



Member's-Only Content: Please Read Before Proceeding .

The case studies and other materials in this section are confidential and for members of the ForGood Framework community only. For Discussion, Not Distribution: These materials are intended to facilitate discussion and are not for public use. Do Not Cite or Share: Please do not cite, share, or distribute these materials outside of this member's section.

Cyborgs, Centaurs & Self-Automators: 3 Modes of Human-GenAI Knowledge Work and Implications for Skilling

Randazzo et al. (2025)

Context

As generative AI enters professional problem-solving, attention is shifting from whether it is used to how. In a field study of 244 BCG consultants analysing 4,975 human-AI interactions, the authors identify three distinct modes of collaboration depending on who sets the task and who executes it: Fused Co-Creation (Cyborgs, 60% of participants), Directed Co-Creation (Centaurs, 14%), and Abdicated Co-Creation (Self-Automators, 26%). Each produces meaningfully different outcomes for accuracy, skill development, and the future of professional expertise.

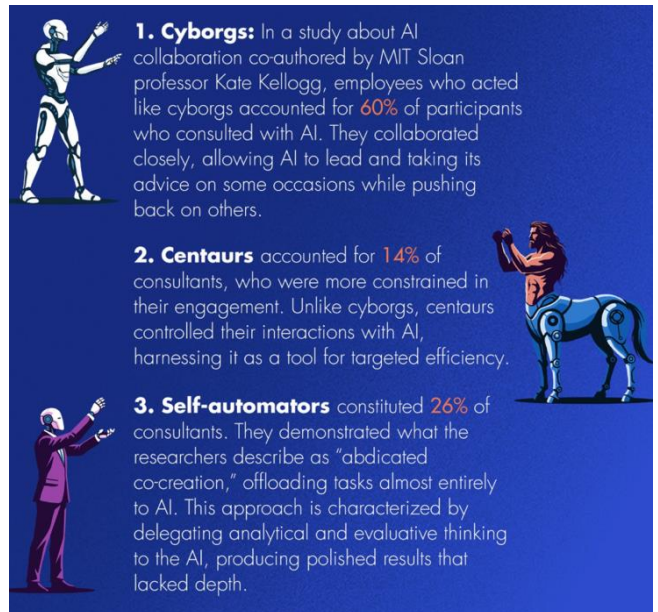
Key Insights

Engagement intensity \neq accuracy

- **Centaurs** scored highest on recommendation accuracy despite using AI least
- **Cyborgs** were persuaded by wrong AI output even while applying validation best practice
- **Self-automators** scored lowest on both accuracy and persuasiveness

Skilling outcomes diverge sharply by mode

- **Cyborgs newskill**: build AI-related capability while maintaining domain expertise (fused co-creation)
- **Centaurs upskill**: deepen task-related expertise but gain limited AI fluency (directed co-creation)
- **Self-automators neither**: no growth in domain or AI skills despite producing output (abdicated co-creation)



Graphic from a MIT Instagram post on the study see [here](#)

FORGOOD check (authors' interpretation, not part of the original source)

- **Fairness & Openness** (*transparency & differential impact*): How might AI co-creation affect different groups at scale? Where is AI co-creation happening without clear disclosure of who shaped the output?
- **Respect & Goals** (*judgement & long-term capability*): Where might heavy AI iteration leave team judgement over-dependent on AI output your team cannot independently verify? What is your team's default mode costing in domain expertise or AI fluency?
- **Opinions, Options & Delegation** (*scrutiny, mode choice, accountability*): How deliberately are teams choosing between modes based on task stakes? Would those choices hold up to external scrutiny? When AI-assisted outputs cause harm or error, is accountability clear to everyone involved before the work begins?