One Size Fits None: Why AI Distrust in Education Depends on Who You Ask

Martha Hubertz
Department of Psychology, University of Central Florida
Orlando, Florida, USA
Email: martha.hubertz@ucf.edu

Alisha Janowsky
Department of Psychology, University of Central Florida
Orlando, Florida, USA
Email: alisha.janowsky@ucf.edu

Abstract — **Adaptive learning systems and AI-powered educational tools are increasingly deployed across diverse student populations, yet they are typically designed and validated with majority students and assumed to work uniformly for all learners. We extend algorithmic bias concerns to the psychological level: if the mental processes by which students form AI trust differ across cultural backgrounds, seemingly neutral design choices may inadvertently widen achievement gaps. Using data from over one thousand university students at a Hispanic-Serving Institution, we find that the anthropomorphism paradox, the counterintuitive finding that viewing technology as human-like predicts lower AI trust, operates robustly among non-UnderRepresented Minority (non-URM) students but is entirely absent among URM students, despite both groups showing identical mean levels of anthropomorphism and AI trust. Additionally, students who received smartphones before age 16 show meaningfully stronger superstitious beliefs and lower AI trust than later adopters. These findings reveal that cultural background and developmental technology exposure create fundamentally different psychological pathways to AI trust, with direct implications for designing equitable adaptive learning systems.**

Keywords- educational AI; adaptive learning; AI trust; anthropomorphism; educational equity; personalized learning; cultural moderation.

I. INTRODUCTION

The Problem: One-Size-Fits-All AI in Diverse Classrooms

Adaptive learning systems and AI-powered educational tools are rapidly becoming standard infrastructure in higher education. Students interact with AI tutors for homework help, receive personalized content recommendations from adaptive platforms, and obtain automated feedback on written work. These systems promise to democratize access to high-quality, individualized instruction at scale, a vision particularly compelling for large public universities serving economically and culturally diverse student populations.

Yet this promise rests on the largely untested assumption that all students respond to AI features in psychologically similar ways. When educational AI systems are designed and validated primarily with majority student populations, then deployed universally without considering cultural or developmental diversity, they risk creating unequal psychological barriers that undermine the very equity goals they aim to serve.

A. Why This Matters: Equity Beyond Access

The digital divide has traditionally focused on access, whether students have devices, connectivity, and technical skills. But equity in educational AI requires more than equal access; it requires that the psychological architecture of these systems works equitably for all learners. If design features that enhance trust and engagement for majority students simultaneously create barriers for UnderRepresented Minorities, then widespread AI deployment may inadvertently widen rather than close achievement gaps.

Recent evidence suggests this concern is not hypothetical. Algorithmic bias research has documented that automated systems consistently produce differential outcomes for Black, Hispanic/Latino, and other UnderRepresented Minority (URM) students, even when systems appear technically neutral [2][3]. These disparities emerge not only from biased training data but from the interaction between system design and diverse user populations. We propose that an analogous phenomenon operates at the psychological level: the mental processes by which students form trust in AI may differ fundamentally across cultural backgrounds, making ostensibly neutral design choices, such as conversational interfaces or anthropomorphic features, inequitable in practice.

B. The Research Gap: Do Trust Mechanisms Generalize?

Prior research has established what we term the "anthropomorphism paradox:" students who attribute human-like qualities to technology (e.g., believing "the computer hates me") paradoxically report lower rather than higher trust in AI systems [4]. This counterintuitive pattern is rooted in superstitious thinking: students with stronger beliefs about bad luck and unpredictable forces anthropomorphize technology more, and that anthropomorphism erodes AI trust.

However, virtually all evidence for this pathway comes from aggregate analyses that implicitly assume uniform psychological mechanisms across all students. No prior research has tested whether the anthropomorphism paradox operates identically for UnderRepresented Minority (URM) and non-UnderRepresented Minority (non-URM) students, nor whether the developmental timing of technology exposure shapes these trust formation processes.

