms precision).

4.2 Empathetic (Non-Mirroring) Response Mapping

The core innovation is non-mirroring empathetic response mapping derived from Gilbert et al.'s research [4]:

TABLE 2: Empathetic Response Mapping

User Emotion	Avatar Response	Rationale
Sad	Compassionate Concern	Not mirrored sadness
Angry	Calm Concern	De-escalation
Fear	Reassuring	Protective stability
Disgust	Validating	Acknowledge aversion
Happy	Happy	Mirror positive
Surprised	Curious	Engaged interest
Neutral	Neutral	Professional

This asymmetric mapping implements therapeutic research showing that:

- **Negative emotions:** Compassionate concern (not mirrored distress) achieves significantly higher empathy ratings
- **Positive emotions:** Mirrored happiness strengthens interpersonal connection
- **Empathy requires different strategies** for positive vs. negative emotions

4.3 Facial Expression Manager

Generates ARKit blendshape weights from FACS-validated patterns. For compassionate response to user distress:

```
COMPASSIONATE_BLENDSHAPES = {
    # Attention/concern (AU4 - brow lowerer)
    'browDownLeft': 0.4,
    'browDownRight': 0.4,
    'browInnerUp': 0.3,  # Slight inner raise

    # Warmth (AU6 - cheek raiser, AU12 - lip corner puller)
    'cheekSquintLeft': 0.3,
    'cheekSquintRight': 0.3,
    'mouthSmileLeft': 0.2,  # Gentle smile
    'mouthSmileRight': 0.2,

    # Relaxed lower face (not tense)
    'mouthPressLeft': 0.0,
    'jawForward': 0.0
}
```

This pattern implements AU4 + AU6 + gentle AU12 identified by Gilbert et al. [4] as conveying compassionate concern without sharing the user's distress.

4.4 Gaze Manager

Controls four coordinated gaze parameters based on therapeutic alliance research:

1. Eye Contact Percentage: Research-based targets [8]:

- Empathetic response to sadness: 70-80% (high presence)
- Empathetic response to anger: 60-70% (respectful, non-challenging)
- Empathetic response to shame: 30-40% (give privacy)

2. Gaze Direction: Approach-avoidance alignment [9]:

- Direct gaze for approach emotions (anger, joy)
- Averted gaze for avoidance emotions (fear, sadness)

3. Pupil Dilation: Arousal-correlated sizing [13]:

- Base dilation by emotion: Disgust (0.9), Anger (0.85), Fear (0.8), Sad (0.7), Happy (0.5), Neutral (0.3)
- Scaled by detected emotion intensity

4. Turn-Taking Synchronization: Gaze marks speaker changes [14]:

- Floor-holding: Avert gaze (continuing to speak)
- Turn-yielding: Make eye contact (offer turn to user)
- Average mutual gaze: 2.2 seconds

4.5 Gesticulation Manager

Generates prosodic-aligned gestures [10] with three innovations:

1. Prosodic Synchronization: Gestures align with pitch peaks and stressed syllables, not keywords, requiring prosody analysis from TTS service.

2. Palm Orientation: Social signaling [11]:

- Palm-up: Openness/trust (empathetic responses)
- Vertical: Equality/cooperation (neutral explanations)
- Palm-down: Dominance (AVOIDED in empathetic contexts)

3. Emotion-Based Gesture Frequency:

- User anger/distress: 3.5 gestures/min (calm, minimal)
- User sadness/fear: 4.5 gestures/min (gentle, reassuring)
- User happiness: 7.0 gestures/min (energetic, positive)

4.6 Accessibility Features

WittyHead incorporates universal design informed by Miyazaki's disability studies research [17]:

Visual: Screen reader optimization, high-contrast modes, adjustable text sizing

Auditory: Visual alerts, comprehensive text alternatives, caption support

Motor: Voice control, switch access, adjustable interaction timing

Cognitive: Progressive disclosure, simplified language modes, "Easy Japanese" integration [18], [19]

Cultural/Linguistic: Culturally-appropriate emotion expression, multilingual support with accessibility-focused translations

4.7 Integration with Digital MOAI

WittyHead serves as the empathetic interface for Digital MOAI mutual aid networks, providing:

Privacy-Preserving Empathy: Local-first emotion processing on AIngle DLT [22], no centralized emotion data collection

Group-Aware Context: Ontology-enriched understanding of MOAI group dynamics, member relationships, and collective goals

User-Controlled Automation: Three automation levels (propose, notify, act independently) for emotional support interactions

Emergency Emotional Support: Coordinated response when MOAI member experiences distress, maintaining privacy while enabling human group support

5 Implementation and Validation

5.1 Research-Based Validation

Our architecture design is validated through existing therapeutic alliance research. Gilbert et al. [4] demonstrated that compassionate concern expressions achieve significantly higher empathy ratings than mirrored distress in clinical contexts. Their study showed

mirrored negative emotions are perceived as "invalidating," while compassionate responses (AU4 + AU6 + gentle AU12) convey understanding without sharing distress.

