

Dissertation Defense: Anticipated Questions and Preparation Guide

AI Readiness Scale (AIRS):
Extending UTAUT2 for Enterprise AI Tool Adoption

Fabio Correa

Doctor of Business Administration
Touro University Worldwide
December 2025

Contents

1	Opening and Overview Questions	3
1.1	Research Motivation	3
1.2	Research Questions	4
2	Theoretical Framework Questions	5
2.1	UTAUT2 Foundation	5
2.2	Theoretical Extension	7
2.3	Theoretical Positioning	9
3	Methodology Questions	11
3.1	Research Design	11
3.2	Sampling	12
3.3	Measurement	15
3.4	Data Collection	17
4	Statistical Analysis Questions	18
4.1	Factor Analysis	18
4.2	Model Fit	21
4.3	Structural Equation Modeling	23
4.4	Measurement Invariance	26
5	Results and Findings Questions	27
5.1	Hypothesis Testing	27
5.2	Moderation Effects	30
5.3	Exploratory Findings	32
6	Unexpected Findings Questions	34
6.1	Non-Significant Traditional Predictors	34
6.2	Dropped Constructs	36
7	Limitations and Validity Questions	38
7.1	Internal Validity	38
7.2	External Validity	40
7.3	Construct Validity	41
7.4	Statistical Conclusion Validity	43
8	Practical Implications Questions	45

8.1	Organizational Application	45
8.2	Stakeholder-Specific	47
9	Future Research Questions	49
9.1	Immediate Extensions	49
9.2	Theoretical Development	50
9.3	Practical Development	51
10	Philosophical and Broader Impact Questions	53
10.1	AI and Society	53
10.2	Broader Implications	55
11	Technical Deep-Dive Questions	57
11.1	Statistical Technicalities	57
11.2	Software and Reproducibility	60
12	Defense of Choices Questions	61
12.1	Methodological Choices	61
12.2	Theoretical Choices	63
12.3	Practical Choices	65
13	Preparation Tips	67
13.1	Before the Defense	67
13.2	During the Defense	67
13.3	Key Numbers to Memorize	67
14	Quick Reference: Potential “Gotcha” Questions	68

1 Opening and Overview Questions

1.1 Research Motivation

In one sentence, what is the central contribution of your dissertation?

Answer: “This dissertation develops and validates the AI Readiness Scale (AIRS), an 8-factor, 16-item diagnostic instrument extending UTAUT2 with AI Trust, revealing that Price Value ($\beta = .505$) rather than Performance Expectancy dominates AI adoption decisions—a significant departure from 30 years of technology acceptance research.”

Reference: §1.5.1 Theoretical Significance; §6.3 Theoretical Contributions

Why did you choose this topic? What personal or professional experiences led you here?

Answer: “Three observations motivated this research: (1) The adoption-value gap—McKinsey reports 88% AI adoption but only 6% achieve meaningful business value (§1.1); (2) MIT Media Lab estimates 90-95% of generative AI pilots fail to scale (§1.1); (3) As a DBA candidate, I recognized that organizations invest heavily in AI infrastructure but lack diagnostic tools to understand why individuals adopt or resist these tools. This represents both a significant business challenge and a research opportunity.”

Reference: §1.1 Background and Context; §1.2 Statement of the Problem

If you had to explain your research to a CEO in 2 minutes, what would you say?

Answer: “Your organization likely has an AI adoption problem—not a technology problem. Industry data shows 90-95% of AI pilots fail to deliver value. My research identified why: organizations lead with AI capabilities when employees actually evaluate AI through a value lens—‘Is this worth my time and effort?’ I developed a validated 16-item diagnostic tool that identifies specific adoption barriers—whether it’s trust deficits, inadequate perceived value, or low social influence—so you can design targeted interventions rather than one-size-fits-all training. The key insight: lead with value propositions, not capability demonstrations.”

Reference: §1.1 Background and Context; §5.5.1 For Organizations; §6.4.1 For Organizations Implementing AI

What gap in the literature does your study address?

Answer: “Three interconnected gaps: - **Theoretical Gap:** UTAUT2 was validated on mobile and consumer technologies, not AI. Venkatesh (2021) identified nine unique research challenges for AI adoption that existing frameworks don’t address (§1.2, §2.2.4). - **Measurement Gap:** No psychometrically validated instrument existed specifically for measuring AI adoption readiness in

professional contexts. Organizations relied on ad hoc surveys or general technology readiness measures (§1.2). - **Practice Gap:** Organizations lack diagnostic tools to identify which employees will adopt AI, which will resist, and why—contributing to the 90-95% pilot failure rate (§1.2, §2.3).”

Reference: §1.2 Statement of the Problem; §2.2.4 Research Gaps

1.2 Research Questions

Can you state your primary research question and explain why it matters?

Answer: “The primary research question is: *How can UTAUT2 be extended with AI-specific constructs to better predict behavioral intention to adopt AI tools in professional and academic contexts?*”

This matters because: (1) Existing technology acceptance models explain 50-74% of variance for traditional technologies but may not capture AI’s unique characteristics—opacity, probabilistic reasoning, autonomy, and ethical implications; (2) Without understanding AI-specific adoption drivers, organizations cannot design effective interventions; (3) The finding that Price Value dominates over Performance Expectancy has direct practical implications for how organizations communicate about AI tools.”

Reference: §1.4.1 Research Questions; §4.2.1 Primary Research Question

How do your secondary research questions build upon each other to answer the primary question?

Answer: “The five secondary research questions form a logical progression: - **RQ1** (Factor structure): Establishes the measurement model—what constructs exist and how they’re measured (answered: 8 factors, 16 items; §4.5.1) - **RQ2** (Measurement invariance): Confirms the instrument works across populations—students and professionals (answered: configural invariance achieved; §4.5.4) - **RQ3** (Strongest predictors): Identifies which factors drive adoption (answered: $PV > HM > SI$; §4.6.2) - **RQ4** (AI Trust extension): Tests the theoretical extension (answered: marginal effect, $p = .064$; §4.6.2) - **RQ5** (Moderators): Examines boundary conditions (answered: experience and population moderate HM; §4.6.3)

Together, these answer *how* UTAUT2 can be extended and *what* the extension reveals about AI adoption.”

Reference: §1.4.1 Research Questions; §4.2.2 Secondary Research Questions (Table 4.1)

Did your research questions evolve during the study? If so, how and why?

Answer: “The core questions remained stable, but the analysis evolved in two ways: - **Dropped constructs:** Four proposed AI-specific constructs (Voluntariness, Explainability, Ethical Risk, AI Anxiety) demonstrated inadequate reliability ($\alpha = .301-.582$) and were excluded. This shifted some hypotheses from confirmatory to ‘not testable’ rather than ‘not supported’ (§4.5.1, Table 4.4). - **Exploratory additions:** The user typology (4 segments) and disability-anxiety association emerged as exploratory findings not originally hypothesized (§4.6.8). These generate hypotheses for future research rather than answering the primary RQ directly.”

Reference: §4.5.1 Exploratory Factor Analysis (Construct Exclusion); §4.6.8 Exploratory Findings

2 Theoretical Framework Questions

2.1 UTAUT2 Foundation

Why did you choose UTAUT2 as your theoretical foundation rather than TAM, TAM3, or other acceptance models?

Answer: “UTAUT2 was selected for four reasons: - **Explanatory power:** UTAUT2 explains up to 74% variance in behavioral intention, compared to TAM’s ~40% (§2.2.2, Table 2.1). - **Theoretical integration:** UTAUT2 synthesizes eight prior models (TRA, TPB, TAM, MM, C-TAM-TPB, MPCU, IDT, SCT), providing comprehensive theoretical coverage rather than requiring separate theories (§2.2.2). - **Consumer/voluntary context:** Unlike original UTAUT (organizational/mandatory), UTAUT2 adds Hedonic Motivation, Price Value, and Habit—constructs highly relevant to AI adoption where use is often discretionary (§2.2.3). - **Established moderators:** UTAUT2’s moderators (age, gender, experience) provide ready hypotheses for boundary conditions (§2.2.3).”

Reference: §2.2.2 UTAUT Development; §2.2.3 UTAUT2 Extension; Table 2.1 (Model Comparison)

What are the seven core constructs of UTAUT2, and how does each relate to AI adoption specifically?

Answer: “The seven constructs and their AI-specific manifestations: - **Performance Expectancy (PE):** Will AI improve my work outcomes? (Non-significant in AIRS: $\beta = -.028$, $p = .791$) - **Effort Expectancy (EE):** Is AI easy to use? (Non-significant: $\beta = -.008$, $p = .875$; §4.6.2) - **Social Influence (SI):** Do important others expect me to use AI? (Significant: $\beta = .136$, $p = .024$; §4.6.2) - **Facilitating Conditions (FC):** Do I have resources and support for AI? (Non-significant: $\beta = .059$,

$p = .338$; §4.6.2) - **Hedonic Motivation (HM)**: Is using AI enjoyable? (Significant: $\beta = .217$, $p = .014$; §4.6.2) - **Price Value (PV)**: Is AI worth the cost/effort? (Dominant predictor: $\beta = .505$, $p < .001$; §4.6.2) - **Habit (HB)**: Not significant in AIRS: $\beta = .023$, $p = .631$ ”

Reference: §2.2.3 UTAUT2 Extension; §4.6.2 Structural Model Results (Table 4.10)

How does UTAUT2 differ from the original UTAUT, and why do those differences matter for your study?

Answer: “Three key differences: - **Context shift**: UTAUT targeted organizational/mandatory adoption; UTAUT2 targets consumer/voluntary adoption. AI adoption in 2024-2025 is predominantly voluntary, making UTAUT2 more appropriate (§2.2.3). - **Added constructs**: UTAUT2 adds Hedonic Motivation (enjoyment), Price Value (cost-benefit), and Habit. Price Value proved to be the dominant predictor ($\beta = .505$), validating this addition (§4.6.2). - **Removed moderator**: UTAUT2 removes Voluntariness as moderator since consumer contexts assume voluntary use. Interestingly, our proposed Voluntariness construct failed empirically ($\alpha = .301$), perhaps because participants in both academic and professional contexts experience varying degrees of organizational pressure (§4.5.1, Table 5.1).”

Reference: §2.2.3 UTAUT2 Extension; §4.5.1 EFA Results; §5.6.2 Dropped Constructs

Venkatesh identified nine research challenges for AI adoption. Which ones does your study address, and which remain unaddressed?

Answer: “Venkatesh et al. (2021) identified nine challenges. AIRS addresses four: - **Challenge 1 (Trust)**: Directly addressed through AI Trust construct (marginal: $\beta = .063$, $p = .064$; §4.6.2) - **Challenge 3 (Measurement)**: Addressed by developing validated 16-item instrument (§4.5) - **Challenge 5 (Individual differences)**: Addressed through moderator analysis (experience, population; §4.6.3) - **Challenge 7 (Value perceptions)**: Strongly addressed—Price Value dominance ($\beta = .505$) is a major finding (§4.6.2)

Remaining challenges for future research include: algorithmic transparency, human-AI collaboration dynamics, longitudinal adoption patterns, and ethical considerations. The failure of our Explainability and Ethical Risk constructs suggests these require different measurement approaches (§5.6.2, §6.6).”

Reference: §2.2.4 Research Gaps; §5.6.2 Dropped Constructs; §6.6 Recommendations for Future Research

2.2 Theoretical Extension

Why did you choose to extend UTAUT2 with AI Trust specifically? What other constructs did you consider?

Answer: “AI Trust was prioritized based on: - **Theoretical foundation:** Glikson & Woolley (2020) establish that AI trust differs fundamentally from technology trust due to AI’s opacity, autonomy, and anthropomorphism (§2.3.2). - **Unique AI characteristics:** Unlike traditional software, AI systems make probabilistic decisions that users cannot fully verify, creating trust-dependence (§2.3.2). - **Empirical support:** Meta-analyses show trust explains significant variance in AI adoption across contexts (§2.3.2).

Other constructs considered but dropped empirically: - **Voluntariness** ($\alpha = .301$): Items conflicted conceptually - **Explainability** ($\alpha = .503$): Measurement complexity - **Ethical Risk** ($\alpha = .572$): Items scattered across factors - **AI Anxiety** ($\alpha = .582$): Reverse-coded items problematic

These remain theoretically important but require different operationalization (§5.6.2, Table 5.1).”

Reference: §2.3.2 AI Trust; §4.5.1 EFA Results; §5.6.2 Dropped Constructs (Table 5.1)

You originally proposed four AI-specific constructs (Voluntariness, Explainability, Ethical Risk, AI Anxiety) but dropped them. Does this weaken your theoretical contribution?

Answer: “No—this finding *strengthens* the contribution in two ways:

First, empirical integrity: Rather than forcing poor-fitting constructs to confirm our hypotheses, we let the data speak. This demonstrates rigorous scale development methodology and avoids Type I errors (§4.5.1).

Second, this is itself a finding: The constructs’ failure reveals that: - These phenomena may be harder to measure than assumed - Current operationalizations conflate multiple sub-dimensions - Different measurement approaches are needed (§5.6.2, Table 5.1)

Third, the 8-factor model that emerged is parsimonious and robust: CFI = .975, TLI = .960, RMSEA = .065—excellent fit with 16 items (§4.5.2). The theoretical contribution is the validated model, not the proposed model.

The dissertation transparently documents this evolution, providing future researchers with specific guidance on what didn’t work and why (§5.6.2).”

Reference: §4.5.1 EFA Results; §4.5.2 CFA Results (Table 4.5); §5.6.2 Dropped Constructs; §6.6.3 Measurement Development

How do you define “AI Trust” conceptually? How does it differ from general technology trust or interpersonal trust?

Answer: “AI Trust in this study draws on Glikson & Woolley’s (2020) framework with three dimensions: - **Cognitive trust:** Belief in AI’s reliability and competence - **Emotional trust:** Comfort with AI’s autonomy and decisions - **Behavioral trust:** Willingness to rely on AI outputs

Differences from other trust types: - **vs. Technology trust:** Traditional technology trust assumes deterministic systems (input → predictable output). AI trust must accommodate probabilistic, sometimes inexplicable outputs (§2.3.2). - **vs. Interpersonal trust:** Interpersonal trust involves reciprocity and moral agency. AI lacks moral agency, yet users often anthropomorphize AI, creating hybrid trust expectations (§2.3.2). - **vs. Institutional trust:** AI trust is more granular—users may trust one AI application but not another, unlike blanket institutional trust.

Our operationalization focused on cognitive and behavioral dimensions given measurement constraints (§3.4.3).”

Reference: §2.3.2 AI Trust; §3.4.3 AI-Specific Extensions (Trust Items)

Your theoretical model assumes Trust directly predicts Behavioral Intention. Could there be indirect effects through other constructs?

Answer: “Yes, and this is a limitation worth acknowledging. Potential indirect pathways include:

- **Trust → Performance Expectancy → BI:** If I trust AI, I may perceive it as more useful
- **Trust → Effort Expectancy → BI:** Trust may reduce perceived effort (I don’t second-guess outputs)
- **Trust → Hedonic Motivation → BI:** Trust may increase enjoyment (less anxiety)

Why we tested direct effects only: (1) Parsimony—adding all indirect paths would require larger samples; (2) Theory—UTAUT2 framework assumes direct effects; (3) The marginal Trust effect ($p = .064$) suggests the relationship may indeed be mediated rather than direct (§4.6.2).

Future research direction: Test Trust as antecedent to UTAUT2 constructs rather than parallel predictor (§6.6.2).”

Reference: §4.6.2 Structural Model Results; §5.4.4 Non-Significance of Traditional Predictors; §6.6.2 Theoretical Extensions

How does your conceptualization of AI adoption differ from adoption of other emerging technologies like blockchain or IoT?

Answer: “AI adoption has unique characteristics:

- **Opacity:** AI’s ‘black box’ nature creates trust challenges absent in blockchain (transparent ledger) or IoT (sensor data) (§2.3.2)
- **Autonomy:** AI makes decisions; blockchain executes rules; IoT collects data. AI’s agency creates different adoption psychology
- **Anthropomorphism:** Users attribute human-like qualities to AI (especially conversational AI), affecting trust and expectations (§2.3.2)
- **Rapid capability evolution:** AI capabilities change dramatically year-over-year, unlike more stable technologies
- **Value ambiguity:** Our finding that Price Value dominates ($\beta = .505$) may reflect uncertainty about AI’s utility—harder to quantify than IoT efficiency gains

The AIRS instrument was designed specifically for these characteristics, though future research should test whether findings transfer to other emerging technologies (§6.6.1).’’

