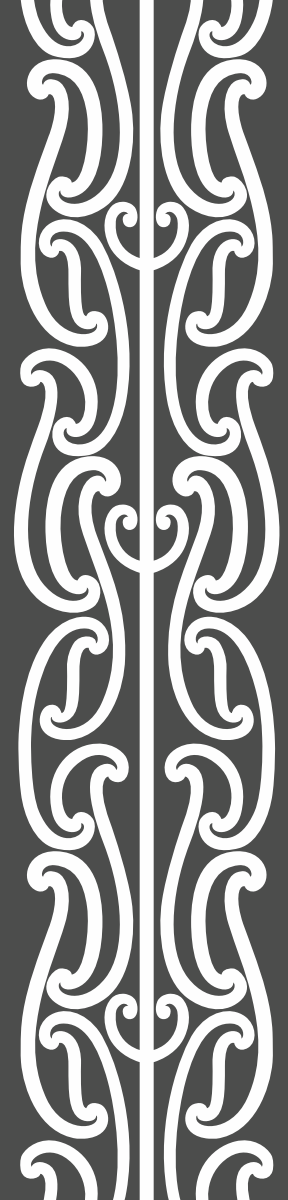
Human-centred design of AI – ecosystems for wellbeing

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Classification: In-Confidence

Tēnā koutou katoa,	**Greetings to you all**
E te mana whenua o tēnei rohe, tēnā koutou	**To the people of this land, greetings to you all**
Nō Kotirana ōku tīpuna	**My ancestors come from Scotland**
Ko Megan tōku ingoa	**My name is Megan**
He kaiako ahau i te hinengaro o te mahi	**I am a lecturer in organisational psychology**
Tēnā koutou, tēnā koutou, tēnā koutou	**Greetings, greetings, greetings to us all**



"Technology is neither good nor bad; nor is it neutral."
(Krantzberg's First Law, 1986)

Classification: In-Confidence

What is wellbeing?

Assumptions:

Well-being is a positive state connected to the experience of emotions and the cognitive assessment of our lives (Deci and Ryan 2008).

Hedonic - understood as pleasant

life (pleasure, enjoyment, comfort),

Eudemonic – valuable life (growth, authenticity, meaning)

AI is not new – Human computer interaction 50's to 70's

What is human-centred design?
The goal is not to make technology more impressive. The goal is to make outcomes better for people.

Wellbeing ecosystems connect to sociotechnical systems theory of AI

Automation can create stress

- witnessing superiority of robots
- pessimistic societal rhetoric = Stress appraisal – a threat that cannot be coped with

- People who witnessed more robots in their workplaces reported more burnout, more job insecurity and more workplace incivility across a variety of contexts

(Yam et al., 2023)

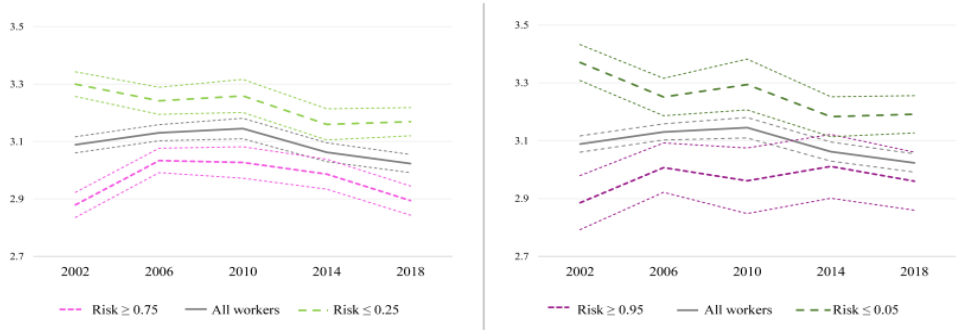
Mixed results on wellbeing - Nazareno & Schiff (2021)

Country-level data

Assessed 402 occupations based on the General Social Survey (2002-2018).