C. Research Questions and Contributions

This paper addresses two questions with direct implications for adaptive learning system design:

Hypothesis 1 (Cultural Moderation): Does the anthropomorphism paradox operate uniformly, or does it differ between UnderRepresented Minority (URM) and non-UnderRepresented Minority (non-URM)?

Hypothesis 2 (Developmental Timing): Does the age at which students first accessed smartphones predict the superstitious beliefs and AI trust they bring to educational AI systems?

Our contributions are threefold:

1. Theoretical: We demonstrate that the anthropomorphism paradox is culturally bound rather than universal, operating robustly among non-URM students but being entirely absent among URM students, despite both groups showing identical mean levels of anthropomorphism and AI trust. This "moderation without mean differences" pattern is invisible to aggregate analyses and represents a novel finding in educational technology research.

2. Developmental: We provide the first empirical evidence that early smartphone access (before age 16) predicts lasting differences in superstitious thinking and AI trust, with medium-sized effects. This finding is particularly timely given recent policy debates in Spain, Greece, and other nations about restricting youth access to algorithmic systems.

3. Practical: We offer actionable design principles for adaptive learning systems, emphasizing the need for configurable interaction styles, disaggregated evaluation metrics, and AI literacy scaffolding that addresses superstition-based anthropomorphism.

II. Related Work

A. Anthropomorphism and Technology Trust

Anthropomorphism, attributing human characteristics, emotions, or intentions to non-human entities, is a fundamental feature of human cognition that varies systematically across individuals and situations [1]. Epley et al. [6] proposed a three-factor theory identifying psychological determinants: the accessibility of anthropocentric knowledge, motivation to understand agent behavior (effectance motivation), and desire for social connection (sociality motivation). These factors predict when people see nonhuman agents, including AI systems, as possessing human-like minds.

In educational contexts, anthropomorphism has emerged as a significant driver of student attitudes toward AI tools. Polyportis and Pahos [7] found that anthropomorphism, alongside trust and perceived novelty, predicts AI chatbot adoption among university students. However, the relationship between anthropomorphism and trust is not straightforward. Jose and Thomas [8] highlighted that digital anthropomorphism shapes epistemic trust in AI tutors in complex ways not yet well understood, particularly for underrepresented learners who may calibrate trust based on different cues than majority students.

1) The Anthropomorphism Paradox

Recent research has revealed a counterintuitive pattern: attributing human-like qualities to AI can reduce rather than enhance trust. Janowsky and Hubertz [4] documented that students who anthropomorphize technology, viewing computers as capable of "hating" users or "seeing all", report lower trust in AI systems. This "anthropomorphism paradox" challenges the widespread assumption in educational technology design that making AI seem more human-like universally improves acceptance and engagement.

The psychological origins of this paradox lie in superstitious thinking. Risen [5] demonstrated that superstitious beliefs persist even when individuals cognitively recognize them as irrational, suggesting they reflect deep-seated intuitive processes rather than simple ignorance. Students high in beliefs about bad luck and unpredictable forces [9] anthropomorphize technology more, experiencing AI systems as unpredictable social agents rather than controllable tools.

B. Superstitious Beliefs and Locus of Control

Fluke and colleagues [9] developed the Belief in Superstition Scale (BSS), demonstrating that superstitious beliefs cluster into three distinct components: belief in bad luck, belief in good luck, and belief that luck can be changed. Critically, they found that external locus of control, the belief that life outcomes are determined by forces beyond personal control, consistently predicts all three types of superstitious beliefs.

The connection between superstition and technology attitudes suggests a hierarchical pathway: external locus of control predicts superstitious beliefs, which predict anthropomorphic interpretations of technology, which in turn predict trust in AI systems. However, this pathway has only been documented in aggregate samples, leaving open the question of whether it operates uniformly across diverse student populations.

C. Algorithmic Bias and Educational Equity

Research on algorithmic bias has documented extensive evidence that automated systems often produce discriminatory outcomes for minority populations. Buolamwini and Gebru's [10] landmark "Gender Shades" study revealed that commercial facial recognition systems showed error rates up to 34.7% for darker-skinned females while achieving near-perfect accuracy (0.8% error) for lighter-skinned males.