Reference: §2.3 AI-Specific Theoretical Considerations; §5.3.1 Price Value Dominance; §6.6.1 Generalizability

2.3 Theoretical Positioning

Where does your work fit within the broader technology acceptance research tradition?

Answer: “This study sits at the intersection of three research streams:

- **Technology acceptance evolution:** Part of the 40-year progression from TRA (1975) → TAM (1989) → UTAUT (2003) → UTAUT2 (2012) → AI-specific extensions (2020s). We extend this lineage to AI contexts (§2.2, Figure 2.1).
- **Scale development tradition:** Following Churchill (1979) and DeVellis (2016) scale development methodology with modern adaptations (split-sample validation, bootstrap CIs) (§3.5).
- **AI-human interaction research:** Contributing to emerging literature on how humans perceive, trust, and adopt AI systems (§2.3).

The dissertation bridges theory and practice—academically rigorous scale development with direct practitioner applications (§6.3, §6.4).’’

Reference: §2.2 Technology Acceptance Models; Figure 2.1 (Theoretical Evolution); §3.5 Data Analysis Approach

How do you respond to critics who argue that UTAUT is atheoretical—merely an empirical aggregation of constructs?

Answer: “This criticism has merit and deserves a nuanced response:

The criticism: UTAUT unified eight models empirically without developing new theoretical mechanisms—it tells us *what* predicts adoption, not *why*.

Partial defense: (1) Synthesis itself has value—UTAUT ended fragmented research using incompatible models; (2) The underlying theories (TRA, TPB, Social Cognitive Theory) do explain mechanisms; UTAUT integrates their predictions (§2.2.2).

Where I agree: UTAUT/UTAUT2 are better described as ‘empirical frameworks’ than ‘theories.’ They predict but don’t fully explain.

My contribution: Rather than defending UTAUT’s theoretical status, this study tests whether its *predictions* hold for AI. The finding that Performance Expectancy (core UTAUT predictor) becomes non-significant while Price Value dominates challenges the framework’s applicability—itsself a theoretical contribution (§5.4.1, §6.3).”

Reference: §2.2.2 UTAUT Development; §5.4.1 Price Value Dominance; §6.3 Theoretical Contributions

What theoretical assumptions underlie your model that might not hold in all contexts?

Answer: “Several assumptions warrant acknowledgment:

- **Intentionality assumption:** We measure Behavioral Intention, assuming intention predicts behavior. For AI, where use may be sporadic or context-dependent, intention-behavior links may be weaker (§5.7.1).
- **Rational actor assumption:** UTAUT assumes individuals weigh factors rationally. AI adoption may involve emotional or intuitive responses not fully captured (§5.7.1).
- **Construct independence:** We assume constructs are theoretically distinct, but Price Value and Performance Expectancy may be conflated when users equate ‘value’ with ‘usefulness’ (§5.7.1).
- **Cross-sectional stability:** We assume relationships are stable over time, but AI’s rapid evolution may shift adoption drivers (§5.7.2).
- **Western/English context:** Survey was English-only in U.S./UK—cultural factors may limit generalizability (§5.7.2).

These limitations are documented transparently in Chapter 5 (§5.7).”

Reference: §5.7.1 Internal Validity Limitations; §5.7.2 External Validity Limitations

3 Methodology Questions

3.1 Research Design

Why did you choose a quantitative survey methodology rather than qualitative or mixed methods?

Answer: “A quantitative survey approach was required for three reasons:

- **Scale development demands:** Creating a psychometrically validated instrument requires quantitative methods—factor analysis, reliability assessment, and validity testing cannot be accomplished qualitatively (§3.1).
- **Theoretical testing:** Extending UTAUT2 requires hypothesis testing with statistical inference, which necessitates quantitative data (§3.1).
- **Generalizability goal:** A primary objective was creating a generalizable diagnostic tool, requiring sample sizes and statistical validation not achievable through qualitative methods (§1.4.2).

That said, the study includes a qualitative component through open-ended feedback questions, which revealed important themes about AI adoption barriers and facilitators (§4.6.8). A mixed-methods follow-up study could enrich understanding of *why* Price Value dominates (§6.6.4).”

Reference: §3.1 Research Paradigm and Approach; §4.6.8 Exploratory Findings; §6.6.4 Methodological Extensions

Justify your cross-sectional design. What are its implications for causal inference?

Answer: “I acknowledge this limitation upfront:

Limitation: Cross-sectional design cannot establish causality. We observe correlations between predictors and intention at a single point, but cannot confirm temporal precedence or rule out reverse causality (§5.7.1).

Partial justification: - Intention → Behavior link is well-established (meta-analytic $\rho = .53-.69$ across UTAUT studies) - Scale validation is appropriately cross-sectional in initial development - UTAUT2 itself was validated cross-sectionally

Important caveat: Price Value ‘causing’ intention assumes users evaluate value, then form intentions. But intentions could shape value perceptions (‘I want to use AI, so I perceive it as valuable’). The $\beta = .505$ finding requires longitudinal replication (§5.7.1, §6.6.4).

Recommended follow-up: Longitudinal panel study tracking actual AI adoption behavior 3–6 months post-survey (§6.6.4).”

Reference: §3.2 Research Design; §5.7.1 Internal Validity Limitations; §6.6.4 Methodological Extensions

Why did you use a split-sample design (EFA/CFA) rather than full-sample CFA or cross-validation?

Answer: “Split-sample design (EFA: n=261; CFA: n=262) is the gold standard for scale development:

- **Prevents capitalization on chance:** If the same data used to explore factor structure is used to confirm it, fit statistics are artificially inflated (§3.5.1).
- **Independent validation:** The CFA sample had never ‘seen’ the EFA results—true cross-validation (§3.5.2).
- **Methodological precedent:** Recommended by Worthington & Whittaker (2006) and followed in UTAUT’s original validation (§3.5).

Alternatives considered: - *Full-sample CFA:* Only appropriate if strong a priori theory specifies exact structure—our AI extensions required exploration - *K-fold cross-validation:* Computationally intensive, more appropriate for predictive models than scale development

Random split verification: Samples were randomly split and verified for demographic comparability (§3.5.1, Table 3.1).”

Reference: §3.5 Data Analysis Approach; §3.5.1 Exploratory Factor Analysis; §3.5.2 Confirmatory Factor Analysis

3.2 Sampling

Describe your sampling strategy. Why did you use Centiment’s panel?

Answer: “Centiment was selected for five methodological advantages:

- **Topic-blinded recruitment:** Participants were invited based on demographics, not topic interest, reducing self-selection bias from AI enthusiasts (§3.3.1).
- **Panel quality:** Centiment maintains a vetted U.S. adult panel with attention checks and quality screening (§3.3.1).
- **Demographic targeting:** Enabled stratified sampling to achieve diverse age, gender, and occupation distribution (§3.3.2).
- **Completion enforcement:** Platform required completion of all items, eliminating missing data on key constructs (§3.3.1).

- **Rapid collection:** N=523 collected in approximately one week, ensuring temporal consistency in responses (§3.3.2).

Sample composition achieved: 54.3% female, mean age 42.5 years, 50.1% students/50.0% professionals, representing 15+ industries (§4.3, Table 4.1).”

Reference: §3.3 Population and Sampling; §3.3.1 Target Population; §3.3.2 Sampling Procedures; §4.3 Participant Characteristics (Table 4.1)

How do you know your sample is representative? What populations might be underrepresented?

Answer: “Representativeness was partially achieved but has limits:

What was achieved: - Gender balance (54.3% female) approximates workforce - Age range 18-75 with mean 42.5 captures working-age spectrum - Mix of students and professionals captures different AI exposure contexts - Multiple industries represented (§4.3, Table 4.1)

Likely underrepresented (acknowledged in §5.7.2): - Non-English speakers (survey English-only) - Non-U.S./UK populations (Western bias) - Blue-collar/manual labor workers (less online panel participation) - Technology-averse individuals (may avoid online surveys) - Older workers (60+) who may face different adoption dynamics

Implication: Findings most confidently generalize to English-speaking knowledge workers in Western contexts. Cross-cultural validation needed (§6.6.1).”

Reference: §4.3 Participant Characteristics; §5.7.2 External Validity Limitations; §6.6.1 Replication Studies

Your sample includes both students and professionals. How do you justify pooling these populations?

Answer: “Pooling was justified through measurement invariance testing:

- **Configural invariance achieved:** Same 8-factor structure holds across both populations—they conceptualize AI adoption similarly (§4.5.4, Table 4.7).
- **Metric invariance partially achieved:** Factor loadings largely equivalent, meaning constructs have similar meaning (§4.5.4).
- **Why both populations matter:** Students represent future workforce; professionals represent current workforce. Both use AI tools (ChatGPT, coding assistants, etc.) in their respective contexts (§3.3.1).

Important nuance: While structure is invariant, *paths* differ. Hedonic Motivation is significant

only for professionals, not students (§4.6.3, Table 4.12). This population moderation is itself a finding, not a threat.

If challenged: We test populations separately where appropriate and report differential effects transparently.”

Reference: §3.3.1 Target Population; §4.5.4 Measurement Invariance (Table 4.7); §4.6.3 Moderation Analysis (Table 4.12)

Is N=523 sufficient for your analyses? How did you determine sample size?

Answer: “Yes, N=523 exceeds multiple thresholds:

- **Factor analysis:** Hair et al. (2019) recommend 10:1 subject-to-item ratio. With 28 initial items, minimum = 280; we have 523 (18.7:1 ratio) (§3.3.2).
- **SEM:** Kline (2016) recommends $N > 200$ for SEM with acceptable fit; we have 262 in CFA sample (§3.3.2).
- **Per-group for invariance:** $N > 100$ per group recommended; we have ~130 students and ~130 professionals per subsample (§4.5.4).
- **Power for small effects:** Post-hoc power for Trust effect ($\beta = .063$): ~.52. This is under-powered for detecting small effects—a limitation acknowledged (§5.7.3).

The Trust effect caveat: With 80% power for small effects, we would need $N \approx 800$. The marginal Trust finding ($p = .064$) may reflect insufficient power rather than true null effect (§5.7.3).”

Reference: §3.3.2 Sampling Procedures; §4.5.4 Measurement Invariance; §5.7.3 Statistical Power Considerations

How did Centiment’s topic-blinded recruitment actually work? Are you confident it prevented self-selection bias?

Answer: “The process worked as follows:

- **Invitation:** Participants received generic invitations (‘Complete a survey for \$X compensation’) without mentioning AI, technology, or adoption (§3.3.1).
- **Screening:** Demographic screening occurred before topic reveal.
- **Topic reveal:** AI context revealed only in survey introduction after participants committed.

Confidence level: Reasonably confident, not certain. - *Strength:* This approach is superior to recruiting from AI forums or tech communities - *Limitation:* Online panel members are inherently more technology-comfortable than general population - *Evidence:* Our sample’s AI experience distribution (42% moderate-to-high experience) seems plausible, not skewed toward early adopters (§4.3, Table 4.2)

If I were to improve, I would compare Centiment sample to probability sample on key demographics (§5.7.2).”

Reference: §3.3.1 Target Population; §3.3.2 Sampling Procedures; §4.3 Participant Characteristics (Table 4.2)

3.3 Measurement

Walk me through the item development process. Where did your items come from?

Answer: “Items followed a systematic development process:

Source 1: UTAUT2 adaptation (§3.4.2): - Core constructs (PE, EE, SI, FC, HM, PV, BI) adapted from Venkatesh et al. (2012) - Modified referent from ‘mobile internet’ to ‘AI tools’ - Example: ‘Using mobile internet increases my productivity’ → ‘Using AI tools increases my productivity’

Source 2: AI Trust items (§3.4.3): - Developed from Glikson & Woolley (2020) trust framework - Focused on cognitive and behavioral trust dimensions - Example: ‘I trust AI tools to perform tasks accurately’

Source 3: AI-specific extensions (§3.4.3): - New items for Voluntariness, Explainability, Ethical Risk, AI Anxiety - These ultimately failed psychometrically but were developed from literature

Pilot testing: Items reviewed by 3 subject matter experts and pilot-tested with 25 participants for clarity (§3.4.4).”

Reference: §3.4 Instrumentation; §3.4.2 Core UTAUT2 Constructs; §3.4.3 AI-Specific Extensions; §3.4.4 Pilot Testing

Why did you use 2-item scales for each construct? What are the psychometric implications?

Answer: “Two-item scales resulted from EFA item reduction, not initial design:

What happened: Started with 3-4 items per construct; EFA reduced to best 2 based on loadings, cross-loadings, and theoretical coherence (§4.5.1).

Psychometric implications: - *Reliability:* Two items is minimum for calculating Cronbach’s alpha; all achieved $\alpha \geq .74$ (Table 4.3) - *AVE:* All constructs achieved AVE > .50, indicating adequate convergent validity (Table 4.6) - *Identification:* Each construct is just-identified in CFA, which is acceptable but limits model flexibility

Tradeoff acknowledged: - *Advantage:* Parsimony—16-item instrument is practical for organizational use - *Disadvantage:* Lower bandwidth; subtle within-construct differences not captured - *Mitigation:* High factor loadings (.75-.91) indicate items well-represent constructs (Table 4.5)

For future development, 3-item scales would improve identification and reliability (§6.6.3).”

Reference: §4.5.1 EFA Results (Table 4.3); §4.5.2 CFA Results (Tables 4.5, 4.6); §6.6.3 Measurement Development

How do you defend using a 5-point Likert scale rather than 7-point?

Answer: “Five-point scale selection was based on:

- **UTAUT2 consistency:** Original UTAUT2 used 5-point scales; maintaining consistency aids comparability (§3.4.1).
- **Adequate discrimination:** Research shows 5-7 point scales yield similar reliability; beyond 7 points, discrimination decreases (§3.4.1).
- **Cognitive burden:** Shorter scales reduce respondent fatigue, important for online surveys (§3.4.1).
- **Statistical treatment:** 5+ points generally allows parametric treatment without serious distortion.

If challenged on sensitivity: - Our reliability coefficients ($\alpha = .74-.89$) are strong, suggesting adequate item discrimination - Variance in responses was sufficient for factor analysis (no ceiling/floor effects observed) - 7-point might marginally improve discrimination but at cost of complexity

This is a defensible choice, not the only valid choice (§3.4.1).”

Reference: §3.4.1 Survey Design

What steps did you take to minimize common method variance?

Answer: “Both procedural and statistical remedies were employed:

Procedural remedies (§3.4.4): - Anonymity assurance to reduce social desirability - Predictor and criterion variables separated in survey - Different scale endpoints for different constructs where possible - Randomized item order within blocks

Statistical remedies (§4.4): - Harman’s single-factor test: First factor explained 38.2% of variance (< 50% threshold) - Correlation matrix examination: No correlations > .90 suggesting common factor - Discriminant validity: HTMT ratios < .85 for all construct pairs (Table 4.6)

Honest assessment: CMV cannot be completely ruled out in single-source, single-method studies. The procedural and statistical remedies reduce but don’t eliminate this threat (§5.7.1).

Best mitigation would be: Multi-source data (e.g., supervisor ratings of AI use) or temporal separation—recommended for future research (§6.6.4).”

Reference: §3.4.4 Pilot Testing; §4.4 Data Screening; §5.7.1 Internal Validity Limitations

How did you operationalize “Behavioral Intention” and how does it relate to actual behavior?

Answer: “Behavioral Intention was operationalized with two items: - ‘I intend to use AI tools in the next 3 months’ - ‘I plan to use AI tools regularly in my work/studies’

Relationship to actual behavior: - Meta-analyses show intention-behavior correlations of $\rho = .53$ (Sheeran, 2002) to $\rho = .69$ (Venkatesh et al., 2016) - UTAUT studies specifically find strong intention → use relationships - Three-month timeframe increases specificity per Ajzen’s (1991) recommendations

Limitation acknowledged (§5.7.1): - We measure intention, not behavior - Intention-behavior gap may be larger for AI due to: - Rapid capability changes (intentions formed about current AI may not transfer) - Organizational constraints (intention exists but access doesn’t) - Novelty (less habituated behavior patterns)

Recommended validation: Longitudinal follow-up measuring actual AI tool usage logs (§6.6.4).”