Therapeutic alliance research [8] establishes that 60-90% eye contact combined with forward lean enhances perceived empathy. Our gaze management system implements these validated parameters. Virtual counselor studies [16] confirm that nonverbal compassion through coordinated eye contact, facial expressions, and head nodding improves counseling effectiveness.

5.2 Multi-Modal Coordination

The orchestrator ensures temporal coordination:

1) **Facial-Gaze Coordination:** Direct gaze enhances approach emotions; averted gaze enhances avoidance emotions [9]
2) **Gesture-Prosody Coordination:** Beat gestures trigger on pitch peaks with amplitude scaling by speech energy [10]
3) **Voice-Facial Coordination:** Vocal pitch correlates with facial intensity
4) **Gaze-Turn-Taking Coordination:** Eye contact marks turn-yielding; gaze aversion holds conversational floor [14]

Synchronization operates at 60 FPS matching human perceptual thresholds.

5.3 Accessibility Validation

WittyHead achieves WCAG 2.1 AAA compliance [2] across all interaction modalities. Integration of "Easy Japanese" principles [19] enables accessible communication for non-native speakers, persons with cognitive disabilities, and older adults. Disability studies perspectives [17] inform universal design benefiting all users, particularly those in high-stress situations.

5.4 Privacy and Safety

Integration with AIngle DLT provides:

- **Local-first architecture:** Emotion processing on user's device
- **Real-time performance:** 0.16ms average latency [22]
- **No central data aggregation:** Privacy sovereignty for vulnerable populations
- **User-controlled automation:** Explicit consent for AI emotional responses

6 Discussion

6.1 Implications for Multi-Agent System Safety

WittyHead addresses three MASST initiative priorities [3]:

1. Context-Aware Behavioral Guard Rails: Ontology-driven emotion validation prevents inappropriate responses. Medical ontologies block celebratory expressions for serious diagnoses. Empathetic response mappings prevent invalidating emotional mirroring.

2. Mutual Observability: Multi-modal reasoning provides explainable empathy. System logs specify: "User emotion: Sad (0.8 intensity) → Avatar response: Compassionate Concern with 75% eye contact, palm-up gestures, soft concerned expression." Reasoning traces enable human oversight.

3. Design-Time Risk Mitigation: Therapeutic alliance research integration prevents empathy failures before deployment. Validated emotion mappings (compassionate concern for distress) avoid uncanny valley responses.

6.2 Supporting Vulnerable Populations

WittyHead's integration with Digital MOAI demonstrates empathetic agents augmenting human mutual aid networks:

Social Isolation: For individuals experiencing loneliness (foster care youth, elderly, refugees), WittyHead provides emotionally authentic companionship while MOAI group provides human connection.

Mental Health Support: Empathetic responses provide validation and support between therapy sessions, with privacy-preserving local processing.

Disability Accommodation: Universal design enables emotional connection for users with diverse sensory, motor, and cognitive abilities.

Cultural Sensitivity: Ontology-enriched context enables culturally-appropriate emotional expression and "Easy Japanese" accessibility.

6.3 Beyond Emotional Mirroring

Therapeutic research confirms emotional mirroring is insufficient—and often counterproductive—for empathetic agents. Gilbert et al. [4] demonstrated mirrored distress is perceived as "invalidating," while compassionate concern achieves significantly higher empathy ratings. McEwan et al. [5] showed compassionate expressions are distinct from happy faces, designed to convey "safeness and security."

Implication: Empathy systems require asymmetric response mappings—different strategies for positive vs. negative emotions. Simple emotional mirroring creates inappropriate responses that undermine trust, particularly for vulnerable populations experiencing distress.