In educational contexts specifically, Baker and Hawn [2] found that across predictive models, AI-powered assessments, and recommendation engines, algorithms consistently produced differential outcomes for Black, Hispanic/Latino, and other URM students. Gándara et al. [3] documented that predictive models used in college student-success systems produced false negatives for 19% of Black students and 21% of Latinx students, systematically under-identifying learners they were designed to support.

These documented experiences with algorithmic bias may fundamentally shape how different student populations calibrate trust in AI systems. If URM students have observed algorithmic systems disadvantaging their communities in practice, their trust formation mechanisms may prioritize factors like perceived fairness, transparency, and historical reliability over anthropomorphic features.

D. The Critical Gap: Universal Mechanisms Assumption

Despite evidence of algorithmic bias and differential system performance across student populations, research on the psychological mechanisms underlying AI trust continues to assume universal processes. Studies documenting relationships between anthropomorphism, superstition, and technology trust have examined aggregate patterns without testing for moderation by race, ethnicity, or other diversity dimensions.

Similarly, developmental psychology research suggests that timing of technology exposure during formative periods may shape lasting attitudes toward digital systems, yet no prior work has examined whether age of first smartphone access predicts AI trust in educational contexts.

Our study addresses these gaps by testing whether the anthropomorphism paradox operates uniformly across URM and non-URM students, and whether developmental timing of smartphone access shapes superstitious beliefs and AI trust.

III. METHODS

A. Study Design Overview

We conducted a cross-sectional survey study to test two primary hypotheses: (1) that the relationship between technology anthropomorphism and AI trust differs between UnderRepresented Minority (URM) and non-URM university students (cultural moderation hypothesis), and (2) that age of first smartphone access predicts superstitious beliefs and AI trust in young adulthood (developmental timing hypothesis).

The study employed a correlational design with both continuous and categorical predictor variables. Cultural background (URM vs. non-URM) and smartphone access timing (before vs. at/after age 16) served as moderator variables, while technology anthropomorphism, superstitious beliefs, and AI trust were measured as continuous outcome variables. We also examined the hierarchical pathway from external locus of control through superstitious beliefs to anthropomorphism to AI trust, testing whether this pathway operates similarly across cultural groups.

Data collection occurred during Fall 2023 through Spring 2024 via an online survey platform. The study was approved by the University of Central Florida Institutional Review Board (protocol #STUDY00005234), and all participants provided informed consent before beginning the survey.

B. Participants

1) Sample Characteristics

Participants were 1,331 undergraduate students enrolled at the University of Central Florida (UCF), a large public research university and designated Hispanic-Serving Institution in Orlando, Florida. UCF enrolls approximately 69,000 students and is one of the most diverse universities in the United States, making it an ideal context for examining cultural moderation in educational technology attitudes.

Of the initial 1,331 participants, 1,306 (98.1%) completed all measures and passed embedded attention checks, forming the final analytic sample. The sample was predominantly female (60.4% female, 38.1% male, 1.5% non-binary or other) with a mean age of 19.48 years (SD = 3.44, range 18–47). First-generation college students, those whose parents did not complete a four-year college degree, comprised 35.2% of the sample, reflecting UCF's mission to serve diverse student populations.

2) Cultural Background Classification

For cultural moderation analyses, participants were classified into two groups based on self-reported race/ethnicity. UnderRepresented Minority (URM) students (n = 234, 17.9%) identified as Black/African American, Hispanic/Latino, Native American, Pacific Islander, or multiracial with at least one URM component, aligning with standard definitions in higher education research and federal reporting. Non-URM students (n = 1,072, 82.1%) identified as White and/or Asian. This classification reflects the documented finding that Asian students in STEM and higher education contexts do not face the same systemic barriers as other minority groups, though we acknowledge this remains a debated categorization.

We recognize that the URM category aggregates diverse ethnic and cultural groups whose relationships with algorithmic systems likely differ. This aggregation was necessary given sample size constraints but represents a limitation we address in the Discussion.