Reference: §3.4.2 Core UTAUT2 Constructs (BI Items); §5.7.1 Internal Validity Limitations; §6.6.4 Methodological Extensions

3.4 Data Collection

What quality controls did you implement to ensure data quality?

Answer: “Multi-layered quality controls:

Platform-level (Centiment; §3.3.1): - Verified panel membership - Historical response quality tracking - Device fingerprinting to prevent duplicates

Survey-level (§3.4.4): - Two attention check items (‘Please select Agree for this item’) - Minimum completion time threshold (< 3 minutes flagged) - Consistency checks on similar items - Forced response on all items (no skipping)

Analysis-level (§4.4): - Univariate outlier detection ($z > |3.29|$) - Multivariate outlier detection (Mahalanobis distance) - Response pattern analysis (straight-lining detection)

Results: 47 responses removed for quality issues; N=523 represents cleaned sample (§4.4).”

Reference: §3.3.1 Target Population; §3.4.4 Pilot Testing; §4.4 Data Screening

What was your response rate, and how might non-response bias affect your findings?

Answer: “Response rate context for online panels differs from traditional surveys:

Panel mechanics: Centiment invited panel members until N=570 completions achieved; ‘response rate’ is less meaningful because: - Invitations are ongoing until quota filled - Panel members self-select into panel initially - Standard response rate calculations don’t apply

Quality-adjusted sample: 570 completions → 523 after quality screening (91.8% retention) (§4.4)

Non-response bias considerations: - Those who ignored invitation may differ from responders - Likely more technology-comfortable sample than general population - Topic-blind recruitment partially mitigates AI-specific self-selection

Honest limitation: We cannot fully characterize non-responders. Findings should be interpreted as applying to individuals willing to complete online surveys about technology—a meaningful but bounded population (§5.7.2).”

Reference: §3.3.2 Sampling Procedures; §4.4 Data Screening; §5.7.2 External Validity Limitations

How did you handle missing data?

Answer: “Missing data was minimal by design:

Prevention strategy (§3.3.1): - Centiment platform required response to all items - No ‘prefer not to answer’ options for core constructs - Survey could not be submitted incomplete

Result: Zero missing values on 16 final AIRS items and Behavioral Intention (§4.4)

Demographics: Minor missingness (< 2%) on optional demographics handled via listwise deletion for analyses using those variables (§4.4)

If asked about imputation: Given complete data, imputation was unnecessary. Had missing data occurred, multiple imputation would have been preferred over listwise deletion for SEM analyses.”

Reference: §3.3.1 Target Population; §4.4 Data Screening

4 Statistical Analysis Questions

4.1 Factor Analysis

Walk me through your EFA process. How did you determine the number of factors to retain?

Answer: “The EFA process followed established psychometric guidelines:

Step 1: Preliminary checks (§4.5.1): - KMO = .923 (excellent sampling adequacy; > .90 is ‘marvelous’) - Bartlett’s test: $\chi^2 = 3,847$, $p < .001$ (correlations exist)

Step 2: Factor extraction: - Used principal axis factoring (appropriate for scale development) - Initial eigenvalues suggested 8-10 factors

Step 3: Number of factors determined by (§4.5.1): - **Parallel analysis:** Compared actual eigenvalues to random data—8 factors exceeded chance - **Scree plot:** Clear elbow at 8 factors (Figure 4.1) - **Theoretical alignment:** 8 factors corresponded to 7 UTAUT2 constructs + AI Trust - **Interpretability:** Each factor was theoretically coherent and nameable

Result: 8-factor solution with 16 items, explaining 71.2% of variance (Table 4.3).”

Reference: §4.5.1 Exploratory Factor Analysis; Table 4.3 (EFA Pattern Matrix); Figure 4.1 (Scree Plot)

Why did you use oblimin rotation rather than varimax?

Answer: “Oblimin (oblique) rotation was chosen because:

- **Theoretical expectation:** UTAUT2 constructs are theoretically correlated—Performance Expectancy should correlate with Effort Expectancy; Social Influence should correlate with Facilitating Conditions. Forcing orthogonality (varimax) would distort these relationships (§3.5.1).
- **Empirical confirmation:** Factor correlations ranged from .21 to .67, confirming constructs are not independent (Table 4.3).
- **Methodological consensus:** Oblique rotation is standard in psychology/IS research when constructs are expected to correlate (Costello & Osborne, 2005).

If forced orthogonality: Simple structure would be harder to achieve; cross-loadings would increase artificially; interpretation would be compromised.

Note: I report the pattern matrix (unique contribution of each factor) rather than structure matrix (§4.5.1).”

Reference: §3.5.1 Exploratory Factor Analysis; §4.5.1 EFA Results

What criteria did you use for item retention in EFA? Any items that were borderline?

Answer: “Item retention criteria (§3.5.1):

- **Primary loading:** $\geq .50$ on intended factor (stricter than .40 minimum)
- **Cross-loadings:** $< .32$ on non-intended factors (Tabachnick & Fidell guideline)
- **Communalities:** $\geq .40$ (item shares sufficient variance with factors)

- **Theoretical coherence:** Item content matches factor interpretation

Items removed: - 12 items eliminated for cross-loadings or low communalities - Entire constructs dropped: Voluntariness ($\alpha = .301$), Explainability ($\alpha = .503$), Ethical Risk ($\alpha = .572$), AI Anxiety ($\alpha = .582$)—items scattered across multiple factors or formed unreliable clusters (§4.5.1, Table 5.1)

Borderline decisions: - FC2 (Facilitating Conditions item 2) loaded at .52—retained for theoretical importance and adequate reliability - One Trust item loaded at .48 initially but improved to .61 after removing poorly performing items

Final 16 items all exceeded .60 primary loadings with no cross-loadings $> .25$ (Table 4.3)."

Reference: §3.5.1 EFA Methodology; §4.5.1 EFA Results (Table 4.3); §5.6.2 Dropped Constructs (Table 5.1)

Explain your CFA model specification. How did you handle correlated errors?

Answer: "CFA model specification (§4.5.2):

Structure: - 8 latent factors (PE, EE, SI, FC, HM, PV, TR, BI) - 16 observed indicators (2 per factor) - Factors allowed to freely correlate (consistent with oblique EFA) - Factor variances fixed to 1.0 for identification

Correlated errors: - **Initial model:** No correlated errors specified (theory-driven approach) - **Model achieved excellent fit without modifications:** CFI = .975, TLI = .960, RMSEA = .065, SRMR = .046 (Table 4.5) - **Therefore:** No post-hoc error correlations added

Rationale for not adding correlated errors: - Data-driven modifications inflate fit artificially - With 2 items per construct, error correlations would be difficult to justify theoretically - Excellent fit without modifications indicates clean measurement model

If fit had been inadequate, I would have examined modification indices but required theoretical justification (e.g., method effects, similar wording) before adding error correlations."

Reference: §3.5.2 Confirmatory Factor Analysis; §4.5.2 CFA Results (Table 4.5)

What estimation method did you use for CFA and why?

Answer: "MLM (Maximum Likelihood with robust standard errors—Satorra-Bentler correction) was selected because:

Multivariate non-normality detected (§4.4): - Mardia's multivariate kurtosis = 47.3 ($p < .001$) - Several items showed univariate skewness (negative skew toward agreement)

Why MLM specifically: - Provides corrected chi-square (Satorra-Bentler scaled χ^2) robust to non-normality - Adjusts standard errors for kurtosis - Maintains interpretability of fit indices - Preferred over ADF (asymptotic distribution-free) which requires $N > 1,000$

Software: semopy (Python) with robust estimation (§3.5.2)

Alternative considered: Bootstrapped estimation—but MLM is more efficient and widely accepted; bootstrap used specifically for confidence intervals on path coefficients (Table 4.10)."

Reference: §3.5.2 Confirmatory Factor Analysis; §4.4 Data Screening; §4.5.2 CFA Results

4.2 Model Fit

Walk me through your model fit indices. Which do you consider most important and why?

Answer: "Model fit assessed using multiple indices (§4.5.2, Table 4.5):

Index	Value	Threshold	Interpretation
CFI	.975	> .95	Excellent
TLI	.960	> .95	Excellent
RMSEA	.065	< .08	Acceptable
SRMR	.046	< .08	Excellent
χ^2/df	1.89	< 3.0	Good

Most important indices (in my view): 1. **CFI:** Compares model to null model; robust to sample size; .975 indicates model recovers 97.5% of covariance 2. **SRMR:** Absolute fit; sensitive to misspecified factor covariances; .046 is excellent 3. **RMSEA:** Parsimony-adjusted; penalizes complexity; .065 acceptable given excellent other indices

Why not rely solely on χ^2 : Chi-square is sensitive to sample size—with $N = 262$, even trivial misfit is significant. χ^2/df ratio (1.89) provides better assessment.

Overall: Converging evidence across incremental (CFI, TLI) and absolute (SRMR, RMSEA) indices supports excellent model fit."

Reference: §4.5.2 CFA Results (Table 4.5); §3.5.2 Fit Index Interpretation

Your RMSEA is .065. Some argue this is mediocre. How do you respond?

Answer: "I acknowledge the criticism and respond on several grounds:

Context for .065: - .065 falls within the .05–.08 ‘acceptable’ range (Browne & Cudeck, 1993) - 90% CI: [.048, .082]—lower bound in ‘good’ range, upper bound still acceptable - Close fit test (RMSEA < .05): $p = .067$ —marginally non-significant but close

Compensating strengths: - CFI = .975 and TLI = .960 are excellent ($> .95$) - SRMR = .046 is well below .08 threshold - RMSEA is known to penalize parsimony—with only 16 items for 8 factors, some penalty is expected

Why RMSEA may be slightly elevated: - 2-item factors create identification constraints - RMSEA can be inflated in models with few degrees of freedom per factor - Kenny, Kaniskan, & McCoach (2015) note RMSEA behaves differently in simple models

Bottom line: No single fit index is definitive. The pattern of indices—excellent CFI/TLI/SRMR with acceptable RMSEA—supports adequate fit. I would not revise the model based on RMSEA = .065 alone.”

Reference: §4.5.2 CFA Results; §5.7.3 Statistical Considerations

How did you assess and establish discriminant validity?

Answer: “Discriminant validity was assessed using multiple criteria (§4.5.3, Table 4.6):

Criterion 1: Fornell-Larcker: - $\sqrt{\text{AVE}}$ for each construct must exceed its correlations with other constructs - All 8 constructs passed: $\sqrt{\text{AVE}}$ ranged .77–.89; highest inter-construct correlation = .67

Criterion 2: HTMT (Heterotrait-Monotrait): - More stringent modern criterion - All HTMT ratios < .85 (strictest threshold) - Highest HTMT = .73 (PE-PV)—these constructs are theoretically related but empirically distinct

Criterion 3: Inter-construct correlations: - All correlations < .85 (no multicollinearity concern) - 95% CIs for correlations did not include 1.0

Interpretation: Constructs measure distinct phenomena despite theoretical relationships. Users differentiate between, for example, AI usefulness (PE) and AI value (PV)—a non-trivial finding given their conceptual similarity.”

Reference: §4.5.3 Construct Validity (Table 4.6)

How did you assess convergent validity?

Answer: “Convergent validity was established through (§4.5.3):

Criterion 1: Factor loadings: - All standardized loadings $> .70$ (range: .75–.91) - This exceeds

the .70 threshold indicating items strongly represent their constructs - All loadings statistically significant ($p < .001$)

Criterion 2: Average Variance Extracted (AVE): - $AVE > .50$ required (construct explains more variance in items than error) - All 8 constructs exceeded: AVE range = .60–.79 (Table 4.6)

Criterion 3: Construct reliability: - Composite reliability (CR) $> .70$ for all constructs - Cronbach's α range: .74–.89 (Table 4.3)

Interpretation: Items within each construct converge—they measure the same underlying phenomenon. The high loadings (.75–.91) are particularly strong evidence that our 2-item scales adequately capture each construct.”

Reference: §4.5.3 Construct Validity (Tables 4.5, 4.6)

4.3 Structural Equation Modeling

Explain your structural model specification. How did you model the relationships?

Answer: “Structural model specification (§4.6):

Endogenous variable: Behavioral Intention (BI)

Exogenous predictors (7 direct paths to BI): - $PE \rightarrow BI$, $EE \rightarrow BI$, $SI \rightarrow BI$, $FC \rightarrow BI$, $HM \rightarrow BI$, $PV \rightarrow BI$, $TR \rightarrow BI$

Control variables: Age, Gender, Experience (entered as observed covariates)

Moderation testing (§4.6.3): - Experience \times HM interaction (continuous moderator) - Multi-group models for population (students vs. professionals)

Estimation: Full structural model with measurement model embedded (not two-step approach)

What was NOT modeled: - Mediation paths between UTAUT2 constructs (parsimony) - Second-order factors (not theoretically justified) - Formative indicators (all reflective measurement)

The model tests whether UTAUT2 predictors + AI Trust jointly predict AI adoption intentions, with R^2 indicating total variance explained.”

Reference: §4.6 Structural Model Analysis; §4.6.1 Model Specification; Figure 4.2 (Structural Model)

Why did you use SEM rather than multiple regression?

Answer: “SEM was chosen over multiple regression for five reasons:

1. **Latent variable modeling:** SEM separates true construct variance from measurement error. Multiple regression treats scale means as error-free, attenuating relationships (§3.5.3).
2. **Simultaneous estimation:** Measurement and structural models estimated together, ensuring correct standard errors and avoiding two-step bias.
3. **Model fit assessment:** SEM provides global fit indices (CFI, RMSEA) unavailable in regression—we can assess whether the theoretical model fits the data pattern, not just whether paths are significant.
4. **Measurement invariance:** SEM enables formal testing of whether the instrument works equivalently across groups—critical for the student/professional comparison (§4.5.4).
5. **Complex modeling:** Moderation and mediation can be tested within an integrated framework rather than separate analyses.

What regression would miss: With 2-item scales, measurement error is substantial. Regression would underestimate relationships due to attenuation. SEM’s latent modeling corrects this.”

Reference: §3.5.3 Structural Equation Modeling; §4.6 Structural Model Analysis

How did you handle the R^2 of .852? Isn’t that suspiciously high?

Answer: “ $R^2 = .852$ (85.2% variance explained) is high but defensible:

Context from UTAUT literature: - Original UTAUT: $R^2 = .69$ (Venkatesh et al., 2003) - UTAUT2: $R^2 = .74$ (Venkatesh et al., 2012) - Blut et al. (2022) meta-analysis: Average $R^2 = .45-.67$ depending on context - Our R^2 is higher but within the upper range for well-specified models

Why our R^2 may be legitimately high: 1. **Latent variable modeling:** SEM corrects for measurement error, increasing explained variance vs. regression 2. **Clean sample:** Quality-screened panel data with complete responses 3. **Strong theoretical model:** UTAUT2 is well-established; our extension added Trust 4. **Intention as DV:** Behavioral Intention is more proximal to predictors than actual behavior

Not evidence of problems: - No multicollinearity ($VIF < 3.0$, all correlations $< .85$) - Discriminant validity established ($HTMT < .85$) - Price Value dominance ($\beta = .505$) alone explains substantial variance

Honest caveat: Common method variance could inflate R^2 —acknowledged in §5.7.1.”

Reference: §4.6.2 Structural Model Results (Table 4.10); §5.7.1 Internal Validity Limitations

How did you test for mediation effects?

Answer: “Mediation was tested using bootstrap confidence intervals (§4.6.4):

Hypothesized mediation: $EE \rightarrow TR \rightarrow BI$ (Effort Expectancy influences intention through Trust)

Method: - 5,000 bootstrap resamples - Bias-corrected 95% confidence intervals for indirect effect
- Sobel test as supplementary

Result: - Indirect effect = .008 - 95% CI: [-.003, .019]—includes zero - Mediation NOT supported (H12; Table 4.14)

Interpretation: Trust does not mediate the $EE \rightarrow BI$ relationship. This suggests: - Ease of use influences intentions directly, not through trust-building - Trust operates independently of perceived ease - Future research could test other mediation pathways (e.g., PV mediating $PE \rightarrow BI$)

Methodological note: We tested only the theoretically specified mediation, not all possible indirect paths (data-driven mediation hunting would capitalize on chance).”

Reference: §4.6.4 Mediation Analysis (Table 4.14); §5.4.4 Non-Significance of Traditional Predictors

Explain your moderation analysis approach.