6.4 Limitations and Future Work

Cultural Adaptation: Current implementation supports Western gaze norms (60-90% eye contact). Eastern cultures prefer 30-60%. Future work: cultural ontology integration for adaptive empathy expression.

Individual Differences: Autism spectrum users may prefer reduced eye contact and gesture frequency. Future work: personalized empathy profiles based on user feedback.

Longitudinal Empathy: Current system operates on single-interaction basis. Future work: multi-session emotional memory enabling relationship-aware empathy.

Empirical Validation: Planned human subjects research with Digital MOAI groups (JSPS KAKENHI Grant JP23K01882) will validate empathetic response effectiveness with vulnerable populations including foster care youth. Future studies will employ Digital

Therapeutic Alliance (DTA) measurement instruments [28], [29] to assess perceived bond, goal alignment, and task agreement with anthropomorphic WittyHead avatars, extending DTA research from text-based chatbots to multi-modal embodied agents.

7 Conclusion

This paper presented WittyHead, a multi-modal empathetic agent architecture designed to support individuals and communities through scientifically-grounded emotional expressivity. Our key insight—empathy requires compassionate concern responses rather than emotional mirroring—derives from therapeutic alliance research demonstrating that mirrored negative emotions are perceived as invalidating while compassionate concern conveys understanding without sharing distress.

Our contributions demonstrate that authentic empathetic human-agent collaboration requires: (1) scientifically-grounded multi-modal architecture validated by FACS, therapeutic alliance, gesture, and prosody research, (2) empathetic (non-mirroring) response mappings implementing asymmetric strategies for positive vs. negative emotions, (3) accessibility-first design incorporating disability studies perspectives and "Easy Japanese" principles, and (4) privacy-preserving integration demonstrating how empathetic agents can augment traditional mutual aid networks.

WittyHead serves as the empathetic interface for Digital MOAI, demonstrating how multi-agent systems can provide authentic emotional support to vulnerable populations—individuals experiencing social isolation, mental health challenges, disability, or community fragmentation—while maintaining privacy sovereignty, cultural sensitivity, and human-centric design. As multi-agent systems increasingly serve healthcare, education, mental wellness, and community support domains, empathetic expressivity becomes critical for trust, engagement, and effective human-agent collaboration.

Future work will validate the architecture through empirical studies with Digital MOAI groups, extend to cultural adaptation and individual personalization, and develop longitudinal emotional memory enabling relationship-aware empathy.

Acknowledgments

The author thanks Eduardo Santaella, Chief Information Officer at T-ROC Global, for early inspiration and for valuable advice and support in evaluating this work's hypotheses in a real-world application. Digital MOAI work is supported by JSPS KAKENHI Grant Number JP23K01882 (PI: Kazuko Kotoku).

References

[1] P. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei, "Deep reinforcement learning from human preferences," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[2] World Wide Web Consortium (W3C), "Web content accessibility guidelines (WCAG) 2.1," W3C Recommendation, June 2018.

[3] J. M. Bradshaw and M. Mahmud, "First International MASST Initiative Workshop: Multi-Agent System Safety and Teamwork," IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology, 2025.

[4] P. Gilbert, C. McEwan, R. Matos, and A. Rivis, "Compassionate faces: Evidence for distinctive facial expressions associated with specific prosocial motivations," *PLOS ONE*, vol. 14, no. 1, e0210283, 2019.

[5] K. McEwan, P. Gilbert, S. Dandeneau, et al., "Facial expressions depicting compassionate and critical emotions: The development and validation of a new emotional face stimulus set," *PLOS ONE*, vol. 9, no. 2, e88783, 2014.

[6] K. Sonnby-Borgström, "Alexithymia as related to facial imitation, mentalization, and internal working models-of-self and -others," *Neuropsychoanalysis*, vol. 11, no. 1, pp. 111-128, 2009.

[7] P. Ekman and W. V. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Palo Alto: Consulting Psychologists Press, 1978.

[8] K. A. Dowell and J. S. Berman, "Therapist nonverbal behavior and perceptions of empathy, alliance, and treatment credibility," *Journal of Psychotherapy Practice and Research*, vol. 3, pp. 214-224, 1994.

[9] R. B. Adams and R. E. Kleck, "Effects of Direct and Averted Gaze on the Perception of Facially Communicated Emotion," *Emotion*, vol. 5, no. 1, pp. 3-11, 2005.