3) Smartphone Access Timing

Participants reported the age at which they first owned or had regular access to a smartphone. Responses ranged from age 5 to age 25, with a mean of 12.30 years (SD = 2.30). For primary analyses, we dichotomized this variable as early access (before age 16; n = 1,235, 94.6%) or late access (at or after age 16; n = 71, 5.4%). The age-16 threshold was selected based on two considerations: (1) it aligns with current policy debates in Spain, Greece, and other nations considering social media age restrictions, and (2) it corresponds to mid-adolescence, a developmentally sensitive period for forming abstract reasoning about agency and causality. We also conducted supplementary analyses treating smartphone access age as a continuous variable to examine dose-response relationships.

C. Measures

All measures used Likert-type rating scales and demonstrated acceptable-to-excellent internal consistency reliability. Internal consistency was assessed using

Cronbach's alpha (α), a statistical measure of how consistently items within a scale measure the same underlying construct. Alpha values above .70 are generally considered acceptable, above .80 good, and above .90 excellent.

1) Technology Anthropomorphism

Technology anthropomorphism was measured using a 3-item composite drawn from the Technology Superstition Scale [4]. Items were selected specifically to capture attributions of human-like qualities to computers and smartphones: (1) "The computer hates me," (2) "The computer sees all and knows all," and (3) "My phone sees all and knows all." Participants rated their agreement on a 5-point scale (1 = Strongly Disagree to 5 = Strongly Agree). Items were averaged to create an overall anthropomorphism score (α = .82). These items capture attributions of intentionality, omniscience, and agency to technological devices — distinct from general anthropomorphism scales by their focus on magical or superstitious thinking about technology possessing human-like awareness.

2) Superstitious Beliefs

Superstitious beliefs were assessed using the 18-item Belief in Superstition Scale (BSS; [9]), consisting of three 6-item subscales. The *Bad Luck subscale* (α = .85) measures beliefs that certain actions or events bring negative outcomes (e.g., "Breaking a mirror brings bad luck," "Friday the 13th is an unlucky day"). The *Good Luck subscale* (α = .89) measures beliefs that certain objects or actions bring positive outcomes. The *Change Luck subscale* (α = .76) measures beliefs that one can actively manipulate luck. Participants rated each item on a 7-point scale (1 = Strongly Disagree to 7 = Strongly Agree); the overall BSS demonstrated excellent internal consistency (α = .90). We focused primarily on Bad Luck beliefs given their theoretical relevance to threat-oriented anthropomorphism. Fluke and colleagues [9] demonstrated that Bad Luck beliefs correlate most strongly with neuroticism and external locus of control, whereas Good Luck beliefs correlate with agreeableness and Change Luck beliefs with proactive coping.

3) AI Trust

Trust in artificial intelligence was measured using a 16-item scale covering four conceptual domains: (1) *Security/Privacy* — trust that AI systems protect personal information and maintain confidentiality (4 items); (2) *Validity/Reliability* — trust that AI systems produce accurate and consistent outputs (4 items); (3) *Capability* — trust that AI systems can successfully perform intended tasks (4 items); and (4) *Understandability* — trust that AI system operations are transparent and comprehensible (4 items). Example items include "I trust AI to keep my personal information secure," "I trust AI to provide reliable recommendations," and "I trust AI to accurately assess my work." Participants rated each item on a 7-point scale (1 = Strongly Disagree to 7 = Strongly Agree). Items were averaged to create an overall AI trust score (α = .91), indicating the four domains cohere into a unified construct. This multidimensional measure is particularly appropriate for educational contexts, where students must trust AI systems across multiple functions.

D. Procedure

Participants were recruited through the UCF Psychology Department's research participation system and received course credit for participation. The survey was administered online via Qualtrics and required approximately 25–30 minutes to complete. After providing informed consent, participants completed measures in a fixed order: (1) demographic questions, (2) smartphone access timing, (3) superstitious beliefs (BSS), (4) technology anthropomorphism, (5) AI trust, (6) external locus of control, and (7) embedded attention checks. Fixed order was used to prevent exposure to AI trust items from priming technology-related superstitious thinking.