Answer: “Two types of moderation were tested (§4.6.3):

Type 1: Continuous moderation (Experience) - Created interaction terms: Experience \times each predictor - Entered interactions in structural model - Significant interaction: Experience \times HM ($p = .009$) - Interpretation: HM’s effect on BI increases with experience (Table 4.12)

Type 2: Categorical moderation (Population) - Multi-group SEM: Separate models for students vs. professionals - Paths freely estimated within groups - Chi-square difference test for path equivalence - Key finding: HM $\beta = .449$ (students) vs. $\beta = -.301$ (professionals)—significant difference (Table 4.12)

Why multi-group for population, interaction for experience: - Population is categorical (binary)—multi-group is natural and allows testing measurement invariance simultaneously - Experience is continuous—interaction terms more efficient and preserve information

Other moderators tested: Age \times predictors, Gender \times predictors—no significant interactions (§4.6.3).”

Reference: §4.6.3 Moderation Analysis (Tables 4.11, 4.12)

4.4 Measurement Invariance

Walk me through your measurement invariance testing. What levels did you test?

Answer: “Measurement invariance testing followed the progressive hierarchy (§4.5.4):

Level 1: Configural invariance ☐ ACHIEVED - Same factor structure (8 factors, same item-factor assignments) in both groups - Model fit: CFI = .961, RMSEA = .071 - Interpretation: Students and professionals conceptualize AI adoption similarly

Level 2: Metric invariance ~ PARTIAL - Constrains factor loadings equal across groups - $\Delta\text{CFI} = .008$ (exceeds .010 threshold marginally) - Mean $\Delta\lambda = .082$ across loadings - Most loadings equivalent; Social Influence showed largest difference ($\Delta\lambda = .326$) - Interpretation: Constructs have largely similar meaning across groups, with some variation

Level 3: Scalar invariance — NOT TESTED - Would constrain intercepts equal - Requires full metric invariance as prerequisite - Given partial metric invariance, scalar testing was not pursued

Practical implication: Configural invariance is sufficient for comparing structural paths (which predictors matter). Full scalar invariance would be needed for mean comparisons (§4.5.4, Table 4.7).”

Reference: §4.5.4 Measurement Invariance (Table 4.7)

You didn’t achieve full metric invariance. What are the implications?

Answer: “Partial metric invariance has specific implications:

What we CAN do: - Compare structural path coefficients across groups (which paths are significant) - Interpret configural invariance as confirming same construct structure - Report group-specific findings with appropriate caveats

What we should be CAUTIOUS about: - Direct comparison of factor means across groups - Assuming perfectly equivalent measurement - Pooling data without acknowledging measurement differences

How we handled it (§4.5.4): - Reported both pooled and group-specific structural models - Noted that Social Influence operates differently across populations - Interpreted population moderation findings in light of measurement differences

Theoretical insight: The loading difference for Social Influence ($\Delta\lambda = .326$) is itself interesting—peer influence may be more salient/consistent in professional contexts (supervisor expectations) than academic contexts (more diffuse social norms).

This is not a failure but a finding about AI adoption's context-dependence.”

Reference: §4.5.4 Measurement Invariance; §5.4.6 Population-Specific Pathways

The Social Influence factor showed the largest loading difference ($\Delta\lambda = .326$). Why might this be?

Answer: “Several theoretical explanations for SI loading differences:

For professionals (higher/more consistent loadings): - Clear organizational hierarchy: ‘supervisor expects’ is concrete - Colleague behavior is observable and relevant - Professional consequences for non-adoption (performance reviews) - SI items tap well-defined referents

For students (lower/more variable loadings): - Peer influence more diffuse: ‘who thinks I should use AI?’ - Professor expectations may vary by course/discipline - Less formal social structure around AI use - SI items may not capture student-specific referents

Theoretical alignment: Social influence effectiveness depends on referent specificity and power. Professional contexts provide clearer referents with accountability; academic contexts have more ambiguous social expectations for AI tool use.

Practical implication: Organizations implementing AI can leverage explicit social expectations (training requirements, team adoption goals); educational contexts may need different influence mechanisms.

Future research: Develop student-specific SI items (e.g., ‘My professor expects me to use AI for assignments’) (§6.6.3).”

Reference: §4.5.4 Measurement Invariance (Table 4.7); §5.4.6 Population-Specific Pathways

5 Results and Findings Questions

5.1 Hypothesis Testing

Summarize your hypothesis testing results. What percentage of hypotheses were supported?

Answer: “Of 12 testable hypotheses (Tables 4.10-4.14):

Outcome	Count	Percentage	Hypotheses
Supported	5	42%	H6 (PV→BI), H5 (HM→BI), H3 (SI→BI), H2 (EE→BI), H10 (Exp×HM)
Partial	2	17%	H8 (TR→BI, p=.064), H4 (FC→BI, p=.065)
Not Supported	5	42%	H1 (PE→BI), H7 (H→BI), H9 (Age mod), H11 (Gender mod), H12 (Mediation)

Four additional hypotheses were not testable due to dropped constructs (Voluntariness, Explainability, Ethical Risk, AI Anxiety)—these represent measurement challenges, not theoretical refutations.

Key insight: The 42% support rate is not a ‘failure’—it reveals that AI adoption differs from traditional technology adoption. The *pattern* of what works and doesn’t work is the theoretical contribution.”

Reference: §4.2.2 Secondary Research Questions (Table 4.1); §4.6.2 Structural Model Results (Table 4.10); §4.6.5 Hypothesis Summary (Table 4.14)

Three of seven UTAUT2 hypotheses were supported. Is this a failure of the theory or your study?

Answer: “Neither—this is the study’s most important finding:

Not a theory failure: - UTAUT2 was validated on mobile technology in consumer contexts - The theory’s constructs remain valid; their *relative importance* shifts for AI - Venkatesh himself (2021) predicted AI would require model adaptation

Not a study failure: - Adequate power, validated measures, quality data - Significant effects where theory predicts (PE, EE) are legitimately non-significant for AI - This is *discovery*, not *disconfirmation*

The finding: AI adoption operates through a **value lens** rather than **utility lens**. Users ask ‘Is AI worth it?’ (PV: $\beta = .505$) rather than ‘Will AI help me perform better?’ (PE: $\beta = .028$, ns).

Theoretical implication: UTAUT2 may need domain-specific adaptations—just as UTAUT2 adapted UTAUT for consumers. An ‘UTAUT-AI’ variant might weight constructs differently (§6.3).”

Reference: §5.3.1 Price Value Dominance; §5.4.1 Theoretical Implications; §6.3 Theoretical Contributions

Walk me through the significant predictors: Price Value, Hedonic Motivation, Social Influence. Why these three?

Answer: “The three significant predictors reveal AI’s adoption psychology:

Price Value ($\beta = .505$, $p < .001$) (§5.3.1): - Dominant predictor—explains most unique variance - AI framed as investment decision with uncertain ROI - Freemium models make cost-benefit salient - Users evaluate ‘Is this worth my time/money/learning effort?’

Hedonic Motivation ($\beta = .217$, $p = .014$) (§5.3.2): - Enjoyment matters—AI has novelty appeal - ‘Fun to use’ differentiates adoption intentions - Stronger effect with experience (interaction: $\beta = .136$, $p = .009$) - Moderated by population—stronger for students

Social Influence ($\beta = .197$, $p < .001$) (§5.3.3): - Peer effects remain powerful - ‘Important others think I should use AI’ - Professional contexts: supervisor expectations - Academic contexts: peer adoption norms

Pattern interpretation: AI adoption is driven by value perceptions, social normalization, and intrinsic enjoyment—not by performance gains or ease of use. This suggests different intervention strategies than traditional technology rollouts.”

Reference: §4.6.2 Structural Model Results (Table 4.10); §5.3.1-5.3.3 Interpretation of Significant Predictors

Why is Price Value ($\beta = .505$) so dominant? What’s the theoretical explanation?

Answer: “Price Value dominance was unexpected but has multiple explanations:

Explanation 1: Value uncertainty (§5.3.1): - AI benefits are hard to quantify—unlike ‘saves 2 hours/week’ - Users can’t confidently assess performance gains, so they focus on evaluable costs - Value acts as proxy for uncertain utility

Explanation 2: Freemium market structure: - ChatGPT, Copilot, etc. have free tiers - ‘Worth paying for?’ becomes the adoption decision point - Creates explicit value calculation moment

Explanation 3: Investment framing: - AI requires learning investment, workflow changes - Users evaluate total cost (money + time + disruption) - PV captures this comprehensive assessment

Explanation 4: Capability commoditization (§5.4.1): - Basic AI capability assumed (‘of course it can write emails’) - Performance is baseline expectation, not differentiator - Competition shifts to value proposition

Practical implication: Organizations should communicate AI's value proposition clearly—ROI, time savings, career advantages—rather than showcasing capabilities (§6.4.1)."

Reference: §5.3.1 Price Value Dominance; §5.4.1 Theoretical Implications; §6.4.1 For Organizations Implementing AI

AI Trust was marginally significant ($p = .064$). How do you interpret this?

Answer: "The marginal Trust effect requires nuanced interpretation:

Statistical reality: - $\beta = .063$, $p = .064$ (just above .05 threshold) - 95% CI: $[-.004, .130]$ —includes zero but barely - Effect size small but non-trivial

Three interpretations (§5.4.3):

1. **Power limitation:** With $N = 262$ (CFA sample), power for small effects $\approx .52$. Need $N \approx 600$ for 80% power. The effect may be real but undetected (Type II error risk).
2. **Context specificity:** Trust may matter more for high-stakes AI (medical, legal, financial) than general tools in our sample. Survey didn't specify AI type.
3. **Mediation possibility:** Trust might influence intention indirectly through other constructs rather than directly—future research should test mediated pathways.

Decision to retain Trust in AIRS: - Theoretical importance (Glikson & Woolley, 2020) - Diagnostic value for organizations - Marginal significance suggests relevance worth monitoring - Future studies with larger N can confirm

We report honestly as 'marginal' rather than claiming significance (§4.6.2, Table 4.10)."

Reference: §4.6.2 Structural Model Results; §5.4.3 Trust as Marginal Predictor; §5.7.3 Statistical Power

5.2 Moderation Effects

Explain the experience moderation effect on Hedonic Motivation. What's the mechanism?

Answer: "Experience moderates $HM \rightarrow BI$ positively ($\beta = .136$, $p = .009$):

The finding: Hedonic Motivation's effect on intention *increases* as AI experience increases (Table 4.12).

Proposed mechanism (§5.4.5):

For low-experience users: - Adoption driven by practical needs (will this help me?) - Fun is ‘nice to have’ but not decision-driving - Focus on learning curve, not enjoyment

For high-experience users: - Competency needs already satisfied - Efficiency gains achieved—now seeking intrinsic satisfaction - ‘Enjoyable to use’ becomes differentiator among AI tools - AI use integrated into professional identity

Theoretical alignment: Self-Determination Theory suggests progression from extrinsic to intrinsic motivation as competency develops. Experienced users have moved beyond instrumental concerns.

Practical implication: Early AI training should emphasize utility; ongoing engagement strategies should emphasize enjoyment and satisfaction (§6.4.1).”

Reference: §4.6.3 Moderation Analysis (Table 4.11); §5.4.5 Experience-Dependent Mechanisms

The population moderation showed HM is stronger for students ($\beta = 0.449$) than professionals ($\beta = -0.301$). Explain this reversal.

Answer: “The HM sign reversal across populations is striking (Table 4.12):

For students (HM $\beta = .449$, $p < .001$): - High autonomy in tool choice - Exploration encouraged (‘try different AI tools’) - Low stakes for experimentation - ‘Fun’ directly influences adoption

For professionals (HM $\beta = -.301$, $p < .05$ negative!): - Tool choice often constrained by organization - Productivity expectations override enjoyment - ‘Fun’ tools may signal lack of seriousness - Negative effect may reflect: ‘If I’m just enjoying it, am I really working?’

Alternative interpretation: The negative coefficient could indicate suppression—professionals who find AI ‘fun’ may already be using it, reducing variance in intention.

Theoretical insight (§5.4.6): Hedonic motivation’s role is highly context-dependent. Consumer adoption models (UTAUT2) may not transfer directly to professional contexts without modification.

Practical implication: Don’t market AI as ‘fun’ to professionals; emphasize productivity and professional development (§6.4.1).”

Reference: §4.6.3 Moderation Analysis (Table 4.12); §5.4.6 Population-Specific Pathways

Why didn’t you find more moderation effects? Was your study underpowered?

Answer: “Limited moderation findings have multiple explanations:

Power considerations (§5.7.3): - Interaction effects typically require larger N than main effects

- N = 523 adequate for main effects; marginal for interactions - Small interactions ($\beta < .10$) may have gone undetected - Post-hoc power for small interactions $\approx .45$

Theoretical explanations: - Age and Gender may be less relevant for AI than traditional technology (digital natives across generations for AI) - AI is sufficiently novel that demographic patterns haven't stabilized - Individual differences (experience, population) more consequential than demographics

What we did find: - Experience \times HM: significant ($p = .009$) - Population moderates HM: significant ($p = .041$) - These effects had adequate power and theoretical foundation

Honest assessment: The null moderation findings for Age and Gender should be interpreted cautiously—absence of evidence \neq evidence of absence given power constraints.

Recommendation: Future studies should oversample for moderation testing ($N > 800$) or focus on specific demographic subgroups (§6.6.4)."

Reference: §4.6.3 Moderation Analysis (Table 4.13); §5.7.3 Statistical Power Considerations

5.3 Exploratory Findings

Describe your four-segment user typology. How stable is this classification?

Answer: "The four-segment typology emerged from K-means cluster analysis (§4.6.8):

Segment	%	Characteristics
Enthusiasts	16%	High on all predictors; early adopters
Cautious Adopters	30%	Moderate-high PV; lower HM; pragmatic users
Moderate Users	37%	Average across factors; mainstream majority
Anxious Avoiders	17%	Low PV, low Trust; AI-hesitant

Source: Compiled by Author

Stability assessment: - Cross-validated using split-half samples - 4-cluster solution optimal (silhouette coefficient = .42) - Segment sizes stable within $\pm 3\%$ across validation

Limitations (§5.7.4): - Exploratory—not hypothesis-driven - K-means assumes spherical clusters - Labels are interpretive, not validated - Requires replication in independent samples

Value: Provides actionable segmentation for practitioners while generating testable hypotheses for future research (§6.6.5)."

Reference: §4.6.8 Exploratory Findings (User Typology); §5.7.4 Exploratory Analysis Limitations

What’s the practical utility of the user typology?

Answer: “The typology enables segment-specific AI implementation strategies (§6.4.2):

For Enthusiasts (16%): - Champion recruitment for AI initiatives - Advanced training and early access programs - Peer influence agents for organizational adoption

For Cautious Adopters (30%): - ROI-focused communication - Concrete use cases with measurable benefits - Gradual rollout with demonstrated value

For Moderate Users (37%): - Mainstream adoption support - Peer normalization messaging - Standard training programs

For Anxious Avoiders (17%): - Trust-building interventions - Low-stakes introduction to AI - Addressing specific concerns (job displacement, errors) - Not forced adoption—may not be appropriate targets

Implementation: AIRS scores can diagnose individual segment membership; organizations can allocate resources accordingly (§6.4.2).

Caveat: Typology is exploratory. Organizations should validate segments in their specific context.”

Reference: §4.6.8 Exploratory Findings; §6.4.2 For Training and Development

The disability-anxiety association ($d = .36$) is an exploratory finding. What do you make of it?

Answer: “This finding requires careful interpretation (§4.6.8):

The observation: - Participants reporting disabilities showed higher AI anxiety ($d = .36$, medium effect) - Association significant despite AI Anxiety scale’s low reliability

Possible explanations: - **Accessibility concerns:** AI interfaces may not accommodate disabilities - **Past negative experiences:** Technology has historically been less accessible - **Fear of replacement:** AI in accessibility tools may raise dependency concerns - **General technology anxiety:** May correlate with disability status

What we CANNOT conclude: - Causality (disability \rightarrow anxiety or confounders) - Mechanism (why this association exists) - Generalizability (sample may not represent disability community)

Importance (§5.5.2): - Highlights AI equity considerations - Signals need for inclusive AI design - Suggests disability community perspectives should inform AI development

Ethical note: We report this finding to surface equity concerns, not to stigmatize. Future research should involve disability community in study design (§6.6.5).”

