



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 Do Not Cite or Share: Please do not cite, share, or distribute these materials outside of this member's section.

How behavioral science can improve the return on AI investments

De Cremer et al. (2025)

Context

Many AI projects fail because leaders default to technosolutionism: the belief that better technology alone will solve organisational challenges. People, however, tend to resist tools that disrupt routines, overreact to visible AI errors, and prefer familiar human judgment, which is why even well-designed systems fail to gain traction. Leaders can address this by applying the “Behavioural Human-Centred AI” approach across 3 stages for higher trust, ROI and faster uptake.

Key Insights

Three Stages to Behavioural Human-Centred AI

Design: Build for cognitive shortcuts

- Co-design with diverse end-users to iterate based on their real needs, increase ownership & adoption
- Add purposeful friction (e.g., harder-to-read fonts) to improve scrutiny of AI outputs
- Require beta-testing to catch bias before launch against inventor's bias (the tendency to overlook unintended consequences of one's own system)

Adoption: Address trust, effort, and perceived control

- Frame AI as augments: highlight how it frees employees for higher-value work, driving innovation
- Counter availability heuristic (here the fixation on rare AI failures which are more salient) by making AI mistakes relatable: position AI as a learning partner that errs like humans do, rather than infallible
- Proactively disclose limitations and safeguards to reduce anxiety and build willingness to engage
- Use explainable AI (XAI) features to demystify AI reasoning and increase perceived control

Management: Address leadership biases and train executives in change management

- Executives should acknowledge their own biases
 - Overconfidence → Underestimating behavioural complexity, skipping pilots assuming employees will “figure it out”
 - Escalation of commitment → Doubling down on failing tools rather than cutting losses
- Make “leading change” a core skill: identifying & addressing resistance, communicating transparently, inviting regular feedback, modelling AI adoption (e.g. demonstrating LLM use to normalise it)
- Track people-centric KPIs: run regular temperature checks measuring whether employees trust the AI, perceive it as fair, believe colleagues are using and like it, as these are early leading indicators of failure

Implications

To what extent do you have these behavioural change management capabilities for higher AI-ROI:

- Executive training in bias awareness and clear exit criteria for underperforming initiatives
- Regular temperature checks on whether AI tools are meeting employees’ and firm’s actual goals, respecting their autonomy in how tools are used (within clear, communicated boundaries)
- Applied behavioural insights in design (e.g. friction, XAI) including diverse user testing and meaningful human override options to ensure fairness