Three attention check items were embedded throughout the survey (e.g., "Please select 'Strongly Agree' for this item"). Participants who failed two or more attention checks were excluded from analyses (n = 25, 1.9% of initial sample). Upon completion, participants were debriefed about the study's purpose and provided with resources for learning about AI literacy and educational technology.

E. Analytic Approach

All analyses were conducted in SPSS. For cultural moderation analyses (Hypothesis 1), we computed separate Pearson correlation coefficients for URM and non-URM students, then tested for differences using Fisher's r-to-z transformation. For developmental timing analyses (Hypothesis 2), we compared early-access and late-access groups using independent-samples t-tests with Welch's correction for unequal variances. Effect sizes were computed using Cohen's d for group comparisons and interpreted using conventional benchmarks (small: d = 0.20, medium: d = 0.50, large: d = 0.80). Missing data were minimal (<2% for any single variable) and addressed through listwise deletion. Alpha was set at .05 for all tests, with exact p-values reported for transparency.

IV. RESULTS

A. The Paradox Is Not Universal: Cultural Moderation

Technology anthropomorphism predicted lower AI trust in the overall sample (r = −.14, p < .001). Critically, however, both groups showed nearly identical mean levels of anthropomorphism (M non-URM = 3.04 vs. M URM = 3.10, d = −0.04) and AI trust (M non-URM = 4.33 vs. M URM = 4.27, d = 0.06), yet the relationship between these constructs differed fundamentally (see Table I).

Among non-URM students (n = 1,072), anthropomorphism was a robust negative predictor of AI trust (r = −.155, p < .001, 95% CI [−.213, −.096]). To interpret this correlation: for every one standard-deviation increase in anthropomorphism among non-URM students, AI trust decreased by approximately 0.15 standard deviations. This replicates the expected "anthropomorphism paradox" — viewing technology as more human-like

predicts lower trust in AI. The narrow confidence interval that does not include zero confirms this is a reliable effect.

TABLE I. ANTHROPOMORPHISM–AI TRUST CORRELATIONS BY URM STATUS

Group	*n*	*r*	*p*	95% CI
Non-URM	1,072	−.155	< .001	[−.213, −.096]
URM	234	.021	.754	[−.108, .148]
Group difference (Fisher's $Z = -2.44$)			.015	—

Note. URM = UnderRepresented Minority. Non-URM = White and/or Asian students. CI = confidence interval.

Among URM students ($n = 234$), the same relationship was near-zero and non-significant ($r = .021$, $p = .754$, 95% CI [−.108, .148]). This correlation is essentially zero, and the wide confidence interval that crosses zero indicates there is no relationship between anthropomorphism and AI trust for URM students. Using Fisher's *r*-to-*z* transformation, which converts correlations to a common scale for comparison, we confirmed that the two groups differ significantly ($Z = -2.44$, $p = .015$).

Table I presents these core moderation findings. The confidence intervals provide additional evidence: the non-URM interval [−.213, −.096] excludes zero entirely, while the URM interval [−.108, .148] comfortably includes zero, confirming these are genuinely different patterns rather than merely weaker versions of the same effect.

This pattern, identical means but different relationships, reveals "moderation without mean differences." Both groups anthropomorphize technology to the same degree and trust AI at the same level on average. However, the psychological significance of anthropomorphism differs fundamentally: for non-URM students, viewing technology as human-like signals unpredictability and reduces trust; for URM students, these beliefs are psychologically unrelated to trust formation.

The correlation between bad luck beliefs and technology anthropomorphism operated similarly across groups (Non-URM: $r = .293$, $p < .001$; URM: $r = .228$, $p = .001$; Fisher's $Z = 0.96$, $p = .336$), indicating that the pathway from superstitious thinking to anthropomorphic technology beliefs does not differ by cultural background. The divergence emerges at the next step: what those anthropomorphic beliefs mean for AI trust.

B. Growing Up Algorithmic: Developmental Timing

Early smartphone access was nearly universal: 94.6% of students ($n = 1,235$) received their first smartphone before age 16 ($M = 12.30$ years, $SD = 2.30$); 48.6% gained access during elementary or middle school (ages 10–12). Despite small cell sizes, comparisons with late-access students (age ≥ 16; $n = 71$) revealed medium-sized effects across all three outcome variables (Table II).

