Beyond the Algorithm:

Comparing Human and AI Systems Analysis of Outcomes of an Engineering Summer Bridge Program

Caitlin Morris · Kavitha Chandra · Susan Thomson Tripathy · Orlando Arias · Elizabeth Often

University of Massachusetts Lowell

LEAD Division | 2026 ASEE Annual Conference | Charlotte, NC | June 21–24, 2026

Acknowledgements

With Gratitude

To the people and program that made this research possible

- RAMP students & industry mentors, for their time, candor, & reflections
- UMass Lowell & the RAMP program team, for institutional & programmatic support
- Colleagues in engineering education & PAR scholarship whose frameworks anchor this work

AI Disclosure

This presentation was developed with the assistance of Claude (Anthropic) for slide design, layout, and figure preparation

All analysis, interpretation & conclusions are the authors' own; AI use was limited to presentation formatting

Note: This mirrors the transparency this study itself calls for

A Preview..

This deck's own production—AI-assisted, human-reviewed—mirrors the hybrid workflow this study recommends for engineering education research.

Research Context & Motivation

The RAMP Program

Research, Academics & Mentoring Pathways to Success at UMass Lowell

- Engineering summer bridge for rising high-school seniors & new college students
- Robotics projects, coding instruction, industry mentorship
- Builds engineering identity, belonging, & leadership
- PAR-based assessment framework

Why This Matters

AI is increasingly used for data processing, coding support, and preliminary interpretation of program evaluation data

Yet engineering education constructs—identity, belonging, mentorship—are deeply contextual and theory-dependent

Gap: No systematic comparison of AI vs. human interpretation for these constructs

Research Question

How do AI systems interpret mentorship and engineering identity data differently from humans, and how does prompt structure influence this interpretation?

Theoretical Background

Engineering Identity & Belonging

- Predicts persistence, engagement & purpose (Godwin, 2016)
- Shaped by recognition, performance opportunities & internalized values (Carlone & Johnson, 2007)
- Belonging supports resilience—especially for students from historically excluded groups (Strayhorn, 2018)

Mentorship Mechanisms

- Industry mentors provide authentic career exposure
- Near-peer mentors scaffold tasks via cognitive apprenticeship
- Both forms support professional identity & leadership orientation formation (Rottmann et al., 2015)
- PAR surfaces psychosocial dimensions of mentoring (Tripathy et al., 2020)

AI Limitations in Education

- Lacks developmental nuance
- May confuse correlation with mechanism
- Misinterprets ceiling effects & small-N datasets
- Ignores qualitative meaning-making
- May overgeneralize from limited evidence (Holmes et al., 2019)

Mentorship as Leadership Development: The LEAD Connection

Engineering Leadership Orientations (ELOs)

- Professional identity configurations shaping how engineers understand authority, collaboration & influence (Rottmann, Sacks & Reeve, 2015)
- Emerge through socialization & project-based learning—not formal leadership instruction alone
- Mentoring relationships expose students to authentic professional practice & responsibility-sharing

Troost iLead Research Program

- Leadership identity develops via experiential learning & relational validation (Reeve et al., 2015)
- Team-based problem solving builds leadership capacity (Chan & Rottmann, 2018)
- Bridge programs function as an early site of leadership socialization
- Shapes students' self-view as capable contributors & emerging leaders

Why This Matters for LEAD

- Industry mentoring & collaborative problem-solving in RAMP mirror experiential leadership-development models
- If mentorship builds identity, it may simultaneously build leadership competency & orientation
- AI causal overclaiming carries elevated stakes when it shapes how leadership outcomes get assessed
- Bridge-program mentoring is an early site of leadership socialization

Methods: Three-Lens Interpretive Framework

01

Non-Contextualized

Structural Auditor

AI restricted to numeric descriptions, distributions, missingness, scale ranges. No causal, developmental, or theoretical language permitted.

02

Contextualized

Theory-Informed Interpreter

AI instructed to consider engineering identity & belonging frameworks, mentorship mechanisms, developmental stage, and qualitative themes.

03

Adjudicator

Synthesis & Critique

AI reviews both its prior outputs: compares lenses, flags blind spots and overreach risks, generates an integrated evidence-bound interpretation.

