

## CISE HCC CAREER: Towards Ethical Emotion AI Futures

Noura Howell, PhD, Assistant Professor, Georgia Institute of Technology, nhowell8@gatech.edu

### Overview

This proposal will apply methods of speculative design and STEAM education to advance more ethical Emotion AI futures. Emotion AI is an emergent, controversial set of techniques that (claim to) classify human psychological states. Emotion AI is proposed, deployed, and sometimes withdrawn after criticism, in high stakes contexts such as education, hiring, or work. Computer scientists, psychologists, and policy experts debate potential benefits and harms, arguing for polarized futures ranging from panoptic expansion to complete bans. Proponents expound benefits for well-being and security, while critics decry privacy and civil liberties risks, cultural bias, and shaky scientific foundations. Undergirding these polarized debates seem to be competing imagined futures of what Emotion AI can/not and should/not do. Without articulating these envisioned futures for Emotion AI, hasty Emotion AI deployments may cause harm, garner criticism, and ultimately hamper AI innovation. Articulating more diverse, beneficial, and ethical futures for Emotion AI can help Emotion AI deliver on its promise for societal benefit and advance responsible AI innovation.

My long-term goal is to articulate futures from diverse stakeholders, to shape inclusive, beneficial AI innovation. Amidst polarized debate, I side with neither extreme; rather, I work to surface diverse perspectives that may guide development or advocate for curtailment of Emotion AI. My approach applies existing best practices from STEAM education with novel adaptations of speculative design. Speculative design's intent is to foster societal debate and ethical reflection to identify desirable, beneficial directions for technology development; yet, too often speculative design risks foregrounding the voices of privileged societal positions with easy access to design and art spaces. This proposal builds on my past methodological advancements and adds novel methodological advancements for more participatory, diverse speculative design. Through this, by articulating imagined futures of Emotion AI from diverse perspectives, AI innovation can be more responsive to societal needs.

### **1. Objective: Develop an Emotion AI programming Toolkit for new coders.**

RQ: How can Emotion AI be made explainable and accessible to new coders for tinkering and prototyping? How can tinkering with Emotion AI support interdisciplinary STEM education?

Method: Develop a graphical programming interface for an existing Emotion AI API. With community partners, host workshops with (Group 1) socioeconomically marginalized teens and young adults introducing programming with the Emotion AI Toolkit. Refine the toolkit based on results.

Evaluation: Partner with university educational evaluation group to measure student learning of CS concepts and sense of belonging in STEM.

Output: Free open-source Emotion AI Toolkit on github, free online lesson plans, tutorials on Instructables and Youtube. With community partners, we will also host trainings for educators to adapt the workshops into their own teaching.

### **2. Objective: Articulate Emotion AI futures with (Group 1) socioeconomically marginalized teens and young adults.**

RQ: What future Emotion AI analyses, applications, and ways of living with Emotion AI does (Group 1) imagine? What do they consider desirable or undesirable, ethical or unethical,

about these futures? With Emotion AI, what do they want or not want for themselves and others?

Method: The workshops with (Group 1) described in Objective 1 will include speculative design activities to imagine many alternative futures with Emotion AI and reflect on benefits and risks of these imagined futures. Participants will represent these imagined futures with prototypes and vignettes. Participants will be asked to find ways to ‘trick’ the Emotion AI into making errors, and suggest guidelines for future Emotion AI development.

Evaluation: Partner with university educational evaluation group to measure student learning around creativity and ethical reflection, and sense of belonging in STEM. Also, qualitative thematic analysis of participants’ discussions and imagined futures (prototypes and vignettes) addressing the RQs.

Output: Taxonomy of alternative futures for Emotion AI, including perceived benefits and risks and desirability or undesirability.

### **3. Objective: Articulate Emotion AI futures with (Group 2) technology and AI workers.**

RQ: What future Emotion AI analyses, applications, and ways of living with Emotion AI does (Group 2) imagine? What do they consider desirable or undesirable, ethical or unethical, about these futures? With Emotion AI, what do they want or not want for themselves and others?

Method: Host workshops with (Group 2) using an adapted version of the workshops described above.

Evaluation: Qualitative thematic analysis of participants’ discussions and imagined futures (prototypes and vignettes) addressing the RQs.

Output: Taxonomy of alternative futures for Emotion AI, including perceived benefits and risks and desirability or undesirability.

### **4. Objective: Identify failure modes of Emotion AI.**

RQ: What are common Emotion AI errors? What are potential consequences of these errors in different applications?

Method: Qualitative coding of the workshops with (Group 1) and (Group 2) focused on failure modes and consequences. Literature review of algorithmic audits of Emotion AI. Informal algorithmic audits with university students.

Output: Taxonomy of failure modes of Emotion AI. This can inform the direction of future algorithmic audits, and enable users of Emotion AI systems to make more informed decisions that consider the limitations of Emotion AI.

### **5. Objective: Develop design recommendations for Emotion AI.**

RQ: What are design considerations, especially those unique to Emotion AI, can help guide Emotion AI to be more ethical and inclusive?

Method: Qualitative coding of the workshops with (Group 1) and (Group 2) focused on design recommendations. Literature review of existing AI ethics guidelines.

Output: Design recommendations for Emotion AI. This can be useful to technology workers designing Emotion AI applications.

### **Intellectual Merit**

Related work studies how a general public imagines Emotion AI, shows that Emotion AI demands unique considerations, and traces imagined futures as pivotal in the development of Emotion AI. This proposal addresses a gap of studying imagined futures of Emotion AI with (Group 1) socioeconomically marginalized teens and young adults and (Group 2) technology workers in computer and design roles, with an explicit emphasis on potential harms and failure modes of Emotion AI. Recent advances in design research methods seek to make speculative design and related approaches more diverse, inclusive, and participatory. This proposal builds on this past work to advance a workshops method for more participatory speculative design.

### **Broader Impacts**

This project addresses an urgent moral imperative to better understand how people want or do not want to live alongside Emotion AI. The perspectives of (Group 1) and (Group 2) are essential: (Group 1) Socioeconomically marginalized teens and young adults navigate biased, opaque AI systems with increasing Emotion AI surveillance in work, education, and hiring. The Emotion AI Toolkit and speculative design workshops will offer enhanced AI and STEM literacy for (Group 1) socioeconomically marginalized teens and young adults, and encourage them to recognize the importance of their perspectives in shaping the future of technology development. (Group 2) technology workers struggle to advocate for ethics in the products they build, yet are uniquely well positioned to directly impact future Emotion AI systems. The workshops will also offer a much-needed space for (Group 2) tech workers to engage in ethical reflections offsite from their employer and the constraints of their direct professional responsibilities. I have been studying the tech industry's engagement with Emotion AI for about a decade through design futuring methods that seek to rework dominant techno-narratives and articulate diverse alternative futures. I also partner with community groups on STEM outreach. Understanding how (Group 1) and (Group 2) imagine Emotion AI futures is key to shaping technology development and curriculum for more beneficial Emotion AI futures.