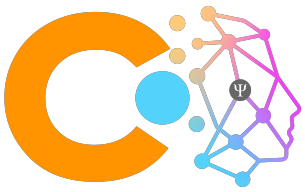

The Clone as Boss

Anthropomorphism Backfire and Blame Attribution in Spatial AR Manager Clones

A within subjects study comparing photorealistic AI manager clones and generic AI agents on Apple Vision Pro

Master's Thesis Exposé

Author: Jan Steinhauer
Supervisor: Dr. David Obremski
Chair: Psychology of Intelligent Interactive Systems
Faculty: Faculty of Human Computer Interaction
University: Julius Maximilians Universität Würzburg
Date: May 2026

1 Introduction and Motivation

The workplace is becoming an increasingly algorithmic environment. Beyond decision support systems and chatbots, recent field research documents a qualitatively new step: the artificial intelligence (AI) *supervisor*. Frontline employees already receive instructions, performance reviews, and disciplinary feedback from AI bots, with measurable effects on motivation, fairness perceptions, and turnover intentions (Lee et al., 2025; Tang et al., 2024).

In parallel, generative AI now enables *photorealistic personal clones*: 3D scanned avatars driven by voice synthesis, large language models, and personalised knowledge bases. Such clones can stand in for their owner during meetings, deliver messages in the owner's likeness, and adopt their communication style (Cheng et al., 2025; Leong et al., 2024). While Leong et al. report that 76 % of collaborators preferred personalised *Dittos* over generic *Delegates*, Lee et al. (2023) simultaneously caution that AI clones trigger *doppelganger phobia*, identity fragmentation, and diffuse accountability.

Two emerging hardware trends amplify the question. First, mass market spatial computing devices such as Apple Vision Pro deliver life size, spatially anchored avatars in the user's own room. Second, GPU streaming pipelines like NVIDIA CloudXR make real time photorealistic rendering on these head mounted displays (HMDs) practical. Yet, despite the convergence of AI managers, personalised proxy agents, and consumer spatial AR, no empirical study has investigated what happens when an employee's *own manager* appears as a photorealistic AI clone in spatial AR, particularly when that clone delivers *negative* feedback or *makes errors*.

Crucially, the cloud streaming architecture chosen for this study is not a one off engineering convenience but a deliberate bet on the *next* generation of HMDs. Industry roadmaps point away from the standalone, compute heavy spatial computer (Vision Pro, Quest 3) and toward light, glasses form factor displays that maximise battery life, thermal headroom, and social acceptability by *offloading* rendering, language model inference, and voice synthesis to a remote backend. The thin client / thick server split realised in this thesis, the device captures audio and head pose, the cloud handles the avatar pipeline, the LLM, and the voice clone, is precisely the architecture such future glasses will require. Empirical results on manager clones obtained on Vision Pro should therefore generalise, with minimal redesign of the backend, to lighter and more ubiquitously wearable AR glasses, which is the form factor in which clone supervision is most likely to enter everyday workplace use.

This exposé proposes such a study. It combines theory on workplace AI supervision (Tang et al., 2024; Yam et al., 2022) and personalised agent proxies (Leong et al., 2024) into a 2 (Agent Type) × 3 (Task Type) within subjects experiment (n=10). The design tests whether the well documented benefits of personalisation survive the asymmetric power context of a real manager subordinate relationship, or whether the *anthropomorphism backfire effect* (Yam et al., 2022) re emerges and is amplified by photorealism.

2 Related Work

2.1 AI Agents as Workplace Supervisors

More and more research in business and technology is looking at how AI is taking on managerial roles. Yam et al. (2022) report the *anthropomorphism backfire effect*: workers behave more *spitefully* toward robot supervisors that appear more human like, because anthropomorphism elevates attributions of agency and therefore perceived malicious intent during punishment. Tang et al. (2024) extend this to language only AI bots in service work, finding that AI supervision lowers job satisfaction and elevates withdrawal behaviour, mediated by reduced perceptions of voice and procedural justice. Lee et al. (2025) survey employee preferences and document a robust *algorithmic aversion* for managerial decisions in relational, but not transactional, contexts. The studies converge on a key prediction: human likeness of an AI supervisor is not monotonically beneficial.