5 RAMP 2025 Datasets (Numeric, Ordinal & Qualitative)

Student pre-survey (n=20)
Student post-survey (n=19)
Mentor pre-survey (n=6)
Mentor post-survey (n=5)
Mentor poll / preference ranking (n=20)

3 AI Models Evaluated

- **Perplexity Sonar** — Retrieval-Augmented Generation
- **ChatGPT 5.1** — General-purpose LLM
- **Claude 3.5 Sonnet** — Narrative-oriented LLM

Figure 1: Interpretation as an Engineered Process

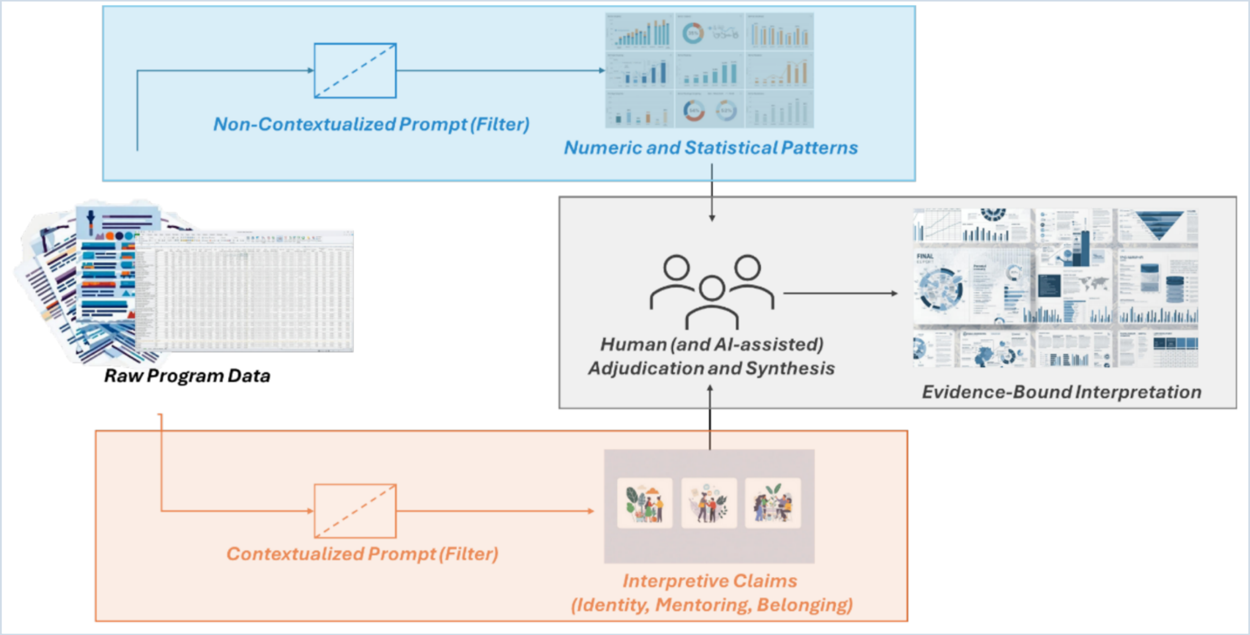

Figure 1 (paper): P&ID-style hybrid human–AI interpretive workflow. The same survey data pass through two engineered prompt filters before human adjudication.

How to Read This Figure

- Same raw program data flows through two engineered prompt conditions
- **Blue path** restricts AI to numeric & structural patterns only
- **Orange path** permits theory-informed claims about identity, mentoring & belonging
- Both outputs converge in human (+AI-assisted) adjudication → one evidence-bound interpretation

Results: Non-Contextualized Outputs

Perplexity Sonar

Data Auditor

Overreach: Low

Strictly structural — distributions, missingness, scale ranges, duration outliers. No interpretation attempted.

> *"Identity items are right-skewed toward agreement at both timepoints."*

ChatGPT 5.1

Statistical Reviewer

Overreach: Low–Mod

Statistical audit: chi-square comparisons, effect sizes, significance thresholds. Cautious, constrained phrasing throughout.

> *"No items reached conventional significance at $\alpha = 0.05$; small sample sizes constrain inference."*

Claude 3.5 Sonnet

Interpretive Drift ⚠

Overreach: High

Despite explicit restrictions, drifted into causal and evaluative language. Treated descriptive changes as evidence of effectiveness.

> *"Coding confidence improved significantly, demonstrating the program's effectiveness."*

Results: Contextualized Outputs

Perplexity Sonar

Overreach: Low–Mod

Conservative theory integration. Grounded in data: program appeared to consolidate an existing engineering identity rather than create new interest—consistent with Carlone & Johnson (2007).

"Students may be consolidating identity rather than forming it."

ChatGPT 5.1

Overreach: Moderate

Strong conceptual synthesis. Noted ceiling effects; raised plausible mechanisms (cognitive apprenticeship, professional socialization) while acknowledging design and power limitations.

"Belonging outcomes were mixed; mentorship supported perceived growth; developmental stage mattered."

Claude 3.5 Sonnet

Overreach: High

Rich narrative—but frequently exceeded the evidence. Claimed causal and durable identity shifts not measurable in survey data. Likely drawing on learned patterns from prior literature.

"Mentoring transformed students' perceptions of engineering."