Reference: §4.6.8 Exploratory Findings; §5.5.2 Equity Implications; §6.6.5 Special Populations

6 Unexpected Findings Questions

6.1 Non-Significant Traditional Predictors

Performance Expectancy—the most robust predictor in UTAUT research—was not significant. How do you explain this?

Answer: “PE non-significance ($\beta = .028$, $p = .464$) is this study’s most theoretically provocative finding:

Explanation 1: Baseline expectation (§5.4.1): - AI capability is assumed—‘of course ChatGPT can write emails’ - Performance gains are expected, not differentiating - When everyone assumes utility, it stops predicting adoption variance

Explanation 2: Evaluation shift (§5.4.1): - Users don’t ask ‘Will AI help me perform better?’ (PE) - Users ask ‘Is AI worth the investment?’ (PV) - Evaluation metric shifted from utility to value

Explanation 3: Uncertainty absorption (§5.4.4): - AI performance gains are hard to quantify (‘How much better?’) - Value captures this uncertainty (‘Worth trying?’) - PV may absorb PE variance through broader framing

Explanation 4: Task specificity: - AI is general-purpose—performance gains vary by task - Hard to form global PE judgment - Value assessment is more holistic

Implication: Don’t showcase AI capabilities; communicate value proposition. Users already believe AI ‘works’—convince them it’s ‘worth it’ (§6.4.1).”

Reference: §4.6.2 Structural Model Results (Table 4.10); §5.4.1 Price Value Dominance; §5.4.4 Non-Significance of Traditional Predictors

Effort Expectancy was NOT significant. Doesn’t this contradict TAM’s “perceived ease of use”?

Answer: “EE was NOT significant ($\beta = -.008$, $p = .875$):

What we found (§4.6.2): - EE effect is essentially zero—not significant - This is a notable departure from TAM/UTAUT2 findings - Ease of use does NOT predict AI adoption intention

Why the null effect: - **Interface maturation:** Modern AI tools (ChatGPT, Copilot) are remarkably user-friendly - **Conversation paradigm:** Natural language interfaces reduce learning curve - **Ceiling effect:** Most AI tools rated ‘easy’—reduced variance eliminates predictive power - **Expectation shift:** Ease is assumed, not a competitive differentiator

TAM compatibility: This finding suggests ease of use becomes non-predictive as interface design improves universally. TAM was developed when software usability varied dramatically; modern AI has largely solved this.

Practical implication: Don’t invest in making AI ‘easier’—it’s already easy enough. Invest in demonstrating value instead (§6.4.1).”

Reference: §4.6.2 Structural Model Results (Table 4.10); §5.4.4 Non-Significance of Traditional Predictors

Facilitating Conditions was NOT significant. What does this mean for organizational AI support programs?

Answer: “FC was NOT significant ($\beta = .059$, $p = .338$):

Interpretation (§5.4.4): - Resources and support do NOT predict adoption intention - Individual psychological factors (PV, HM, SI) dominate - Infrastructure is table stakes, not a differentiator

Why FC doesn’t matter for AI: - **Ubiquitous access:** AI tools available via browser—no special infrastructure - **Low technical barrier:** No software installation, minimal training required - **Self-serve learning:** YouTube, documentation readily available - **Consumer-ization:** AI tools designed for individual use without IT support

Implications for organizations (§6.4.1): - Basic infrastructure is necessary but not sufficient - Don’t expect support programs alone to drive adoption - Focus on value communication and social normalization - FC may become more important for specialized/enterprise AI tools requiring integration

Caveat: FC’s marginal effect may reflect our sample (tech-comfortable online panel). Populations with less technical resources might show stronger FC effects.”

Reference: §4.6.2 Structural Model Results (Table 4.10); §5.4.4 Non-Significance of Traditional Predictors; §6.4.1 For Organizations

Habit was not significant. Given AI tools are relatively new, is this expected?

Answer: “Habit non-significance ($\beta = .043$, $p = .287$) is theoretically expected:

Why non-significant (§5.4.4): - **Temporal requirement:** Habit forms through repeated use over time - **AI novelty:** ChatGPT launched November 2022; data collected 2024 - **Insufficient time:** < 2 years for habitual patterns to develop - **Evolving tools:** AI capabilities changing faster than habits form

UTAUT2 context: - Habit was added for mature technologies (mobile internet) - Assumes stable technology and repeated use patterns - AI is neither stable nor has long usage history

Prediction for future (§6.6.4): - Habit will become significant as AI matures - Longitudinal studies should find increasing habit effects - Early adopters may already show habit formation - Organizations should design for habit formation (daily prompts, workflow integration)

Methodological note: We retained Habit in AIRS for future longitudinal use—it’s theoretically important even if currently non-significant.”

Reference: §4.6.2 Structural Model Results (Table 4.10); §5.4.4 Non-Significance of Traditional Predictors; §6.6.4 Longitudinal Research

6.2 Dropped Constructs

Four constructs were dropped due to low reliability. What went wrong?

Answer: “Four AI-specific constructs failed psychometrically (Table 5.1):

Construct	α	Issue
Voluntariness	.301	Items conflicted conceptually
Explainability	.503	Multi-dimensional (process vs. output)
Ethical Risk	.572	Items scattered across factors
AI Anxiety	.582	Mixed approach/avoidance anxiety

Source: Compiled by Author

What went wrong (§5.6.2):

1. **Two-item limitation:** Complex constructs need 3-4+ items to capture breadth
2. **Conceptual conflation:** Items measured different sub-dimensions that don’t cohere
3. **Insufficient piloting:** Cognitive interviews might have revealed item problems
4. **Novel constructs:** Less precedent for AI-specific measurement

Example—AI Anxiety: - Item 1: ‘AI makes me nervous’ (avoidance anxiety) - Item 2: ‘I worry about missing out on AI’ (FOMO—approach anxiety) - These are conceptually distinct, even opposite, phenomena

This is a finding, not a failure: We learned that these constructs require more sophisticated measurement. Future scale developers have guidance on what doesn’t work (§6.6.3).”

Reference: §4.5.1 EFA Results; §5.6.2 Dropped Constructs (Table 5.1); §6.6.3 Measurement Development

AI Anxiety had $\alpha = .582$. That’s terrible. How did this happen?

Answer: “The .582 reliability reveals conceptual, not just psychometric, problems:

Root cause analysis (§5.6.2):

The items measured two distinct phenomena: - **Avoidance anxiety:** Fear, nervousness, threat appraisal - **FOMO anxiety:** Fear of missing out, competitive concern, opportunity loss

These are psychologically different: - Avoidance anxiety → reduced intention - FOMO anxiety → increased intention - Combining them creates measurement incoherence

Why this wasn’t caught earlier: - Face validity seemed adequate (‘both relate to anxiety about AI’) - Pilot sample (N=25) too small for reliability testing - Literature on AI anxiety is nascent—less guidance available

Theoretical insight: ‘AI Anxiety’ is not unidimensional. Future measures need: - Separate FOMO scale (§6.6.3) - Separate threat anxiety scale - Separate performance anxiety scale (fear of AI errors)

The $\alpha = .582$ tells us something important: People’s anxious reactions to AI are multifaceted and sometimes contradictory.”

Reference: §5.6.2 Dropped Constructs (Table 5.1); §6.6.3 Measurement Development

If you could redesign the dropped constructs, what would you do differently?

Answer: “Specific redesign recommendations for each construct (§6.6.3, Table 5.1):

Voluntariness ($\alpha = .301$): - Separate organizational pressure items from personal choice items - Context-specific: ‘My employer requires...’ vs. ‘I freely choose...’ - 3-4 items per dimension - May need different scales for students vs. professionals

Explainability ($\alpha = .503$): - Separate process explainability (‘I understand how AI works’) - From output explainability (‘I understand why AI gave this answer’) - From explainability importance (‘Explainability matters to me’) - Each sub-dimension: 3 items minimum

Ethical Risk ($\alpha = .572$): - Separate bias concerns, privacy concerns, job displacement concerns
- Each ethical dimension distinct - Domain-specific items (healthcare AI ethics \neq marketing AI ethics) - 3-4 items per ethical sub-dimension

AI Anxiety ($\alpha = .582$): - FOMO scale: 3-4 items on competitive concern - Threat scale: 3-4 items on fear/nervousness - Performance anxiety: 3-4 items on AI error concerns - Cognitive interviews to ensure item interpretation

Process improvements: - Larger pilot sample ($N \geq 100$) for preliminary reliability - Cognitive interviewing before full deployment - Literature review of similar scales in adjacent domains

These recommendations inform future AIRS iterations (§6.6.3)."

Reference: §5.6.2 Dropped Constructs (Table 5.1); §6.6.3 Measurement Development Recommendations

7 Limitations and Validity Questions

7.1 Internal Validity

What threats to internal validity exist in your study?

Answer: "I acknowledge four primary internal validity threats (§5.7.1):

1. Common Method Variance (CMV): - Single source (self-report), single method (survey), single time point - Mitigation: Harman's test (38.2% first factor), procedural controls, discriminant validity - Residual risk: Cannot fully rule out; R^2 may be inflated

2. Self-Selection Bias: - Online panel members may be more technology-comfortable - Centiment's topic-blind recruitment partially mitigates - Sample may underrepresent technology-averse populations

3. Social Desirability: - Respondents may overstate AI interest or understate anxiety - Mitigation: Anonymity assurance, indirect items - Price Value's dominance may partially reflect this (socially acceptable to be value-conscious)

4. Recall/Attribution Bias: - Retrospective reports of AI use may be inaccurate - Intentions measured at single point may not reflect stable attitudes

Honest assessment: These are inherent to survey research. We implemented standard mitigations but cannot eliminate these threats. Longitudinal, multi-method follow-up recommended (§6.6.4)."

Reference: §5.7.1 Internal Validity Limitations; §4.4 Data Screening

How do you rule out reverse causality? Maybe behavior influences intention rather than vice versa?

Answer: “I cannot rule out reverse causality with cross-sectional data—this is explicitly acknowledged (§5.7.1):

The concern: - We assume: $PV \rightarrow BI$ (Value perceptions cause intentions) - Alternative: $BI \rightarrow PV$ (People who intend to adopt retrospectively justify via value) - Cross-sectional design cannot distinguish these

Partial defense: 1. **Theoretical precedent:** TRA, TPB, TAM, UTAUT all posit beliefs \rightarrow intention \rightarrow behavior, supported by longitudinal studies 2. **Meta-analytic support:** Intention \rightarrow behavior correlation of $\rho = .53-.69$ across hundreds of studies 3. **Temporal framing:** Intention items ask about future (‘intend to use in next 3 months’), reducing retrospective bias

What would strengthen causal inference: - Panel design: Measure predictors at T1, intention at T2, behavior at T3 - Experimental manipulation of Price Value communication - Natural experiment during AI tool rollout

Bottom line: Causality is assumed based on theory, not demonstrated. This is typical for scale validation studies but should inform interpretation (§5.7.1).”

Reference: §5.7.1 Internal Validity Limitations; §6.6.4 Methodological Extensions

Could there be unmeasured confounders affecting your results?

Answer: “Yes, several potential confounders were not measured (§5.7.1):

Likely confounders: - **Personality:** Openness to experience, need for cognition may affect both predictors and intention - **Prior AI experiences:** Negative past experiences could drive both low trust and low intention - **Organizational culture:** Innovation-friendly vs. risk-averse environments affect both social influence and intention - **Job characteristics:** Role autonomy, task complexity, performance pressure - **Digital literacy:** May affect both effort expectancy and adoption

Partial controls: - Age, gender, experience entered as covariates - Population (student/professional) analyzed separately - But not comprehensive confound control

Impact if confounders exist: - Path coefficients may be biased (likely upward if confounders correlate positively with both predictor and DV) - $R^2 = .852$ may be partially explained by unmeasured variables - Relative ordering of predictors may shift

Recommendation: Future studies should include personality measures (Big Five), organizational climate scales, and digital literacy assessments (§6.6.4).”

Reference: §5.7.1 Internal Validity Limitations

7.2 External Validity

To what populations do your findings generalize?

Answer: “Findings confidently generalize to a bounded population (§5.7.2):

Strong generalization: - U.S. adults in professional and academic contexts - English-speaking knowledge workers - Technology-accessible populations (internet users) - Age 18-65 in education or white-collar employment

Uncertain generalization: - Blue-collar/manual labor workers - Non-English speakers - Older adults (65+) outside workforce - Populations without reliable internet access - Government or highly regulated sectors

Likely non-generalizable: - Non-Western cultures (different value orientations) - AI-specific contexts (healthcare AI, autonomous vehicles) - Populations with no AI exposure

The pragmatic interpretation: AIRS is validated for the populations most likely to encounter AI in near-term professional contexts—a meaningful and useful scope, even if not universal.”

Reference: §5.7.2 External Validity Limitations; §4.3 Participant Characteristics

Would your findings replicate in a non-Western context?

Answer: “Unknown—cross-cultural validity was not tested and is an important limitation (§5.7.2):

Constructs likely affected by culture:

- **Price Value:** Collectivist cultures may weigh collective benefit over personal value; price sensitivity varies by economic context
- **Social Influence:** Stronger in collectivist cultures (e.g., East Asia); different referent groups (family vs. colleagues)
- **Trust:** Trust formation differs culturally; institutional trust vs. interpersonal trust emphasis varies
- **Hedonic Motivation:** May be less acceptable to prioritize enjoyment in high power-distance cultures

Constructs likely more stable: - Performance Expectancy (functional benefits are relatively universal) - Effort Expectancy (usability perceptions may be stable)

Research needed: - Translation and back-translation of AIRS - Multi-country data collection - Measurement invariance testing across cultures - Hofstede dimension correlations with path coefficients

Practical note: Organizations operating globally should not assume AIRS findings transfer without local validation (§6.6.1).”

Reference: §5.7.2 External Validity Limitations; §6.6.1 Replication Studies

Your sample is from 2024-2025. AI is evolving rapidly. Will your findings still be relevant in 5 years?

Answer: “Partially—different aspects have different temporal stability (§5.7.2):

Likely stable (psychological mechanisms): - Value-based evaluation process - Social influence in adoption decisions - Trust importance for uncertain technologies - Hedonic motivation’s role in discretionary adoption - The UTAUT2 framework structure

Likely to evolve (specific perceptions): - Price Value: As AI becomes commoditized, PV dominance may decrease - Performance Expectancy: May become more predictive as AI differentiates - Habit: Will likely become significant as AI matures - Trust: May increase or decrease depending on AI track record

Instrument implications: - AIRS structure likely stable - Item wording may need updating (‘AI tools’ → specific technologies) - Normative benchmarks will shift - Dropped constructs may become measurable with improved items

Recommendation: Periodic revalidation (every 2-3 years) to track construct stability and update norms (§6.6.4).”

Reference: §5.7.2 External Validity Limitations; §6.6.4 Longitudinal Research

7.3 Construct Validity

How do you know your constructs measure what you claim they measure?

Answer: “Construct validity was established through multiple approaches (§4.5.3):

1. Content validity: - Items adapted from validated UTAUT2 scales (Venkatesh et al., 2012) - AI-specific items grounded in Glikson & Woolley (2020) trust framework - Expert review by 3 subject matter experts - Pilot testing for clarity and interpretation

2. Convergent validity: - All factor loadings > .70 (.75–.91 range) - All AVE > .50 (.60–.79 range) - Items within constructs correlate strongly

3. Discriminant validity: - Fornell-Larcker: $\sqrt{\text{AVE}} > \text{all inter-construct correlations}$ - HTMT ratios $< .85$ for all pairs - No correlations $> .85$ suggesting construct overlap

4. Nomological validity: - Constructs relate to intention as theory predicts (mostly) - Pattern of significant/non-significant paths interpretable - R^2 consistent with UTAUT literature

Limitation acknowledged: We rely on self-report for all constructs. Behavioral measures of actual AI use would strengthen validation (§5.7.1).”

Reference: §4.5.3 Construct Validity (Tables 4.5, 4.6); §3.4 Instrumentation

Is “Behavioral Intention” a good proxy for actual behavior?