2.2 Personalised Proxy Agents

Leong et al. (2024) introduce *Dittos*: embodied, personalised proxy agents that attend meetings on behalf of an absent colleague. Their evaluation shows that visual and vocal resemblance strongly increase perceived presence, that collaborators trust the proxy's information more than a generic delegate's, and that first person language is a particularly powerful cue of proxy authenticity. Cheng et al. (2025) broaden the design space, conceptualising shared autonomy between user and agent in voice based multitasking. Both lines of work treat the represented person as a peer; neither investigates a hierarchical relationship where the proxy's principal holds formal authority over the participant.

2.3 Risks of AI Clones to Selfhood and Relationships

Lee et al. (2023) surface three risks of AI clones: *doppelganger phobia* (eeriness intensified by familiarity), *identity fragmentation* (clones and their principals diverge in ways the principal cannot control), and *living memories* (clones outlive or drift from the person they represent). Particularly relevant to the present work, the authors highlight an *accountability gap*: when a clone errs, blame attribution is ambiguous between the clone's principal, the system, and the developers.

2.4 Synthesis and Gap

The four threads above predict opposite signed effects of photorealistic personalisation in a managerial context. Personalised proxy work predicts higher trust, presence, and acceptance for a manager clone. Anthropomorphism backfire and doppelganger phobia predict heightened perceived abuse during negative feedback and asymmetric blame attribution after clone errors. No prior study tests these competing predictions in a single within subjects design that holds the human principal constant.

3 Research Gap and Questions

While prior research has examined AI agents in workplace settings (Lee et al., 2025; Tang et al., 2024; Yam et al., 2022) and personalised meeting proxies (Leong et al., 2024), no study has investigated photorealistic manager clones delivered through spatial AR head mounted displays in authentic workplace hierarchies. The proposed thesis addresses this gap with the following research questions:

RQ1

How does agent personalisation (manager clone vs. generic agent) affect employee trust, social presence, and acceptance in AR mediated workplace interactions?

RQ2

Does the anthropomorphism backfire effect (Yam et al., 2022) manifest when a manager clone delivers negative feedback in spatial AR?

RQ3

How do employees attribute blame and accountability when personalised vs. generic agents make errors?

RQ4

Does task type (relational vs. transactional) moderate preferences for manager clones over generic agents?

4 Hypotheses

Eight directional predictions operationalise the four research questions. Given the small sample (Section 5.2), they are framed as *exploratory and hypothesis generating*: the study estimates effect sizes and yields qualitative evidence for follow up confirmatory work, rather than offering definitive null hypothesis significance tests. Primary predictions (H1–H4) target the main effect of agent personalisation; secondary predictions (H5–H8) target interactions with feedback valence, error attribution, preference, and task type. The table below summarises the predictions; the subsections that follow document, hypothesis by hypothesis, the empirical and theoretical reasoning that produced each one and the specific inferential step that links prior findings to the present spatial AR manager clone context.

ID	Statement (Direction: Clone vs. Generic)	Theoretical anchor
H1	Higher *social presence* for the manager clone.	Leong et al. 2024
H2	Higher *trust* and *information credibility* for the clone.	Leong et al. 2024; Mayer et al. 1995
H3	Higher *proxy authenticity* for the clone.	Leong et al. 2024
H4	Higher *uncanny-valley discomfort* for the clone, moderated by relationship quality.	Weidner et al. 2023; Lee et al. 2023
H5	During negative feedback, the clone elicits stronger *perceived abuse* and *negative affect*.	Yam et al. 2022
H6	On agent error, employees attribute more *blame to the real manager* for clone errors than for generic errors.	Lee et al. 2023
H7	Stronger overall *preference* and *intention to use* for the clone.	Leong et al. 2024
H8	Clone preference is stronger for *relational* than for *transactional* tasks (Agent × Task interaction).	Cheng et al. 2025