Students who received smartphones before age 16 reported significantly stronger beliefs in bad luck ($M = 28.86$, $SD = 8.84$ vs. $M = 25.79$, $SD = 9.48$), $t(78.33) = 2.65$, $p = .010$, Cohen's $d = 0.35$. To contextualize this effect: the difference between early and late-access groups is approximately one-third of a standard deviation, translating to moving from the 50th percentile to approximately the 64th percentile in bad luck beliefs. This medium-sized effect suggests that early exposure to unpredictable algorithmic systems may shape enduring beliefs about uncontrollable forces in technology.

Students with early smartphone access also reported meaningfully lower AI trust ($M = 4.32$, $SD = 0.88$ vs. $M = 4.60$, $SD = 1.03$), $t(76.74) = -2.28$, $p = .026$, $d = -0.32$. On a 7-point scale, the 0.28-point difference represents approximately one-third of a standard deviation, a meaningful shift that could influence how students engage with educational AI throughout their university careers. Notably, this effect size ($d = -0.32$) is comparable to the bad luck effect ($d = 0.35$), suggesting that developmental timing has similarly strong impacts on both superstitious thinking and AI trust.

The difference in anthropomorphism did not reach conventional statistical significance ($M = 3.05$, $SD = 1.01$ vs. $M = 2.85$, $SD = 1.07$), $t(77.05) = 1.54$, $p = .129$, $d = 0.20$, though the effect size falls in the small-to-medium range and trends in the expected direction. The lack of statistical significance likely reflects the small late-access sample size ($n = 71$), which limits statistical power to detect small-to-medium effects.

Continuous analyses showed earlier access predicted stronger bad luck beliefs ($r = -.171$, $p < .001$) and greater anthropomorphism ($r = -.124$, $p < .001$), while the correlation with AI trust was near-zero ($r = .030$, $p = .277$), suggesting the trust effect operates as a developmental threshold rather than a linear gradient. The pattern where the categorical comparison (before vs. at/after age 16) yields significant effects on trust ($p = .026$) but the continuous correlation does not ($p = .277$) supports the interpretation that age 16 represents a critical developmental boundary rather than part of a smooth continuum.

TABLE II. KEY OUTCOMES BY AGE OF FIRST SMARTPHONE ACCESS

Outcome	Early (<16) M (SD)	Late (≥16) M (SD)	*t*	*p*	*d*
Bad Luck Beliefs	28.86 (8.84)	25.79 (9.48)	2.65	.010	0.35
AI Trust	4.32 (0.88)	4.60 (1.03)	−2.28	.026	−0.32
Anthropomorphism	3.05 (1.01)	2.85 (1.07)	1.54	.129	0.20

Note. Values are M (SD). Welch's correction applied for unequal variances. n early = 1,235; n late = 71.

V. DISCUSSION

A. The Paradox Has Boundaries

The anthropomorphism paradox is not a universal psychological mechanism. As demonstrated in Table I, it operates robustly among non-URM students ($r = -.155$, $p < .001$) but is entirely absent for URM students ($r = .021$, $p = .754$), despite both groups showing nearly identical mean levels of anthropomorphism and AI trust. This is exactly the kind of subgroup effect invisible to aggregate analyses, and it has a direct consequence for learning analytics: institutions monitoring average AI trust or average system engagement across student populations will see no equity problem. The problem is in the relationship structure underneath those means.

The statistical evidence in Table I is compelling: the Fisher's Z test confirms that these correlations differ significantly ($Z = -2.44$, $p = .015$), and the non-overlapping confidence intervals reinforce that this is a genuine moderation effect rather than measurement noise. We propose that URM students employ fundamentally different trust calibration strategies that are orthogonal to anthropomorphism. Trust for these students may depend more on perceived fairness, cultural representation, transparency, and historical reliability, all factors shaped by documented experiences with algorithmic bias in high-stakes domains [2][3]. When automated systems have disadvantaged one's community in practice, the question of whether a computer personally "hates" you may be psychologically irrelevant; trust is calibrated on different criteria entirely. For non-URM students, whose cultural contexts may emphasize individual predictability and control, an anthropomorphized AI becomes a potential social adversary, perceived as unpredictable and evaluative, triggering the distrust effect.