Answer: “Imperfect but standard and defensible (§5.7.1):

Empirical support: - Meta-analysis (Sheeran, 2002): Intention-behavior $r = .53$ - UTAUT-specific meta-analysis: $r = .69$ (Venkatesh et al., 2016) - Intention explains ~25-50% of behavior variance

Why intention, not behavior: - Scale validation study—intention is appropriate DV - Actual behavior harder to measure reliably (self-report, logs vary) - Intention captures psychological readiness independent of opportunity - Standard in TAM/UTAUT tradition

Known intention-behavior gaps: - Implementation intentions not formed - Situational constraints prevent action - Competing priorities override intention - Time delay between intention and opportunity

For AI specifically: - Gap may be larger (rapid AI evolution, organizational constraints) - Or smaller (AI easily accessible, low barrier) - Empirical question for longitudinal research

Recommendation: Future studies should measure actual AI tool usage (log data, observed behavior) to validate intention-behavior link (§6.6.4).”

Reference: §5.7.1 Internal Validity Limitations; §3.4.2 Core UTAUT2 Constructs

How did you ensure content validity of your AI-specific items?

Answer: “Content validity was established through systematic development (§3.4.3, §3.4.4):

Step 1: Literature grounding: - AI Trust items based on Glikson & Woolley (2020) three-dimensional framework - AI-specific constructs informed by Venkatesh (2021) AI research challenges - Reviewed 50+ AI adoption papers for construct coverage

Step 2: Item generation: - Generated 3-4 candidate items per construct - Ensured items captured construct breadth (not single facet) - Balanced positive/negative wording where appropriate

Step 3: Expert review: - Three experts reviewed items for: - Relevance to construct definition - Clarity of wording - Appropriateness for target population - Items revised based on feedback

Step 4: Pilot testing (N = 25): - Cognitive interviews with subset - Clarity and interpretation checked - Problematic items revised

Limitation: Dropped constructs (Voluntariness, Explainability, Ethical Risk, AI Anxiety) suggest content validity was insufficient for these—items may not have captured construct breadth adequately. Future development should expand item pools and pilot more extensively (§5.6.2).”

Reference: §3.4.3 AI-Specific Extensions; §3.4.4 Pilot Testing; §5.6.2 Dropped Constructs

7.4 Statistical Conclusion Validity

Given your sample size, what effects might you have missed?

Answer: “Power analysis suggests small effects may be undetected (§5.7.3):

Power for different effect sizes (N = 262 CFA sample): - Large effects ($\beta > .30$): Power $> .99$ —very unlikely to miss - Medium effects ($\beta = .15-.30$): Power $\approx .80-.95$ —adequately powered - Small effects ($\beta = .05-.15$): Power $\approx .35-.70$ —may miss - Very small effects ($\beta < .05$): Power $< .35$ —likely to miss

Effects potentially missed: - AI Trust ($\beta = .106$, $p = .064$)—may be real but underpowered - Small moderation effects for Age, Gender - Indirect/mediation effects

Required N for 80% power: - Trust effect ($\beta = .063$): $N \approx 600$ - Very small effects ($\beta = .05$): $N \approx 800$

Implication: The marginally significant Trust and FC effects should not be dismissed as ‘null.’ They may represent real but small effects. Future studies with larger N can confirm (§6.6.4).

What we’re confident about: Large effects (PV: $\beta = .505$, SI: $\beta = .197$) are robust—highly unlikely to be Type I errors.”

Reference: §5.7.3 Statistical Power Considerations

Did you adjust for multiple comparisons? Why or why not?

Answer: “No formal adjustment was made, for principled reasons (§3.5.3):

Why no Bonferroni/FDR adjustment:

1. **Hypothesis-driven testing:** 12 a priori hypotheses, not fishing expedition. Adjustments penalize confirmatory tests unnecessarily.

2. **SEM context:** Path coefficients estimated simultaneously in integrated model, not independent tests. Traditional correction assumptions violated.
3. **Trade-off consideration:** Bonferroni with 12 tests would require $p < .004$. This risks missing real effects (Type II errors) in scale validation where detecting true relationships matters.

How we addressed multiplicity: - Clearly distinguished confirmatory (12 hypotheses) from exploratory analyses - Exploratory findings (typology, disability-anxiety) flagged as requiring replication - Effect sizes reported alongside p-values - Confidence intervals provided for interpretation

Alternative view: A reviewer might legitimately argue for adjustment. With Bonferroni ($\alpha = .05/12 = .004$): - PV, SI, HM remain significant - EE becomes non-significant - Trust, FC remain non-significant

Core findings robust to adjustment; marginal findings require replication regardless.”

Reference: §3.5.3 Structural Equation Modeling; §5.7.3 Statistical Considerations

How sensitive are your results to outliers or influential observations?

Answer: “Results are robust to outlier influence (§4.4):

Outlier detection performed: - **Univariate:** z-scores $> |3.29|$ checked—0 extreme outliers on key variables - **Multivariate:** Mahalanobis distance at $p < .001$ —0 multivariate outliers identified - **Influential observations:** Cook’s $D < 1.0$ for all cases

Sensitivity checks conducted: - Compared results with and without top/bottom 1% of scores - Path coefficients stable within $\pm .02$ - Significance patterns unchanged

Robust estimation: - MLM estimation provides corrected standard errors - Bootstrap CIs (5,000 resamples) less sensitive to distributional assumptions - Results consistent across estimation methods

Why relatively clean data: - Centiment’s panel screening removes careless responders - Attention checks filtered low-quality responses - Complete-case design prevented missing data issues

Confidence statement: Results are not driven by outliers or influential cases. The PV dominance finding is robust across subsamples and estimation approaches.”

Reference: §4.4 Data Screening; §4.6.2 Structural Model Results (Bootstrap CIs in Table 4.10)

8 Practical Implications Questions

8.1 Organizational Application

If an organization wanted to use the AIRS tomorrow, what would you tell them?

Answer: “AIRS is ready for diagnostic use with appropriate caveats (§6.4):

What they CAN do now: - Administer the 16-item instrument to employees - Calculate mean scores for each of 8 factors - Identify relative strengths and weaknesses (e.g., ‘Low PV, high SI’) - Use qualitative patterns to inform strategy - Compare departments or teams on factor profiles

What they should be CAUTIOUS about: - No normative benchmarks yet (‘What’s a good score?’) - Individual-level prediction requires more validation - Intervention effectiveness not yet tested

Recommended approach: 1. Use AIRS as diagnostic input, not sole decision tool 2. Combine with qualitative interviews for context 3. Focus on factor patterns, not absolute scores 4. Partner with researchers for rigorous evaluation

My offer: Academic collaboration welcome for organizations wanting rigorous AIRS implementation (§6.5).”

Reference: §6.4 Practical Contributions; §6.5 Research-Practice Partnerships

How should organizations interpret AIRS scores? What’s a “good” score?

Answer: “Interpretation guidance without normative benchmarks (§6.4.3):

Current interpretation approach: - **Within-organization comparison:** Compare departments, roles, experience levels - **Factor profile analysis:** Look for patterns (e.g., high Trust + low PV = different intervention than low Trust + low PV) - **Relative priority:** Rank factors by score to identify intervention targets - **Change tracking:** Measure before/after interventions

Why no absolute benchmarks: - Single validation sample (N = 523) - Population-specific norms needed (different for healthcare vs. tech industry) - Temporal instability (benchmarks from 2024 may not apply in 2026)

Preliminary guidance from our sample (use cautiously): - Mean scores ranged ~2.8–4.2 across constructs (5-point scale) - Score < 3.0 likely indicates adoption barrier - Score > 4.0 suggests facilitative factor

Future development: Normative database planned with industry-specific benchmarks (§6.6.6).”

Reference: §6.4.3 For Practitioners; §4.3 Participant Characteristics (Descriptive Statistics)

What specific interventions would you recommend based on your findings?

Answer: “Factor-specific intervention hypotheses based on structural model results (§6.4.1, §6.4.2):

For low Price Value ($\beta = .505$ —highest priority): - Clear ROI communication with specific metrics - Time savings documentation (‘Save 2 hours/week’) - Career advancement framing (‘Competitive skill’) - Free/subsidized access to reduce barrier

For low Social Influence ($\beta = .197$): - Peer champion programs - Visible adoption by respected leaders - Team-based AI initiatives - Sharing success stories

For low Hedonic Motivation ($\beta = .110$): - Gamification elements - Low-stakes exploration opportunities - Emphasize creative applications - Note: Less important for professionals (population moderation)

For low Trust (marginal but important): - Transparency about AI limitations - Human oversight emphasis - Gradual trust-building with low-stakes tasks - Error recovery demonstrations

Critical caveat: These are theoretically derived recommendations. Intervention effectiveness requires randomized controlled trials—not yet conducted (§6.6.5).”

Reference: §6.4.1 For Organizations Implementing AI; §6.4.2 For Training and Development

Your study suggests “lead with value, not capabilities.” How would this change a typical AI rollout?

Answer: “The ‘value lens’ finding (PV $\beta = .505$, PE $\beta = .028$ ns) suggests reframing AI communication (§5.5.1):

Traditional rollout (capability-focused): - ‘Look what AI can do!’ (demos, features) - ‘AI will improve your performance!’ (PE emphasis) - Training on functionality - Assumes users will recognize value

Value-focused rollout (recommended): - ‘Here’s what AI will save you’ (time, effort quantified) - ‘The ROI for your role is...’ (personalized value) - Training on workflow integration - Explicit cost-benefit presentation

Specific changes: 1. **Pre-launch:** Document specific value metrics by role 2. **Launch communication:** Lead with ‘worth it’ message, not ‘look at this’ 3. **Training:** Include ROI module before feature training 4. **Follow-up:** Track and communicate realized value 5. **Support:** Help users see their personal value realization

Why this matters: Users already assume AI ‘works’ (PE). The adoption decision is ‘Is it worth my investment?’ Capability demonstrations don’t answer that question (§5.3.1).”

Reference: §5.3.1 Price Value Dominance; §5.5.1 For Organizations; §6.4.1 For Organizations Implementing AI

8.2 Stakeholder-Specific

What should AI tool vendors learn from your research?

Answer: “Vendor implications from AIRS findings (§5.5.3):

Pricing strategy: - Price Value dominates—pricing decisions are adoption decisions - Freemium models may be optimal (let users discover value before paying) - Clear value proposition at every price tier - Consider value-based pricing over feature-based

Product development: - Hedonic Motivation matters—enjoyable UX differentiates - Trust-building features (transparency, explanation, human oversight options) - Easy onboarding (though EE less important than expected) - Integration with existing workflows (demonstrate value quickly)

Marketing communications: - Lead with value/ROI, not capabilities - Social proof (adoption by respected others) - Segment messaging: Students (fun), Professionals (efficiency)

Trust considerations: - Transparency about limitations - Clear data handling practices - Reliability track record - Error recovery demonstrations

Key insight: Vendors competing on features may be optimizing the wrong dimension. Compete on value demonstration and enjoyable experience (§5.3.1, §5.3.2).”

Reference: §5.5.3 For AI Vendors; §5.3.1 Price Value Dominance

How should HR and training departments use your findings?

Answer: “HR/L&D applications from AIRS (§6.4.2):

Pre-training assessment: - Administer AIRS to identify training cohort profiles - Segment learners by dominant barriers - Customize training approach by segment

Segment-specific training design: - **Enthusiasts (16%):** Advanced content, peer mentor roles - **Cautious Adopters (30%):** ROI-focused, practical applications - **Moderate Users (37%):** Standard curriculum, peer support - **Anxious Avoiders (17%):** Anxiety-aware design, gradual exposure, low stakes

Training content priorities: - Value realization module (address PV) - Social learning components (leverage SI) - Hands-on, enjoyable activities (build HM) - Trust-building exercises (AI limitations, human oversight)

Accessibility considerations: - Disability-anxiety association ($d = .36$) suggests accessibility awareness - Multiple modalities for training delivery - Accommodation-friendly AI tool selection

Evaluation: - Pre/post AIRS to measure intervention impact - Track adoption behavior, not just satisfaction

Warning: Don't use AIRS to screen out employees—use to support them (§5.5.2)."

Reference: §6.4.2 For Training and Development; §4.6.8 Exploratory Findings (User Typology)

What's the message for C-suite executives?

Answer: "Executive summary of AIRS implications (§5.5.1):

Key message: AI adoption is a human psychology problem, not a technology problem.

Why AI pilots fail (90-95%): - Not because AI doesn't work (it does) - Because organizations lead with capabilities when employees evaluate value - One-size-fits-all approaches ignore psychological diversity

What executives should understand: 1. **Investment priority:** Budget for change management and value communication, not just technology 2. **Metrics that matter:** Track perceived value and adoption intent, not just training completion 3. **Segmentation:** 17% are Anxious Avoiders who need different approaches—don't lose them 4. **Social dynamics:** Adoption spreads through influence—identify and support champions

Strategic implications: - AI ROI depends on adoption; adoption depends on psychology - Competitor advantage from AI requires human adoption success - Early investments in understanding barriers pay off in deployment

The diagnostic question: 'Do our employees see AI as worth their investment?' If not, capability demonstrations won't help.

AIRS utility: Provides diagnostic data for strategic decisions about AI implementation (§6.4)."

Reference: §1.1 Background and Context; §5.5.1 For Organizations; §6.4 Practical Contributions

9 Future Research Questions

9.1 Immediate Extensions

What's the single most important study that should follow from your dissertation?

Answer: “The single most important follow-up study would be a **larger-sample replication** ($N > 600$) to definitively establish the AI Trust effect. While my study found Trust marginally significant ($\beta = .106$, $p = .064$), a larger sample would provide the statistical power needed to confirm whether this represents a genuine population effect or sampling variability. With 80% power targeting a small effect ($\beta \approx .10$), approximately 610 participants would be needed.

Alternatively, a **longitudinal validation study** tracking the intention-behavior relationship over 6-12 months would address the cross-sectional limitation and confirm whether my strong behavioral correlation ($\rho = .69$) holds prospectively. This would transform AIRS from a predictive tool to a validated forecasting instrument.”

Reference: §6.3.2 Future Research Directions; §5.4.3 AI Trust Findings

How would you redesign the dropped constructs?

Answer: “For **AI Anxiety**, I would reconceptualize the construct by separating two distinct dimensions that were conflated: (1) **performance-related anxiety** (worry about AI making mistakes or job replacement) and (2) **engagement avoidance** (reluctance to interact with AI systems). Each dimension would have 4+ items with clear behavioral anchors rather than general affect statements.

For **AI Explainability**, the issues were the ‘black box’ metaphor and conflation of different concepts. I would separate: (1) **understanding preference** (desire to know how AI reaches conclusions) and (2) **explanation necessity** (requirement for explanations before trusting output). Items would reference concrete scenarios rather than abstract concepts, and avoid technical jargon like ‘black box’ that loaded inconsistently.”

Reference: §5.2.2 Construct Elimination; §6.3.2 Future Research Directions

What would a longitudinal follow-up study look like?

Answer: “A rigorous longitudinal design would include: - **Timeline:** 6-12 month panel study with assessments at baseline, 3 months, 6 months, and 12 months - **Sample:** Minimum $N = 400$ at baseline, accounting for 30% attrition - **Measures:** AIRS-16 at each wave, plus **objective usage metrics** (e.g., AI tool login frequency, feature utilization from organizational systems) - **Key analyses:** Cross-lagged panel models to establish temporal precedence; habit formation trajectories; threshold effects for intention → behavior conversion - **Design features:** Include organizational

AI implementation events as natural experiments; track both new and experienced users

This would address the primary limitation of cross-sectional data while examining whether the strong BI-Usage correlation ($\rho = .69$) represents genuine predictive validity.”

Reference: §6.3.2 Future Research Directions; §6.2.1 Cross-Sectional Limitations

9.2 Theoretical Development

Should there be an “AI-TAM” or “AI-UTAUT” that’s distinct from general technology acceptance models?

Answer: “My findings strongly suggest yes—AI requires its own acceptance framework distinct from general technology. The evidence includes:

1. **Price Value dominance** ($\beta = .505$): Economic considerations dominate AI adoption in ways not seen with simpler technologies
2. **Performance Expectancy non-significance** ($\beta = -.028$, $p = .791$): The traditional ‘usefulness’ construct that dominates TAM/UTAUT doesn’t predict AI adoption
3. **Trust emergence:** AI-specific trust considerations require dedicated measurement
4. **Construct measurement challenges:** Anxiety and Explainability required AI-specific reconceptualization

An ‘AI-TAM’ or ‘AI-UTAUT’ would likely feature: (1) Economic rationality as the primary driver, (2) Trust and transparency as core constructs, (3) Hedonic elements emphasized, and (4) Traditional performance expectations reconceptualized around augmentation rather than replacement.”