4.1 Primary Hypotheses (H1–H4)

H1 – Social presence. Leong et al. (2024) report that personalised *Dittos* evoked a "significantly stronger sense of the absent colleague's presence" than generic delegates, with vocal resemblance acting as "a particularly powerful trigger" of co presence. The Networked Minds model (Biocca et al., 2003) traces this effect to identity recognition cues (face, voice, idiosyncratic phrasing) that the perceiver uses to construct a mental model of the *specific other*. The manager clone supplies all three cue classes simultaneously, whereas the generic delegate supplies none. Holding LLM backbone, knowledge, and dialogue scripts constant, any between condition difference in social presence is therefore attributable to identity representation. We predict the effect to replicate in spatial AR on Apple Vision Pro because head mounted stereoscopic rendering anchors the avatar in the user's physical space, which prior HMD work shows further amplifies presence (Weidner et al., 2023). H1 thus serves as a manipulation check that the personalisation manipulation is psychologically operative before the more risk laden secondary hypotheses are tested.

H2 – Trust and information credibility. H2 reflects two converging derivations. First, in the empirical Ditto study (Leong et al., 2024), collaborators "trusted the information more and gave more weight to the agent's opinions" when the agent was personalised. Second, the Mayer et al. (1995) integrative trust model decomposes trustworthiness into ability, benevolence, and integrity; each of these dimensions is normally inferred from *cumulative experience* with a specific person. A photorealistic clone of the participant's actual manager may transfer trust priors built up over prior interactions to the agent, whereas a generic AI assistant must be evaluated from scratch. Information credibility (Flanagin & Metzger, 2000) hinges on perceived source expertise, which is similarly inherited from the cloned identity. We therefore predict higher Trust in Agent and Information Credibility scores for the clone, even when both agents output the identical message. The hypothesis also functions as a constructive replication of Leong et al. (2024) in a hierarchical workplace relationship rather than a peer collaboration, where transferred trust priors might in principle be diluted by power asymmetry.

H3 – Proxy authenticity. Leong et al. (2024) introduce *proxy authenticity* as the extent to which a digital agent is experienced as a faithful stand in for its principal, and identify first person language and visual/vocal resemblance as its strongest antecedents. The Manager Clone condition is engineered to maximise all three antecedents (3D scan, cloned voice, first person narration calibrated to the manager's communication patterns), whereas the Generic Agent is engineered to minimise them (stylised avatar, synthetic voice, neutral third person register). The custom Proxy Authenticity scale (developed for this study) measures this construct with six items that probe identity recognition, stylistic congruence, and the participant's subjective sense of conversing with the manager rather than with an AI. H3 is the most direct test of whether the *Ditto* construct generalises from peer absentee proxies to hierarchical clone supervision, and it provides the validity backbone for any downstream interpretation of H5 and H6.

H4 – Uncanny valley discomfort, moderated by relationship quality. H4 inverts the optimistic predictions of H1–H3. Two literatures predict elevated eeriness for the clone. First, the visualisation review by Weidner et al. (2023) shows that high fidelity HMD avatars consistently raise eeriness ratings relative to stylised counterparts, in line with Mori et al. (2012) and the index refinements of Ho and MacDorman (2017). Second, Lee et al. (2023) introduce *doppelganger phobia*, the finding that AI clones become more disturbing as familiarity with the cloned person increases. Combined, these two threads imply both a main effect (clone > generic on the Eeriness subscale of the Uncanny Valley Index) and a moderation by Leader Member Exchange quality (Graen & Uhl-Bien, 1995): closer manager subordinate relationships should magnify, not attenuate, the uncanny response, because the participant has more accurate priors against which the clone can deviate. H4 is critical because, paired with H1, it articulates the central tension of the thesis: spatial AR personalisation simultaneously increases the cues that drive presence and the cues that drive eeriness. H1 and H4 are therefore not mutually exclusive; the literature shows that presence and eeriness correlate positively at high realism levels. We will report the *joint distribution* of presence and eeriness scores across participants, not only their marginal means, so that participants who score high on both can be identified.