Figure 2: Prompt Engineering as Interpretive Filtering

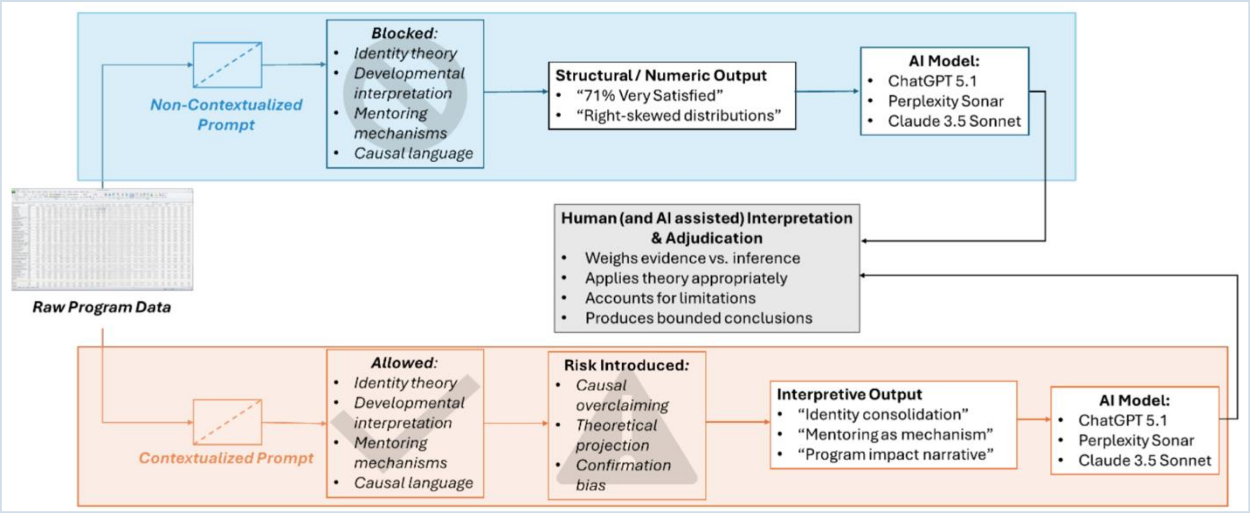

Figure 2 (paper): The same survey & qualitative data pass through two engineered prompt conditions; a human adjudication step integrates both into evidence-bound conclusions.

Where the Risk Comes In

- **Blocked (blue):** identity theory, developmental interpretation, mentoring mechanisms & causal language
- **Allowed (orange):** the same, but with risk of causal overclaiming, theoretical projection & confirmation bias
- Human adjudication weighs evidence vs. inference & produces bounded conclusions

Where Human Interpretation Diverged from AI

AI Models	Human Researcher (Sociologist, PAR expert)
Ceiling Effects ChatGPT & Claude inferred Likert increases near max as evidence of identity growth.	Recognized ceiling as self-selection bias — students entered RAMP already highly motivated. Likert gains near maximum are not meaningful growth indicators.
Developmental Differences AI inconsistently noticed HS vs. 1st-year differences; tended to collapse into one narrative.	Clearly distinguished exploration (HS: 'learning what engineering is') from consolidation (college: 'building networks, strengthening major prep') aligned with identity theory.
Mentor Perceptions Claude: treated mentor reports of student engagement as direct causal evidence of program impact.	Treated mentor perceptions as valuable but subjective accounts; adjusted for n≈5, social desirability bias, and absence of longitudinal tracking.

Discussion: Implications for Engineering Education Research

Key Interpretive Risks

Causal overclaiming
Descriptive pre–post differences stated as program effectiveness.

Theoretical projection
Identity theory applied beyond what survey instruments can measure.

Confirmation bias
Positive signals treated as objective evidence of program impact.

Small-N overconfidence
Patterns in n≈5 mentor data treated as robust findings.

Recommended Guardrails

1. Label every claim as descriptive vs. interpretive
2. Require evidence trace for each interpretive claim
3. Prohibit causal language without appropriate research design
4. Enforce limitations-first reporting structure
5. Require human adjudication aligned with established theory

AI = accelerant & hypothesis generator, not interpreter — human expertise and theory must remain central.

Conclusions & Future Directions

Key Takeaways

1 AI models have distinct interpretive identities that persist across prompt conditions

2 Contextualization makes AI more useful AND more dangerous

3 Numeric-only AI analysis is insufficient for identity, belonging, and mentorship constructs

4 Human interpretation is essential—not optional—for engineering education research

5 Transparency in reporting AI model, prompt, and edits is a research integrity requirement

Future Work

→ Open-source & domain-tuned educational AI models

→ Longitudinal tracking of interpretive drift across model versions

→ Equity implications: race, gender, first-gen intersectionality

→ Structured hybrid human–AI workflow frameworks

→ AI for qualitative first-cycle coding at scale

→ Larger, more diverse bridge program datasets

Caitlin Morris | caitlin_morris@uml.edu | UMass Lowell | Paper #52442 | Questions welcome!