Reference: §6.1.1 Theoretical Contribution; §5.4.1 Price Value Dominance

How might your model need to change as AI becomes more autonomous and high-stakes?

Answer: “As AI moves from assistive tools to autonomous decision-makers, the model would likely need several modifications:

1. **Trust becomes central:** The marginal effect I found ($\beta = .106$, $p = .064$) would likely strengthen substantially when AI makes consequential decisions without human oversight
2. **New constructs needed:**
 - **Perceived autonomy threat** (concern about loss of control)
 - **Ethical alignment** (confidence AI decisions match personal values)
 - **Accountability clarity** (understanding who is responsible for AI decisions)

3. **Risk tolerance moderators:** Individual differences in risk acceptance would become significant
4. **Context-specificity:** The model may need domain-specific variants (healthcare AI vs. financial AI vs. creative AI)

My current model captures AI in its ‘tool’ phase; autonomous AI may require a fundamentally different theoretical architecture.”

Reference: §6.3.2 Future Research Directions; §6.1.2 AI Trust Discussion

What role might personality traits play in AI adoption?

Answer: “While not measured in my study, several personality dimensions are plausible moderators or predictors:

1. **Openness to experience:** Likely positive predictor; correlates with technology curiosity and willingness to try novel tools
2. **Need for cognition:** May moderate Hedonic Motivation effects—high NFC individuals might value intellectual challenge of AI
3. **Risk tolerance:** Particularly relevant for Trust relationships; risk-averse individuals may require stronger trust before adoption
4. **Locus of control:** Internal locus may correlate with adoption confidence; external may increase anxiety
5. **Perfectionism:** May create barriers through fear of AI errors or uncertainty

Future research could incorporate Big Five personality measures or domain-specific traits (e.g., technology anxiety as personality dimension) to explain variance beyond AIRS constructs. My user typology (Enthusiasts, Pragmatists, etc.) may partly reflect underlying personality differences.”

Reference: §6.3.2 Future Research Directions; §5.5.2 User Segment Characteristics

9.3 Practical Development

Describe your research roadmap from validated scale to organizational applications.

Answer: “I envision a four-phase research program:

Phase 1 - AIRS Scoring System (1-2 years): - Develop normative data across populations - Create standardized scoring algorithms (T-scores, percentiles) - Establish clinically meaningful cut-points for low/medium/high readiness - Validate scoring against adoption outcomes

Phase 2 - Diagnostic Protocols (2-3 years): - Develop factor-level interpretation guidelines - Cre-

ate organizational assessment frameworks - Design automated reporting systems - Validate segment classification algorithms

Phase 3 - Intervention Frameworks (3-4 years): - Design segment-specific interventions (Anxious Avoider support, Enthusiast leverage) - Test intervention effectiveness through RCTs - Develop training curricula tied to AIRS profiles - Create organizational change management protocols

Phase 4 - Comprehensive System (4-5 years): - Integrated assessment-intervention platform - Continuous monitoring and adaptation - Cross-organizational benchmarking - ROI validation studies”

Reference: §6.4.3 Research Roadmap; §6.4.1 Organizational Applications

What would intervention effectiveness research look like?

Answer: “Rigorous intervention testing would require:

Design: Randomized controlled trials with: - **Treatment groups:** Segment-specific interventions based on AIRS profiles - **Control groups:** Standard training or no intervention - **Sample size:** Minimum N = 200 per condition for adequate power

Interventions to test: - *For Anxious Avoiders:* Anxiety reduction + hands-on exposure + peer support - *For Cautious Adopters:* Clear ROI demonstrations + value communication - *For Moderate Users:* Low-stakes experimentation opportunities + social proof - *For Enthusiastic Adopters:* Advanced training + mentor role assignments

Outcomes: - Primary: Change in AIRS scores (pre-post) - Secondary: Actual AI tool adoption metrics (usage frequency, feature utilization) - Tertiary: Productivity measures, satisfaction, sustained usage at 3-6 months

Analysis: Mixed-effects models accounting for organizational clustering; cost-effectiveness analysis.”

Reference: §6.4.1 Organizational Applications; §6.3.2 Future Research Directions

How could your typology be used for personalized AI training?

Answer: “The four-segment typology enables differentiated training approaches:

Enthusiastic Adopters (16%): - Role: Deploy as AI champions and peer mentors - Training: Advanced features, integration strategies, troubleshooting - Engagement: Early access to new tools, feedback collection - Caution: Monitor for overconfidence and risky usage

Cautious Adopters (30%): - Approach: Lead with concrete value demonstrations - Training: Prac-

tical use cases, efficiency gains, ROI evidence - Engagement: Clear guidelines, structured support
- Key message: ‘This is worth your time investment’

Moderate Users (37%): - Approach: Low-stakes experimentation, social proof - Training: Sand-boxed environments, peer success stories, gradual exposure - Engagement: Buddy systems, optional participation initially - Key message: ‘Others like you are succeeding’

Anxious Avoiders (17%): - Approach: Address emotional barriers before skill training - Training: Anxiety management, confidence building, job security reassurance - Engagement: One-on-one support, validated concerns - Key message: ‘Your concerns are heard; this is supportive, not threatening’

This framework requires experimental validation but aligns with differentiated instruction principles.”

Reference: §5.5 User Typology Results; §6.4.2 Practitioner Recommendations

10 Philosophical and Broader Impact Questions

10.1 AI and Society

What are the ethical implications of measuring and potentially manipulating AI adoption readiness?

Answer: “This is an important ethical consideration. There are genuine tensions between:

1. **Enabling vs. coercing adoption:** AIRS should be used to identify barriers and provide appropriate support, not to pressure or manipulate employees into adoption. The goal is informed choice, not compliance.
2. **Individual autonomy:** Employees have the right to resist technology they find objectionable. Assessment should identify concerns to address, not weaknesses to exploit.
3. **Organizational interests vs. individual welfare:** Organizations may want rapid adoption; individuals may have legitimate concerns about job security, surveillance, or values misalignment.

Ethical guidelines I would recommend: - AIRS results should be used for support allocation, not performance evaluation - Individual profiles should remain confidential - ‘Low readiness’ should never be treated as a deficiency - Participation in assessment should be voluntary where possible

Researchers have a responsibility to anticipate misuse and provide guidance for ethical implementation.”

Reference: §6.4.2 Practitioner Recommendations; §6.3.3 Ethical Considerations

Your “Anxious Avoider” segment might resist AI for valid reasons. Should we try to change them?

Answer: “This is a nuanced question. My position is that we should **address barriers without invalidating concerns**. Anxious Avoiders (17% of my sample) may resist AI for entirely legitimate reasons:

- **Valid concerns:** Job displacement fears, privacy worries, algorithmic bias, loss of human judgment in important decisions
- **Healthy skepticism:** Critical evaluation of AI claims and limitations is valuable
- **Personal values:** Some may prioritize human connection over efficiency

What we should do: - Listen to and validate their concerns - Address concrete barriers (training, support, job security assurances) - Provide accurate information about AI capabilities and limitations - Create opt-out pathways where appropriate - Leverage their critical perspective for identifying AI risks

What we should NOT do: - Treat resistance as a problem to be fixed - Use peer pressure or performance penalties - Dismiss concerns as ‘technophobia’ or ignorance

Some resistance may represent the kind of healthy skepticism organizations need to implement AI responsibly.”

Reference: §5.5.2 Anxious Avoider Profile; §6.4.2 Practitioner Recommendations

Could your instrument be misused to discriminate against employees who are AI-resistant?

Answer: “Yes, this is a real risk that must be addressed proactively. Potential misuse scenarios include:

- **Hiring discrimination:** Using low AIRS scores to screen out job candidates
- **Performance evaluation:** Treating low readiness as poor performance
- **Promotion decisions:** Favoring ‘AI-ready’ employees regardless of job requirements
- **Termination justification:** Using assessment results to build cases for dismissal

Safeguards I recommend: 1. **Clear policy statements:** AIRS is a developmental tool, not an evaluative one 2. **Confidentiality protections:** Individual results should not be shared with managers making personnel decisions 3. **Aggregate-only reporting:** Organizations receive group-level in-

sights, not individual profiles 4. **Voluntary participation:** Assessment should be opt-in where possible 5. **Ethical review:** Organizations should establish oversight for AI readiness programs

These concerns apply to any psychometric instrument; AIRS is not unique in this regard, but the AI context makes vigilance especially important.”

Reference: §6.3.3 Ethical Considerations; §6.4.2 Implementation Guidelines

What responsibility do researchers have in an AI adoption arms race?

Answer: “Researchers have significant ethical responsibilities in this space:

1. **Honest communication:** Don’t oversell findings or promise organizational transformations that aren’t supported by evidence
2. **Acknowledge limitations:** Be transparent about what the research can and cannot tell us
3. **Consider equity:** AI adoption may widen gaps between those with resources and those without; consider implications for workforce segments
4. **Anticipate misuse:** Think through how findings could be weaponized and provide guidance
5. **Question assumptions:** Interrogate whether ‘more AI adoption’ is inherently good

My specific responsibilities with AIRS: - Provide ethical implementation guidelines alongside the instrument - Acknowledge that low readiness may reflect valid concerns, not deficiencies - Avoid framing resistance as something to be overcome - Support balanced discourse about AI’s benefits and risks - Make research accessible beyond organizations that can afford consulting services

Science should inform decisions, not drive predetermined agendas.”

Reference: §6.3.3 Ethical Considerations; §6.4.3 Research Responsibility

10.2 Broader Implications

What does your research say about the future of human work in an AI-enabled world?

Answer: “Several key insights emerge from my findings:

1. **Adoption is not automatic:** Despite AI availability, human psychology creates substantial barriers and facilitators. Technology alone doesn’t determine outcomes.
2. **Economic rationality dominates:** The Price Value dominance ($\beta = .505$) suggests workers evaluate AI through cost-benefit lenses—will this help me more than it costs me in time, learning, and disruption?
3. **Organizational context matters:** Social Influence ($\beta = .136$) and Facilitating Conditions

(implied in practical implications) shape individual decisions. Workplaces, not just technologies, determine adoption.

4. **Individual differences persist:** My four-segment typology shows enduring variation in how people approach AI—some enthusiastically, some anxiously, some pragmatically.

Implications for future of work: - AI won't simply 'replace' workers—implementation is mediated by human acceptance - Organizations that attend to psychological factors will have adoption advantages - Workers who develop AI fluency have strategic value, but 'resistance' may also be valued - The human-AI relationship will be negotiated, not determined"

Reference: §6.1.3 Practical Implications; §6.4.1 Organizational Applications

If Price Value dominates, does this mean AI will only be adopted when it's free or cheap?

Answer: "No—Price Value is about perceived **value relative to costs**, not absolute cost. The construct captures whether the benefits of using AI are worth the investment required. This includes:

Benefits considered: - Time savings and efficiency gains - Quality improvement in outputs - Competitive advantage - Learning and skill development - Reduced cognitive burden

Costs considered: - Financial costs (subscriptions, training) - Time investment in learning - Disruption to established workflows - Risk of errors during transition - Cognitive effort to adopt new tools

Practical implications: - Expensive AI tools can achieve high adoption if perceived benefits are substantial - Free tools may be rejected if learning costs seem too high - Organizations should emphasize value proposition, not just price - ROI communication should be concrete and personally relevant

My findings suggest that professionals are **sophisticated evaluators** of technology value—they're not simply cost-averse, but value-conscious."

Reference: §5.4.1 Price Value Dominance; §6.4.2 Value Communication

Your findings suggest experienced professionals value enjoyment more. What does this mean for workforce development?

Answer: "The Hedonic Motivation finding ($\beta = .217$, second strongest predictor) with usage moderation suggests important workforce development implications:

The pattern: More experienced AI users show stronger relationships between enjoyment and adoption intention, suggesting that as expertise develops, intrinsic motivation becomes more important.

Implications for different career stages:

Early career/novice: - Focus on utility and practical benefits - Provide clear ROI demonstrations - Structured training with measurable outcomes

Mid-career/developing expertise: - Balance utility with engagement - Introduce more autonomous exploration - Highlight intellectual challenge

Experienced professionals: - Emphasize enjoyment and creative potential - Allow experimentation and customization - Position AI as intellectually stimulating, not just efficient

Broader insight: Technology training should evolve with employee development. One-size-fits-all approaches miss the motivational shift from extrinsic (what will this do for me?) to intrinsic (is this interesting and enjoyable?) drivers.”

Reference: §5.4.4 Moderation Results; §6.4.2 Training Recommendations

11 Technical Deep-Dive Questions

11.1 Statistical Technicalities

Explain the difference between Cronbach’s alpha, composite reliability, and AVE. Why report all three?

Answer: “These three metrics assess different aspects of measurement quality:

Cronbach’s alpha (α): - Measures internal consistency - Assumes **tau-equivalence** (all items contribute equally) - Range: 0-1; acceptable $\geq .70$ - My values: .743-.909 across factors - Limitation: Underestimates reliability when loadings vary

Composite Reliability (CR): - Also measures internal consistency - **Accounts for different factor loadings** (more accurate for CFA) - Range: 0-1; acceptable $\geq .70$ - My values: .752-.910 across factors - Preferred for SEM contexts

Average Variance Extracted (AVE): - Measures **variance captured by latent factor** vs. measurement error - Range: 0-1; acceptable $\geq .50$ - My values: .505-.772 across factors - Critical for convergent validity (factor explains majority of item variance)

Why all three? Each provides unique information: α for comparability with prior research, CR for accurate SEM assessment, AVE for convergent validity. Together, they provide comprehensive reliability evidence. All my factors exceed thresholds on all three metrics.”

Reference: §4.6.1 Reliability Assessment; Table 4.9 Reliability Summary

Your factor correlations range from .25 to .72. What does this tell us?

Answer: “The factor correlation range (.25 to .72) tells us several important things:

1. **Constructs are related but distinct:** Moderate correlations indicate factors measure related aspects of AI readiness without being redundant
2. **No multicollinearity concerns:** The highest correlation (.72) is well below the .85 threshold that would suggest constructs are not distinct (Kline, 2016)
3. **Discriminant validity supported:** When combined with HTMT analysis (all values < .85), this confirms factors represent genuinely different constructs
4. **Nomological coherence:** Theoretically related constructs (e.g., Hedonic Motivation and Performance Expectancy) show expected positive relationships

Specific patterns: - Strongest: HM ↔ EE (.72)—enjoyment relates to ease, as expected - Weakest: TR ↔ PV (.25)—trust and value are relatively independent - BI correlates moderately with predictors (.45-.65)—supporting criterion validity

The pattern supports both convergent validity (related constructs correlate) and discriminant validity (distinct constructs are separable).”

Reference: §4.6.2 Validity Assessment; Table 4.10 Factor Correlations

Walk me through exactly how you calculated the confidence intervals for your path coefficients.

Answer: “The confidence intervals were calculated using standard asymptotic methods in semopy:

Process: 1. **Estimate model:** Maximum likelihood estimation produces point estimates (β) and standard errors (SE) 2. **Standard errors:** Derived from the inverse of the information matrix (asymptotic covariance matrix) 3. **95% CI calculation:** $\beta \pm 1.96 \times SE$

Example with Price Value: - $\beta = .505$, $SE = .077$ - Lower bound: $.505 - (1.96 \times .077) = .354$ - Upper bound: $.505 + (1.96 \times .077) = .656$ - 95% CI: [.354, .656]

For marginal Trust effect: - $\beta = .106$, $SE = .057$ - 95% CI: [-.006, .218] - Note: CI includes zero, consistent with $p = .064$

Bootstrap alternative: For more robust estimates, bootstrap standard errors (1000+ resamples) could be used, which doesn’t assume normality. My primary analyses used asymptotic methods, which are appropriate for $N = 262$ with approximately normal data.”

Reference: §4.7.2 Structural Model Results; Table 4.12 Path Coefficients

What's the interpretation of a negative path coefficient like PE → BI ($\beta = -.028$)?

Answer: “The negative path coefficient for Performance Expectancy → Behavioral Intention ($\beta = -.028$, $p = .791$) requires careful interpretation:

What it means: - The coefficient is **essentially zero** (very close to 0) - The negative sign is **not meaningful** given non-significance - There is **no suppression effect** (PE doesn't become negative when controlling for other predictors) - PE simply **doesn't predict BI** in the presence of other variables

Why this matters theoretically: - In original UTAUT/UTAUT2, PE is typically the strongest predictor - For AI, traditional ‘usefulness’ perceptions don't drive adoption - This is a **null finding, not a negative finding**—PE isn't harmful, just irrelevant - Price Value ($\beta = .505$) absorbs the variance PE would typically explain

Statistical note: With $p = .791$, we cannot reject the null hypothesis that the true population coefficient is zero. The observed $-.028$ is within sampling error of zero. If we ran the study again, the estimate might be slightly positive or negative—the key finding is that it's not meaningfully different from zero.”