4.2 Secondary Hypotheses (H5–H8)

H5 – Anthropomorphism backfire during negative feedback. Yam et al. (2022) demonstrate the anthropomorphism backfire effect: workers behave more spitefully toward, and perceive more malicious intent in, robot supervisors that appear more human like, because human likeness raises perceived agency and therefore attributions of *intentionality* to negative actions. In their studies, the manipulated cue is generic anthropomorphism (eyes, name, voice). The present design escalates this manipulation to its logical upper bound: an avatar that not only *looks human* but is the participant's actual manager. Mind perception theory (Gray et al., 2007) predicts that the Agency dimension of mind attribution will be highest in the clone condition, which by the Yam et al. logic should magnify perceived abusiveness when the *same* critical message is delivered. H5 therefore predicts a condition main effect specifically within Task 2 (Performance Feedback): higher Perceived Supervisor Abuse (adapted from Yam et al., 2022) and higher PANAS Negative Affect (Watson et al., 1988) in the clone condition, despite identical script content. A null result would indicate that the backfire effect is bounded by abstraction

and does not survive personalisation, which would itself be a theoretically informative boundary condition.

H6 – Asymmetric blame attribution after errors. Lee et al. (2023) identify an *accountability gap* in clone interactions: when a clone errs, blame can plausibly be assigned to the principal, the system, or the developers, and participants in their speculative study expressed concern that clone errors would damage the principal's reputation. We move the question from speculation to measurement by injecting a controlled factual error into the Advisory task in both conditions and administering a custom Blame Attribution scale plus a forced allocation Responsibility Pie Chart. The first person framing of the clone is hypothesised to act as a *principal signature*: it binds the utterance to the manager's identity, which by attribution theory increases the likelihood that the error is encoded as a property of the manager rather than of the system. The generic agent's third person delivery affords no such signature, so blame should default to the AI system or its developers. H6 is consequential for organisational deployment because it predicts that the same technical malfunction can carry different reputational costs depending on the personalisation choice.

H7 – Overall preference and intention to use. Leong et al. (2024) report that 76 % of collaborators preferred the personalised Ditto over the generic Delegate when forced to choose. We predict the same direction here, but with a non trivial caveat: the Leong et al. sample evaluated proxies of *absent peers*, not of present superiors. Two opposing forces shape H7. The presence, trust, and authenticity advantages hypothesised in H1–H3 should pull preference toward the clone. The eeriness and abuse mechanisms in H4 and H5 should pull it toward the generic agent. H7 predicts that the first set dominates on the aggregate forced choice and Net Promoter style ratings, because preference questions are typically dominated by the most recent, most positively valenced sub task (Information Sharing), not by the high stakes feedback episode. The strength of H7 (or its absence) is therefore informative about which mechanism generalises better to repeated organisational use.

H8 – Task type moderation (Agent × Task interaction). Cheng et al. (2025) show, in the shared autonomy voice domain, that proxy agents are evaluated most favourably in contexts that mix social and transactional content, and least favourably in purely social contexts where authentic human contact is normatively expected. Lee et al. (2025) corroborate this with employee preference data: algorithmic aversion is strongest for relational managerial decisions, weakest for transactional ones. Combining the two findings yields a non monotonic prediction. For the present three task design, we expect (i) a small or null clone advantage on Information Sharing (purely transactional, where the generic agent is already "good enough"); (ii) the largest clone advantage on Advisory / Problem Solving (which mixes expertise, judgement, and relational rapport, the sweet spot identified by Cheng et al.); and (iii) an attenuated or even reversed clone advantage on Performance Feedback (high relational stakes, where the abuse mechanism of H5 is active).

5 Method

5.1 Design

The study uses a 2×3 within-subjects factorial design. **Agent Type** (Manager Clone vs. Generic AI Agent) and **Task Type** (Information Sharing vs. Performance Feedback vs. Advisory) are manipulated within participants. Agent Type order is counterbalanced across participants (AB vs. BA, 5+5). Tasks within each condition are presented in a fixed narrative order from low to high to medium stakes (Information Sharing, Performance Feedback, Advisory) so that the high stakes feedback episode is bracketed by lower stakes interactions; parallel A/B content versions across the two agent conditions avoid item level confounds.