[10] D. P. Loehr, "Temporal, structural, and pragmatic synchrony between intonation and gesture," *Laboratory Phonology*, vol. 3, no. 1, pp. 71-89, 2012.

[11] A. Pease and B. Pease, *The Definitive Book of Body Language*, Bantam, 2006.

[12] K. R. Scherer, "Vocal communication of emotion: A review of research paradigms," *Speech Communication*, vol. 40, no. 1-2, pp. 227-256, 2003.

[13] M. M. Bradley, L. Miccoli, M. A. Escrig, and P. J. Lang, "The pupil as a measure of emotional arousal and autonomic activation," *Psychophysiology*, vol. 45, no. 4, pp. 602-607, 2008.

[14] R. Vertegaal, R. Slagter, G. van der Veer, and A. Nijholt, "Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes," *Proceedings of ACM CHI*, pp. 301-308, 2001.

[15] D. McNeill, *Hand and Mind: What Gestures Reveal about Thought*, University of Chicago Press, 1992.

[16] D. S. Choi, J. Park, M. Loeser, and K. Seo, "Improving counseling effectiveness with virtual counselors through nonverbal compassion involving eye contact, facial mimicry, and head-nodding," *Scientific Reports*, vol. 13, article 5892, 2023.

[17] Y. Miyazaki, "Constructing 'development': A historical discourse analysis of newspapers regarding the creation of terminology and public discourse on autism and hattatsu shogai in japan," Ph.D. dissertation, Kwansei Gakuin University, 2017.

[18] Y. Miyazaki, "Yasashii nihongo (easy japanese) on community media: Focusing on radio broadcasting," *KGPS Review: Kwansei Gakuin Policy Studies Review*, vol. 8, pp. 1-14, March 2007, (in Japanese).

[19] T. Matsuura, A. Yamashita, and N. Iwaoka, "Effectiveness of 'easy japanese' in disaster prevention radio broadcasts," *Journal of Language Education and Multilingualism*, vol. 29, no. 1, pp. 24-25, 2022.

[20] K. Satoh and Sociolinguistics Research Lab, Hirosaki University, "Easy japanese (yasashii nihongo) for disaster information," Hirosaki University Faculty of Humanities, 1995.

[21] K. Kotoku and Y. A. Tijerino, "Artificial intelligence (ai) in nursing practice: Current status and challenges," *Regional Caring (Chiiki Caring)*, vol. 23, no. 4, pp. 39-45, 2021, in Japanese.

[22] "First responder advanced technologies for safe and efficient emergency response," EU H2020 FASTER Project, Grant Agreement No. 833507, 2019-2022.

[23] D. Buettner, *The Blue Zones: Lessons for Living Longer from the People Who've Lived the Longest*, National Geographic Society, 2008.

[24] M. Suzuki, B. J. Willcox, and D. C. Willcox, "Implications from and for food cultures for cardiovascular disease: longevity," *Asia Pacific Journal of Clinical Nutrition*, vol. 10, no. 2, pp. 165-171, 2001.

[25] S. I. Serengil and A. Ozpinar, "LightFace: A hybrid deep face recognition framework," *Innovations in Intelligent Systems and Applications Conference (ASYU)*, pp. 23-27, 2020.

[26] S. D'Alfonso, O. Santesteban-Echarri, S. Rice, et al., "The Digital Therapeutic Alliance and Human-Computer Interaction," *JMIR Mental Health*, vol. 7, no. 11, e21895, 2020, doi: 10.2196/21895.

[27] R. Lederman, T. Wadley, J. Gleeson, S. Alvarez-Jimenez, and M. Alvarez-Jimenez, "The Digital Therapeutic Alliance: Prospects and Challenges," *JMIR Mental Health*, vol. 8, no. 3, e27691, 2021, doi: 10.2196/27691.

[28] C. Beatty, T. Malik, V. Meheli, and S. Sinha, "Evaluating the Therapeutic Alliance With a Free-Text CBT Conversational Agent (Wysa): A Mixed-Methods Study," *Frontiers in Digital Health*, vol. 4, article 847991, 2022, doi: 10.3389/fdgth.2022.847991.

[29] H. L. Tong, L. Quiroz, K. Karyotaki, et al., "Conceptualizing the Digital Therapeutic Alliance in Fully Automated Mental Health Apps: A Thematic Analysis," *Clinical Psychology & Psychotherapy*, vol. 30, no. 6, pp. 1329-1344, 2023, doi: 10.1002/cpp.2898.