B. Early Smartphone Exposure Has Lasting Consequences

The 94.6% of students who received smartphones before age 16 enter university carrying stronger superstitious bad luck beliefs and lower AI trust than peers with later access, representing medium-sized effects ($d \approx 0.32–0.35$) that are not trivial. Table II reveals the specific pattern: early-access students score nearly one-third of a standard deviation higher on bad luck beliefs and one-third of a standard deviation lower on AI trust compared to late-access peers. Students who encountered recommendation algorithms, social media content curation, and AI-driven interfaces during developmentally sensitive periods appear to have formed mental models of technology that are less trusting and more infused with magical thinking.

For adaptive learning system designers, this matters: nearly the entire incoming student cohort is carrying a trust deficit shaped by their pre-university algorithmic experiences. The threshold pattern shown in Table II, where the categorical comparison (before vs. after age 16) yields significant effects on bad luck beliefs ($p = .010$) and AI trust ($p = .026$) while continuous analyses show weaker linear relationships, aligns with developmental research identifying mid-adolescence as a critical period for forming abstract reasoning about agency and causality, suggesting that early exposure shapes deep cognitive frameworks, not just surface attitudes.

C. Design Implications

Together, these findings (Tables I and II) expose the cost of one-size-fits-all educational AI design. Anthropomorphic features are typically justified by evidence from majority populations that they increase engagement. But if anthropomorphism reduces trust for non-URM students while being neutral for URM students (Table I), then the same design choice creates differential barriers to AI engagement, potentially widening achievement gaps even as educators intend the opposite. Similarly, the developmental timing effects in Table II suggest that current students arrive at university with baseline trust levels already shaped by childhood algorithmic exposure, meaning that educational AI systems must account for this pre-existing trust deficit rather than assuming a blank slate.

Three principles follow:

1. **Build adaptability in from the start.** Offer configurable interaction styles, either conversational and personality-rich for students who benefit, or neutral and function-focused for others, driven by student preference, not demographic assumptions.
2. **Disaggregate evaluation data.** Average engagement metrics mask real subgroup differences. As Table I demonstrates, URM and non-URM students can have identical means yet fundamentally different psychological relationships with AI features. Equity-centered evaluations should routinely examine whether AI systems create differential barriers across race/ethnicity, first-generation status, and technology exposure history.
3. **Scaffold AI literacy to address superstition-based anthropomorphism.** Since the superstition → anthropomorphism link is consistent across groups, helping students develop accurate mental models of how AI operates (e.g., pattern recognition, not consciousness) may reduce the anthropomorphic misattributions that drive distrust among non-URM students. Table II suggests this is particularly important given that 95% of students arrive with early algorithmic exposure that has already shaped their beliefs about technology as unpredictable and driven by forces beyond their control.

D. Limitations and Future Directions

Several limitations warrant consideration. The cross-sectional design precludes causal inference: we cannot establish that early smartphone access causes stronger superstitious beliefs and lower AI trust, nor that anthropomorphism causes reduced trust among non-URM students. Unmeasured variables such as parenting style, socioeconomic resources, personality traits, or prior experiences with AI system failures may confound observed relationships. The single-institution sample (a large Hispanic-Serving Institution in the southeastern United

States) limits generalizability to other institutional contexts, and the 94.6% prevalence of early smartphone access reflects current American college students but may not apply to older cohorts or international populations. Unequal group sizes, particularly for URM students (n = 234) and late smartphone access (n = 71), reduce statistical power for detecting small effects and create less stable estimates for these groups. Additionally, all measures relied on self-report, introducing potential social desirability bias and recall inaccuracy for smartphone access timing, and our AI trust measure assessed trust in "AI" generically rather than specific educational AI applications, which may show different patterns.