5.2 Participants

The target sample is n=10 full time employees who are direct reports of a single participating manager, with at least three months of working relationship. Exclusion criteria are an active performance improvement plan, planned departure within three months, severe motion sickness. Recruitment occurs through internal channels with explicit safeguards that responses are not shared with the manager. The within subjects design maximises statistical power given the low n, and the design is explicitly framed as effect size oriented (see Section 5.7).

5.3 Conditions

Both agents share the same large-language model backbone, knowledge base, and factual content; they differ *only* in identity representation.

	Manager Clone (Ditto)	**Generic Agent (Delegate)**
Visual	Photorealistic 3D scan of the manager	Stylised humanoid avatar (gender-neutral)
Voice	Cloned voice (e.g. ElevenLabs)	High quality synthetic neutral voice
Animation	Real time lip sync, generic gestures	Real time lip sync, generic gestures
Language	First person ("In my experience . . . ")	Neutral ("The recommendation is . . . ")
Style	Calibrated to the manager's patterns	Professional, neutral, consistent

5.4 Apparatus

The system streams photorealistic avatars to an Apple Vision Pro via NVIDIA CloudXR. A Google Cloud virtual machine hosts the Unity rendering pipeline, an ElevenLabs voice cloning and speech to text service, an LLM that supplies the dialogue policy, a Firebase backend that persists task state and per session performance traces, and a central *Manager Agent* module that coordinates these services. Spoken user input is transcribed on the device, forwarded to the Manager Agent, dispatched to the LLM and voice cloning service, and the resulting animated

avatar is rendered in Unity, streamed via CloudXR, and displayed life size in the user's physical room on the Vision Pro. The full architecture is shown in Figure 1.

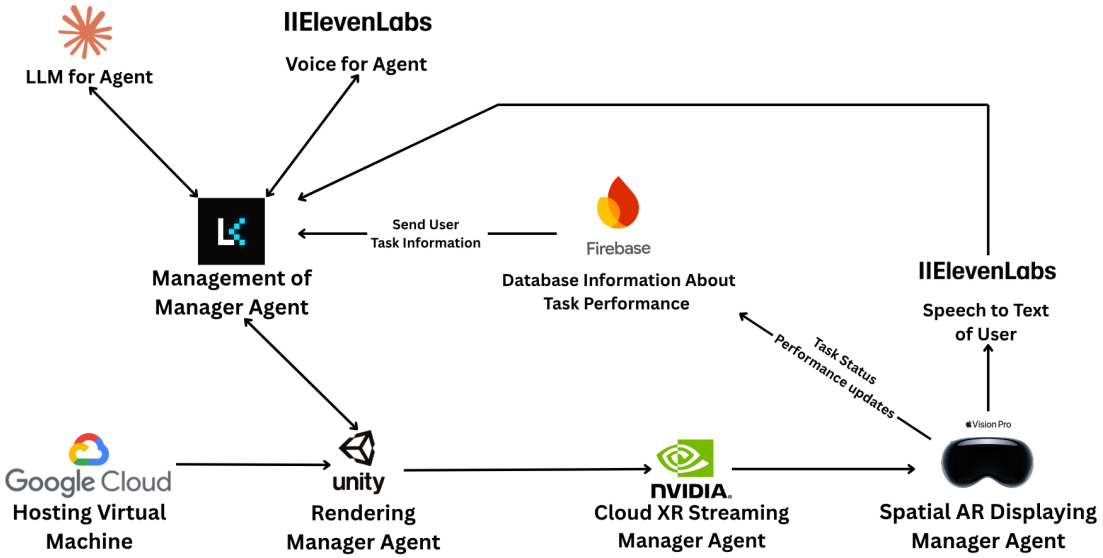

Figure 1: System architecture. A Google Cloud virtual machine hosts the Manager Agent orchestrator, which integrates an LLM (dialogue policy), ElevenLabs (voice synthesis and speech to text), and a Firebase database (task state and performance logs). Unity renders the photorealistic or stylised avatar; the rendered frames are streamed to the Apple Vision Pro via NVIDIA CloudXR, while the device returns the user's head pose and transcribed speech in real time.