Reference: §5.4.2 Performance Expectancy Non-Significance; Table 4.12

How did you handle the non-normality detected by Mardia's test?

Answer: “Mardia's test detected significant multivariate non-normality in my data, which I addressed through robust estimation:

Detection: - Mardia's skewness: Significant ($p < .05$) - Mardia's kurtosis: Significant ($p < .05$) - Individual items showed mild to moderate non-normality

Solution: Robust Maximum Likelihood (MLM): - Used MLM estimation in semopy, which provides: - **Satorra-Bentler scaled chi-square:** Adjusts for non-normality - **Robust standard errors:** Corrects SE estimates - **Unbiased parameter estimates:** ML estimates remain consistent

Why MLM over alternatives: - More powerful than asymptotic distribution-free (ADF) methods - Doesn't require large samples like bootstrapping - Appropriate for mild-to-moderate non-normality - Widely recommended for psychometric data (Finney & DiStefano, 2013)

Verification: Sensitivity analyses comparing ML vs. MLM showed minimal differences in parameter estimates ($< .02$ change in standardized coefficients), indicating non-normality did not substantially bias results.”

Reference: §4.5.3 Estimation Methods; §4.4.2 Data Screening

11.2 Software and Reproducibility

What software did you use for analysis? Are your analyses reproducible?

Answer: “Yes, my analyses are fully reproducible. I used an open-source Python-based analysis pipeline:

Software stack: - **Python 3.11:** Core programming environment - **factor_analyzer:** Exploratory factor analysis (EFA) - **semopy:** Confirmatory factor analysis (CFA) and structural equation modeling (SEM) - **scipy/pingouin:** Statistical tests, correlations - **pandas/numpy:** Data manipulation - **matplotlib/seaborn:** Visualization - **Jupyter notebooks:** Documented analysis workflow

Reproducibility features: - **Random seed = 67:** Set for all stochastic processes - **Version control:** All notebooks tracked in Git - **requirements.txt:** Package versions documented - **Sequential notebooks:** 00-12 series covering complete pipeline - **Inline documentation:** Methods explained within code

Access: Notebooks available in the dissertation repository. Data available with appropriate institutional permissions. Any researcher with Python environment can reproduce all analyses exactly.”

Reference: §3.5 Analysis Software; Appendix: Analysis Notebooks

Why did you choose semopy over other SEM packages like lavaan or Mplus?

Answer: “I chose semopy for several practical and methodological reasons:

Primary reasons: 1. **Python integration:** Seamless connection with my data pipeline (pandas, numpy, factor_analyzer) 2. **Open source:** Free, transparent, reproducible without license constraints 3. **Adequate functionality:** Supports CFA, SEM, robust estimation—sufficient for scale development 4. **Active development:** Well-maintained with responsive developer community

Comparison with alternatives:

Feature	semopy	lavaan (R)	Mplus
Language	Python	R	Proprietary
Cost	Free	Free	\$895+
Robust SE	Yes	Yes	Yes
Bootstrap	Limited	Yes	Yes
Multi-group	Basic	Yes	Yes

Trade-offs acknowledged: - Mplus offers more advanced features (mixture models, complex missing data) - lavaan has longer track record and larger user base - For scale development, semopy's capabilities were sufficient

Validation: Key results were cross-validated with factor_analyzer for EFA, confirming consistency across tools."

Reference: §3.5 Analysis Software; §4.5.1 CFA Methodology

Could someone replicate your analyses with your code and data?

Answer: "Yes, complete replication is possible. Here's what's available:

Code availability: - 13 Jupyter notebooks (00-12 series) documenting complete analysis pipeline - Each notebook is self-contained with markdown explanations - All random seeds set (seed = 67) for exact reproducibility - requirements.txt specifies exact package versions

Data availability: - Clean dataset available (AIRS_clean.csv) - Item-level responses for all 523 participants - Demographic variables included - Data dictionary documenting all variables

Replication steps: 1. Clone repository 2. Create Python virtual environment 3. Install requirements: pip install -r requirements.txt 4. Run notebooks sequentially (00 → 12) 5. Results should match dissertation exactly

Permissions note: Survey data collected under IRB approval with participant consent for research use. Data sharing follows institutional guidelines. Researchers may need to request access through appropriate channels.

This level of transparency exceeds typical dissertation practice and reflects commitment to open science principles."

Reference: §3.6 Data Management; Appendix: Code Repository

12 Defense of Choices Questions

12.1 Methodological Choices

Why didn't you include a behavioral measure of actual AI tool adoption rather than just self-reported usage?

Answer: “This is a valid methodological concern that reflects practical constraints balanced against validation priorities:

Why self-reported usage: - **Survey methodology:** My design relied on Prolific panel data without access to participants’ actual software usage logs - **Privacy constraints:** Objective behavioral tracking raises ethical and practical challenges - **Cross-organizational sampling:** Participants came from diverse organizations with different AI tool ecosystems

Validation evidence despite limitation: - Self-reported usage **correlates strongly with intention** ($\rho = .69$) - This correlation provides **criterion validity** evidence for BI as meaningful outcome - Meta-analytic research shows intention-behavior correlations of .50-.60 for technology adoption (Venkatesh et al., 2012) - My .69 correlation actually **exceeds** typical meta-analytic findings

Future research recommendation: Organizational studies with objective usage metrics (login frequency, feature utilization, time-on-task) would strengthen behavioral validity. This is explicitly included in my research roadmap as a Phase 2 priority.”

Reference: §5.4.5 Behavioral Validation; §6.2.2 Self-Report Limitations; §6.3.2 Future Directions

Why didn’t you use an experimental design to establish causality?

Answer: “My cross-sectional design was appropriate for the primary research goal—scale development and validation—rather than causal inference:

Scale development priorities: 1. **Psychometric validation** requires diverse samples, not experimental control 2. **Factor structure** is assessed cross-sectionally 3. **Reliability and validity** don’t require temporal manipulation 4. **Nomological network** testing uses correlational evidence

Why not experimental: - **Timing:** Experiments test interventions; I first needed a validated instrument to measure outcomes - **Complexity:** Manipulating 8 psychological constructs isn’t feasible - **Ecological validity:** Cross-sectional surveys capture real-world variation better than lab manipulations

Causal claims I DO NOT make: - I describe **predictive relationships**, not causal effects - Path coefficients show **association strength**, not causation - The model is **correlational**, interpreted within theoretical framework

Future experimental work: Once AIRS is established, randomized controlled trials testing segment-specific interventions would establish causal evidence for adoption strategies. This is Phase 3 of my research roadmap.”

Reference: §3.2 Research Design; §6.2.1 Cross-Sectional Limitations; §6.3.2 Future Directions

You tested 12 hypotheses. Did you preregister your analysis plan?

Answer: “I did not formally preregister, though my analyses were hypothesis-driven based on established theory. Here’s the context:

Theory-driven approach: - All 12 hypotheses derived from **UTAUT2 framework** with AI-specific extensions - Hypotheses specified **before** data collection based on literature review - Analysis plan followed **standard psychometric procedures** (EFA → CFA → SEM) - No ‘fishing’ for significant results—I report all 12 hypotheses including 8 non-supported

Exploratory elements acknowledged: - User typology (cluster analysis) was exploratory - Some moderation analyses were data-driven - Post-hoc model comparisons (7- vs. 8-factor) informed recommendations

Why not preregistered: - Scale development involves iterative decisions (item reduction) that are difficult to fully prespecify - Preregistration is more critical for confirmatory hypothesis testing than instrument development - DBA programs typically don’t require preregistration

Recommendation for replications: Future studies testing AIRS predictions should preregister specific hypotheses, analysis plans, and decision rules. This strengthens confirmatory evidence.”

Reference: §3.3 Hypotheses Development; §4.5 Analysis Strategy; §6.3.2 Replication Recommendations

12.2 Theoretical Choices

Why extend UTAUT2 rather than develop a completely new AI adoption theory?

Answer: “Extending UTAUT2 was the more scientifically responsible choice for several reasons:

Building on established theory: - UTAUT2 has **extensive empirical support** across technology contexts - Meta-analyses confirm **robust predictive validity** ($R^2 = .40-.70$) - Provides **validated measurement items** as starting point - Enables **comparison** with prior technology adoption research

Parsimony principle: - Science prefers **extending existing frameworks** over proliferating new ones - AI adoption shares mechanisms with general technology acceptance - Differences can be documented through **differential hypothesis support** - My findings (PV dominance, PE non-significance) show what’s different without abandoning what works

Practical benefits: - Practitioners familiar with UTAUT can **immediately apply** findings - Organizational assessment tools can **build on AIRS** rather than starting fresh - Research community

can **integrate findings** with existing literature

The extension approach revealed important insights: By using UTAUT2 as baseline, I demonstrated that AI differs from general technology (Price Value dominates, Performance Expectancy doesn't predict). This finding would be invisible with a completely novel framework."

Reference: §2.3 Theoretical Framework; §6.1.1 Theoretical Contribution

Why include Behavioral Intention as an outcome rather than actual usage?

Answer: "Behavioral Intention is the standard and appropriate outcome for technology acceptance research:

Theoretical justification: - BI is the **proximal determinant** of behavior in all reasoned action theories (TRA, TPB, TAM, UTAUT) - Meta-analyses show **consistent intention-behavior relationships** ($r = .50-.60$) - BI captures **psychological readiness** which is my construct of interest - Actual usage is influenced by many factors beyond individual readiness (access, opportunity, organizational requirements)

Measurement advantages: - BI can be assessed **consistently across contexts** (different organizations, tools, industries) - Usage measures are **context-dependent** and hard to standardize - BI predicts **future behavior**, making it actionable for interventions

Validation evidence provided: - Strong BI-Usage correlation ($\rho = .69$) confirms **criterion validity** - This correlation **exceeds** typical meta-analytic findings - Supports BI as meaningful proxy for behavioral outcomes

My position: BI is not a limitation—it's the appropriate outcome for understanding psychological readiness. Usage would add value but requires organizational access and standardized measurement approaches not feasible in my design."

Reference: §2.3.2 Outcome Variable Selection; §5.4.5 Behavioral Validation; §6.2.2 Outcome Limitations

Why didn't you include usage as a moderator throughout the model, as in original UTAUT?

Answer: "The original UTAUT included experience/usage as a moderator of multiple paths, but I tested limited moderation for practical reasons:

Complexity constraints: - Full UTAUT moderation involves **12+ interaction terms** across age, gender, experience, voluntariness - Each moderation test requires **adequate subgroup sample sizes** - With $N = 262$ for CFA/SEM, full moderation framework would be **underpowered** - Rule of thumb: ~20 observations per estimated parameter; full moderation would exceed capacity

What I did test: - Usage moderation of key relationships (HM → BI strengthens with experience)
- Subgroup comparisons where sample sizes permitted - **Correlation patterns** across usage levels

Findings from limited moderation: - HM → BI moderated by usage (significant interaction) - Suggests experienced users rely more on enjoyment - Full moderation framework would likely reveal more nuanced patterns

Future research direction: Studies with $N > 600$ could test comprehensive moderation, examining whether AI adoption follows UTAUT2's age, gender, and experience moderation patterns or shows AI-specific differences. This is noted as a Phase 2 research priority.”

Reference: §5.4.4 Moderation Results; §6.2.3 Sample Size Limitations; §6.3.2 Future Directions

12.3 Practical Choices

Why recommend the 8-factor model when the 7-factor model (without Trust) has better fit indices?

Answer: “This decision reflects prioritizing **practical utility** over pure statistical parsimony:

Fit comparison: - 7-factor model (no Trust): CFI = .978, RMSEA = .062 - 8-factor model (with Trust): CFI = .975, RMSEA = .065 - Difference is **marginal** ($\Delta\text{CFI} = .003$)

Why retain Trust despite marginal significance:

1. **Diagnostic value:** Organizations implementing AI need to assess trust; a 7-factor AIRS without trust leaves a critical blind spot
2. **Theoretical importance:** Trust is extensively discussed in AI literature; omitting it ignores substantial research tradition
3. **Marginal effect suggests relevance:** $\beta = .106$, $p = .064$ approaches significance; with larger sample, effect likely emerges
4. **Future AI contexts:** As AI becomes more autonomous and consequential, trust will likely become more important
5. **Practical utility:** Better to have a Trust measure that shows marginal effects than no Trust measure at all

My recommendation is nuanced: - For **research** (maximum parsimony): 7-factor model is defensible - For **practice** (comprehensive assessment): 8-factor model preferred - AIRS should be **adaptable** based on use case

This reflects the applied (DBA) nature of my research—practical utility sometimes trumps statistical elegance.”

Reference: §5.3.3 Model Comparison; §6.1.2 Trust Discussion; §6.4.1 Practical Recommendations

Why didn’t you develop formal AIRS scoring algorithms as part of this dissertation?

Answer: “Scoring algorithm development was beyond dissertation scope but is planned for future work:

Why not included:

1. **Scope constraints:** Dissertation focused on psychometric validation—establishing that AIRS measures what it claims to measure reliably and validly
2. **Normative data requirements:** Formal scoring (T-scores, percentiles) requires:
 - Larger normative samples
 - Demographic stratification
 - Cross-validation across populations
 - Clinical cut-point validation
3. **Developmental sequence:** You must validate the instrument **before** developing scores; putting the cart before the horse risks scores based on flawed measurement
4. **Practical complexity:** Scoring algorithms need additional research on:
 - How to weight factors
 - Whether total score is meaningful
 - Cut-points for ‘low/medium/high’ readiness
 - Segment classification algorithms

What IS included: - Factor-level mean score interpretation - Relative standing guidance (compare to sample means) - Cluster profiles for segment identification - Qualitative interpretation framework

Research roadmap: - **Phase 1** (1-2 years): Develop normative data and standardized scoring - **Phase 2:** Validate scoring against outcomes - **Phase 3:** Create organizational diagnostic protocols

This is a **foundation** dissertation; scoring is the next step in the research program.”

Reference: §6.4.3 Research Roadmap; §6.3.2 Scoring Development; §5.5 Interpretation Guidelines

13 Preparation Tips

13.1 Before the Defense

Re-read your entire dissertation the week before **Practice explaining** key concepts without jargon **Prepare a 15-20 minute summary** presentation **Review your statistical outputs** in detail **Anticipate your committee’s specific interests** based on their expertise **Prepare visual aids** for complex statistical concepts **Have your data and analyses accessible** for ad-hoc queries

13.2 During the Defense

Take your time with complex questions—thinking is expected **It’s okay to say “I don’t know”** or “that’s a limitation” **Acknowledge valid criticisms** rather than being defensive **Connect answers to your larger contribution** **Ask for clarification** if a question is unclear **Stay positive**—you know your study better than anyone

13.3 Key Numbers to Memorize

Metric	Value
Total N	523
EFA sample	261
CFA sample	262
Final items	16
Final factors	8
CFI	.975
TLI	.960
RMSEA	.065
R ²	.852
PV β	.505
HM β	.217
SI β	.136
TR β	.106 (p = .064)
Reliability range	.743-.909
BI-Usage correlation	ρ = .69
User segments	4 (16%, 30%, 37%, 17%)

Table 1: Key Numbers to Memorize. *Source: Compiled by Author*

14 Quick Reference: Potential “Gotcha” Questions

“Your R^2 is .85—that’s impossibly high” → Consistent with UTAUT meta-analyses; latent variable modeling; strong theoretical model

“Most hypotheses weren’t supported” → That’s the finding—AI is different; drives theoretical contribution

“Trust wasn’t significant” → Marginal ($p = .064$); power limitation; retained for diagnostic value

“Your dropped constructs are a failure” → Empirical finding about measurement complexity; informs future research

“Cross-sectional can’t show causation” → Acknowledged; behavioral validation ($p = .69$); longitudinal recommended

“Western sample limits generalizability” → Acknowledged; cross-cultural research needed; U.S. findings still valuable

“Why should we trust self-reported intention?” → Standard in field; meta-analytic support; usage correlation validates

Good luck with your defense!