The URM classification aggregates Black/African American, Hispanic/Latino, Native American, Pacific Islander, and multiracial students into a single category, masking important heterogeneity in historical relationships with algorithmic systems, cultural values, and experiences with institutional discrimination. The observed null anthropomorphism-trust relationship among URM students may reflect averaging across subgroups with different patterns. Future research should examine specific racial/ethnic groups separately with larger samples, employ longitudinal designs to strengthen causal inference, incorporate behavioral measures of actual AI engagement rather than self-reported trust, test whether AI literacy interventions reduce superstition-based anthropomorphism, and experimentally manipulate anthropomorphic design features to validate recommendations for configurable systems. Intersectional analyses examining how race/ethnicity, gender, socioeconomic status, and first-generation status jointly shape AI trust formation would provide deeper understanding of psychological diversity in educational technology responses.

VI. CONCLUSION

The anthropomorphism paradox is real, but it is not universal, and its boundaries have direct implications for the design and evaluation of adaptive learning systems. Superstitious beliefs about bad luck feed anthropomorphic views of technology across all student groups, but the downstream effect on AI trust applies only to non-URM students; for UnderRepresented Minorities, anthropomorphism and AI trust are psychologically unrelated even when group means are identical. Separately, growing up with smartphones before age 16, which describes the experience of 95% of current students, leaves a lasting imprint: early algorithmic exposure predicts stronger superstitious technology beliefs and lower AI trust in young adulthood, suggesting that students arrive in higher education with meaningfully different baseline readiness for AI-powered learning.

The practical message for learning analytics and adaptive system design is direct: one size does not fit all, and aggregate metrics will not reveal the problem. Institutions deploying AI tutors, intelligent feedback systems, and adaptive curricula must disaggregate trust and engagement data by student population, design for configurability rather than fixed anthropomorphic features, and scaffold AI literacy that helps students develop accurate mental models of how algorithmic systems actually work. Educational equity in the age of AI is not just about access; it is about whether the psychological architecture of these systems works equitably for every learner in the room.

REFERENCES

[1] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on AI applications in higher education," International Journal of Educational Technology in Higher Education, vol. 16, Article 39, 2019.

[2] R. S. Baker and A. Hawn, "Algorithmic bias in education," International Journal of Artificial Intelligence in Education, vol. 32, no. 4, pp. 1052–1092, 2022.

[3] D. Gándara, H. Anahideh, M. P. Ison, and L. Picchiarini, "Inside the black box: Detecting and mitigating algorithmic bias across racialized groups in college student-success prediction," AERA Open, vol. 10, no. 1, pp. 1–15, 2024.

[4] A. Janowsky and M. Hubertz, "When students think AI 'hates them': How anthropomorphic attributions shape trust in educational technology," in Proc. HCI International 2026.

[5] J. L. Risen, "Believing what we do not believe: Acquiescence to superstitious beliefs," Psychological Review, vol. 123, no. 2, pp. 182–207, 2016.

[6] N. Epley, A. Waytz, and J. T. Cacioppo, "On seeing human: A three-factor theory of anthropomorphism," Psychological Review, vol. 114, no. 4, pp. 864–886, 2007.

[7] A. Polyportis and N. Pahos, "Understanding students' adoption of the ChatGPT chatbot in higher education: The role of anthropomorphism, trust, design novelty and institutional policy," Behaviour & Information Technology, vol. 44, no. 2, pp. 315–336, 2024.

[8] B. Jose and A. Thomas, "Digital anthropomorphism and the psychology of trust in generative AI tutors: An opinion-based thematic synthesis," Frontiers in Computer Science, vol. 7, Article 1638657, 2025.

[9] S. M. Fluke, R. J. Webster, and D. A. Saucier, "Methodological and theoretical improvements in the study of superstitious beliefs and behaviour," British Journal of Psychology, vol. 105, no. 1, pp. 102–126, 2014.

[10] J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," Proc. Machine Learning Research, vol. 81, pp. 77–91, 2018.

[11] A. Waytz, J. Cacioppo, and N. Epley, "Who sees human?" Perspectives on Psychological Science, vol. 5, no. 3, pp. 219–232, 2010.