Forward compatibility. The architecture is deliberately split into a *thin client* on the head mounted display and a *thick server* in the cloud. Heavy workloads, photogrammetric avatar rendering in Unity, neural voice synthesis, large language model inference, and persistence, all run remotely on a Google Cloud virtual machine. The Vision Pro itself only has to capture audio, transcribe speech, transmit head pose, and decode the incoming CloudXR video stream. This division of labour matters beyond the present study. A central trend in consumer HMD development is the move from compute heavy *spatial computers* toward light, glasses form factor displays that maximise battery life, thermal headroom, and wearability by *offloading* computation. Cloud streamed rendering pipelines such as the one used here scale naturally to such future devices: the same backend can drive a future pair of lightweight AR glasses without redesign, because the device is only a display and a sensor, not a renderer. Conversely, the same backend can be upgraded to larger or more capable language models, higher fidelity voice clones, or more realistic avatar rigs without touching the client. The thesis therefore contributes results that should remain relevant as HMD hardware continues to shrink and as the cloud side of the stack continues to grow.

5.5 Procedure

A pre study session with the participating manager (~2 h, one time) captures the photogrammetry scan, voice samples, a knowledge interview, and clone validation. Each participant session lasts ~70 minutes and follows seven phases: (1) informed consent and baseline questionnaires; (2) Vision Pro familiarisation with a neutral tutorial agent; (3) the first counterbalanced condition with the three tasks and a post condition questionnaire; (4) a five minute washout break; (5) the second condition with parallel content tasks and questionnaire; (6) a direct comparative assessment block; (7) a semi structured exit interview with debriefing.

5.6 Measures

All measures are validated where available; two custom instruments (*Proxy Authenticity*, *Blame Attribution*) extend the Leong et al. (2024) and Lee et al. (2023) frameworks to the manager clone context. Because the main study is underpowered for psychometric validation, both custom instruments are treated as *exploratory indices*: items are reviewed in a small cognitive interviewing pilot ($n \approx 15$–20) external to the main sample, internal consistency is reported as Cronbach's α, and analysis proceeds at the item level as well as on a mean composite. Exploratory factor analysis is reported only if a larger validation sample becomes available; otherwise the indices are interpreted descriptively, not as validated scales. Table 1 summarises the instruments.

Table 1: Core measurement instruments. Pre-study measures are collected once; post-condition measures are administered after each of the two agent conditions.

Phase	**Construct (instrument)**	**Source**
Pre	Demographics	custom
Pre	AI literacy (MAILS)	Carolus et al. 2023
Pre	LMX-7 (relationship with manager)	Graen & Uhl-Bien 1995
Pre	Trust in Automation; TRI 2.0	Parasuraman & Colby 2015
Pre	Individual anthropomorphism tendency (IDAQ)	Waytz et al. 2010
Post	Networked Minds Social Presence	Biocca et al. 2003
Post	IPQ Spatial Presence	Schubert et al. 2001
Post	Trust in Agent (TiA)	Körber 2019
Post	Message Credibility	Appelman & Sundar 2016
Post	Robotic Social Attributes (RoSAS)	Carpinella et al. 2017
Post	Mind Perception (Agency)	Gray et al. 2007
Post	Proxy Authenticity (custom, exploratory index)	developed for this study
Post	NASA-TLX	Hart & Staveland 1988
Post	PANAS (Negative subscale)	Watson et al. 1988
Task 2	Perceived Supervisor Abuse (adapted)	adapted from Yam et al. 2022
Error	Blame Attribution (custom, exploratory index)	developed for this study

5.7 Data Analysis Plan

The analysis is explicitly *exploratory and effect size oriented* and combines two data sources: the questionnaire battery of Section 5.6 and a *semi structured exit interview* administered in Phase 7

of the session. Within subjects quantitative comparisons (H1–H7) are evaluated with paired samples t-tests, with the Wilcoxon signed rank test as a non parametric fallback if normality is violated; H8 is evaluated with a 2×3 repeated measures ANOVA. Effect sizes (Cohen's d_z, partial η^2) and bias corrected bootstrap confidence intervals are reported for all comparisons regardless of statistical significance, since the present sample is powered (80 % at α=.05) only for very large effects. The corresponding minimum detectable effect (MDE) at n=10, 80 % power, two sided α=.05 for a paired comparison is $d_z \approx 0.94$; for the 2×3 repeated measures interaction it is $f \approx 0.49$ (partial $\eta^2 \approx 0.19$). Effects below these thresholds will not reach significance in the main study by design and are therefore reported as estimates with confidence intervals to inform a future, adequately powered confirmatory replication. Leader Member Exchange quality (Graen & Uhl-Bien, 1995) is added as a covariate where statistically defensible, and descriptive patterns are reported by LMX tertile.

Semi structured exit interview. A 20–30 minute interview (audio recorded, verbatim transcribed, pseudonymised) follows a *funnel* protocol with eight thematic blocks that map one to one onto the hypotheses: overall impressions, social presence and identity recognition (H1, H3), trust and information credibility (H2), uncanny valley and eeriness (H4), the negative feedback experience (H5), the error episode and blame attribution (H6), preference and task fit (H7, H8), and design and ethical implications. Each block uses three to five open questions with non leading probes (e.g. "Can you give a concrete example?", "What did the agent do or say right before that shift?").

6 Ethical Considerations

The study involves biometric data (3D scan, voice clone) and an asymmetric workplace relationship; both warrant heightened care. The manager provides explicit, revocable consent for clone creation, reviews the clone before deployment, and is fully informed about participant protections. Participants give informed consent that addresses (a) interaction with their manager's digital replica, (b) the right to withdraw without consequence, and (c) the strict confidentiality of individual responses, which are never shared with the manager and reported only in aggregate. All biometric data are handled under GDPR (lawful basis, purpose limitation, encrypted storage, and a defined destruction timeline). Feedback scenarios are clearly framed as simulated, the design includes a debriefing phase, and signposting to support resources is provided in case of distress. Approval from the JMU Institutional Review Board is obtained prior to data collection.

7 Expected Contributions

Theoretical. The thesis (i) extends Leong et al. (2024)'s Ditto framework from peer meetings to hierarchical managerial contexts in spatial AR; (ii) provides a direct empirical test of Yam et al. (2022)'s anthropomorphism backfire effect under photorealistic personalisation rather than abstract robot embodiment; and (iii) operationalises Lee et al. (2023)'s identity fragmentation and accountability gap concepts via the new *Blame Attribution* scale.

Practical. The work delivers (i) design guidelines for AI manager representations on consumer spatial AR HMDs, (ii) context specific recommendations on *when* a manager clone is preferable to a generic agent (and when not), and (iii) implications for asynchronous and remote management practices that are increasingly common in distributed workplaces.

8 Limitations

The study trades external generalisability for internal control. n=10 and a single manager limit between subjects inference; the within subjects design and effect size focus are explicit mitigations. Novelty effects on the Vision Pro are addressed through familiarisation and washout phases plus a post hoc novelty check. Demand characteristics are reduced through neutral framing and a hypothesis guessing item. The lab setting compromises ecological validity, mitigated by tasks and scripts derived from authentic work scenarios. These limitations frame the study as hypothesis generating for follow up multi organisation field work.

References

Appelman, A., & Sundar, S. S. (2016). Measuring message credibility: Construction and validation of an exclusive scale. *Journalism & Mass Communication Quarterly*, *93*(1), 59–79. https://doi.org/10.1177/1077699015606057

Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators and Virtual Environments*, *12*(5), 456–480. https://doi.org/10.1162/105474603322761270

Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E., & Wienrich, C. (2023). MAILS – meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change- and meta-competencies. *Computers in Human Behavior: Artificial Humans*, *1*(2), 100014. https://doi.org/10.1016/j.chbah.2023.100014

Carpinella, C. M., Wyman, A. B., Perez, M. A., & Stroessner, S. J. (2017). The robotic social attributes scale (RoSAS): Development and validation, 254–262. https://doi.org/10.1145/2909824.3020208

Cheng, Y.-F., Shirado, H., & Kasahara, S. (2025). Conversational agents on your behalf: Opportunities and challenges of shared autonomy in voice communication for multitasking. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–18. https://doi.org/10.1145/3706598.3713251

Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of Internet information credibility. *Journalism & Mass Communication Quarterly*, *77*(3), 515–540. https://doi.org/10.1177/107769900007700304

Graen, G. B., & Uhl-Bien, M. (1995). Relationship-based approach to leadership: Development of leader-member exchange (LMX) theory of leadership over 25 years: Applying a multi-level multi-domain perspective. *The Leadership Quarterly*, *6*(2), 219–247. https://doi.org/10.1016/1048-9843(95)90036-5

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619. https://doi.org/10.1126/science.1134475

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183, Vol. 52). North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9

Ho, C.-C., & MacDorman, K. F. (2017). Measuring the uncanny valley effect: Refinements to indices for perceived humanness, attractiveness, and eeriness. *International Journal of Social Robotics*, *9*(1), 129–139. https://doi.org/10.1007/s12369-016-0380-9

Körber, M. (2019). Theoretical considerations and development of a questionnaire to measure trust in automation. *823*, 13–30. https://doi.org/10.1007/978-3-319-96074-6_2

Lee, P. Y.-K., Ma, N. F., Kim, I. J., & Yoon, D. (2023). Speculating on risks of AI clones to selfhood and relationships: Doppelganger-phobia, identity fragmentation, and living memories. *Proceedings of the ACM on Human-Computer Interaction (CSCW1)*, *7*, 1–28. https://doi.org/10.1145/3579524

Lee, S. Y., Puranam, P., & Mai, K. M. (2025). Would you let an AI be your boss? Employee preferences for AI versus human managers. *Working Paper, INSEAD.*

Leong, J., Tang, J., Cutrell, E., Junuzovic, S., Baribault, G. P., & Inkpen, K. (2024). Dittos: Personalized, embodied agents that participate in meetings when you are unavailable. *Proceedings of the ACM on Human-Computer Interaction (CSCW2)*, *8*, 1–28. https://doi.org/10.1145/3686964

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics & Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Parasuraman, A., & Colby, C. L. (2015). An updated and streamlined technology readiness index: TRI 2.0. *Journal of Service Research*, *18*(1), 59–74. https://doi.org/10.1177/1094670514539730

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments*, *10*(3), 266–281. https://doi.org/10.1162/105474601300343603

Tang, P.-M., Koopman, J., Mai, K. M., De Cremer, D., Zhang, J. H., Jörling, M., & Yam, K. C. (2024). When your boss is an AI bot: Frontline employees' reactions to AI supervision [Advance online publication]. *Journal of Applied Psychology.*

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Waytz, A., Cacioppo, J. T., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, *5*(3), 219–232. https://doi.org/10.1177/1745691610369336

Weidner, F., Boettcher, G., Arboleda, S. A., Diao, C., Sinani, L., Kunert, C., Gerhardt, C., Broll, W., & Raake, A. (2023). A systematic review on the visualization of avatars and agents in AR & VR displayed using head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, *29*(5), 2596–2606. https://doi.org/10.1109/TVCG.2023.3247072

Yam, K. C., Goh, E.-Y., Fehr, R., Lee, R., Soh, H., & Gray, K. (2022). When your boss is a robot: Workers are more spiteful to robot supervisors that seem more human. *Journal of Experimental Social Psychology*, *102*, 104360. https://doi.org/10.1016/j.jesp.2022.104360