Calculated a % of automatization risk index, utilizing Frey & Osborne's (2017) methodology.

Predicted stress levels by year and risk level



A1. Above 75% and below 25% risk.

A2. Above 95% and below 5% risk.

Fig. 3. Predicted stress levels by year and risk level.
Notes: 1. Predicted values and confidence intervals from Diff-in-Diff model reported in Table 3, column 6. 2. All models control for the standard covariates and factor variables. 3. Results for job satisfaction and health are available in Fig. B1 in the appendix.

Interesting: Mixed Wellbeing effects!

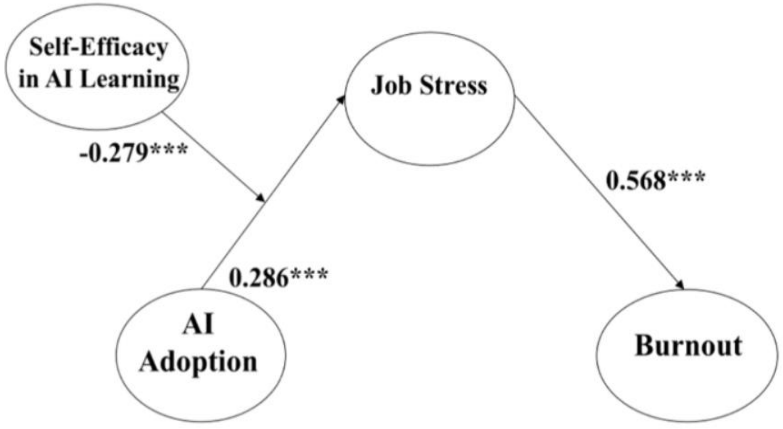
- Workers with high automation risk experience less stress.
- However, they report worse health and mixed/negative job satisfaction.

Impact on wellbeing

Plenty of research suggests that AI decreases wellbeing at work, specifically through

- Loss of meaning
- Mattering
- Job insecurity
- Increased vigilance
- Technostress

Fig. 3



Coefficient Values of Our Research Model (***$p < .001$. All values are standardized).

Pace and pervasiveness of change

- How radical is the functional change? Does it shift the norm of what 'doing a good job' means?
- Does it affect multiple tasks simultaneously? Are not just tasks but also roles, hierarchies, job groups affected? How has the change be implemented?

Human-centred work design needs to stand central (Parker & Grote, 2022)

- Job autonomy – both decision making and where and when we work
- Skill variety
- Job feedback
- Social and relational
- Job demands

Parker and Grote – work design

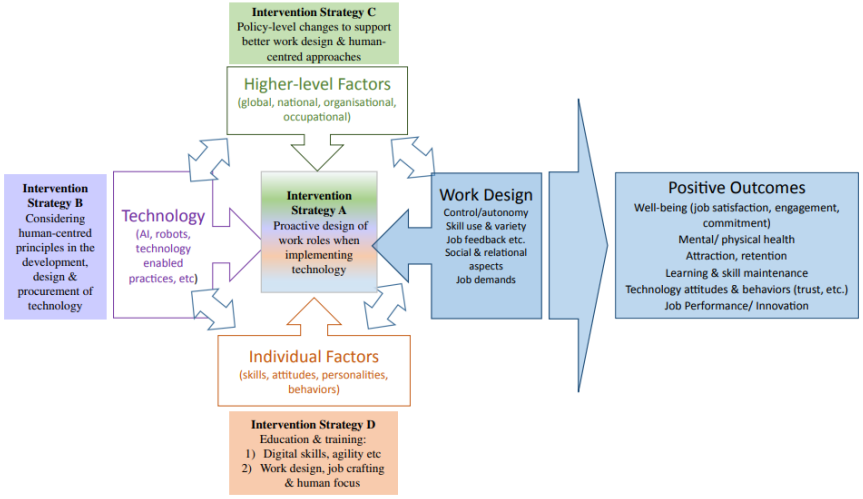


FIGURE 1 Work design as key to achieving benefits of digital technologies

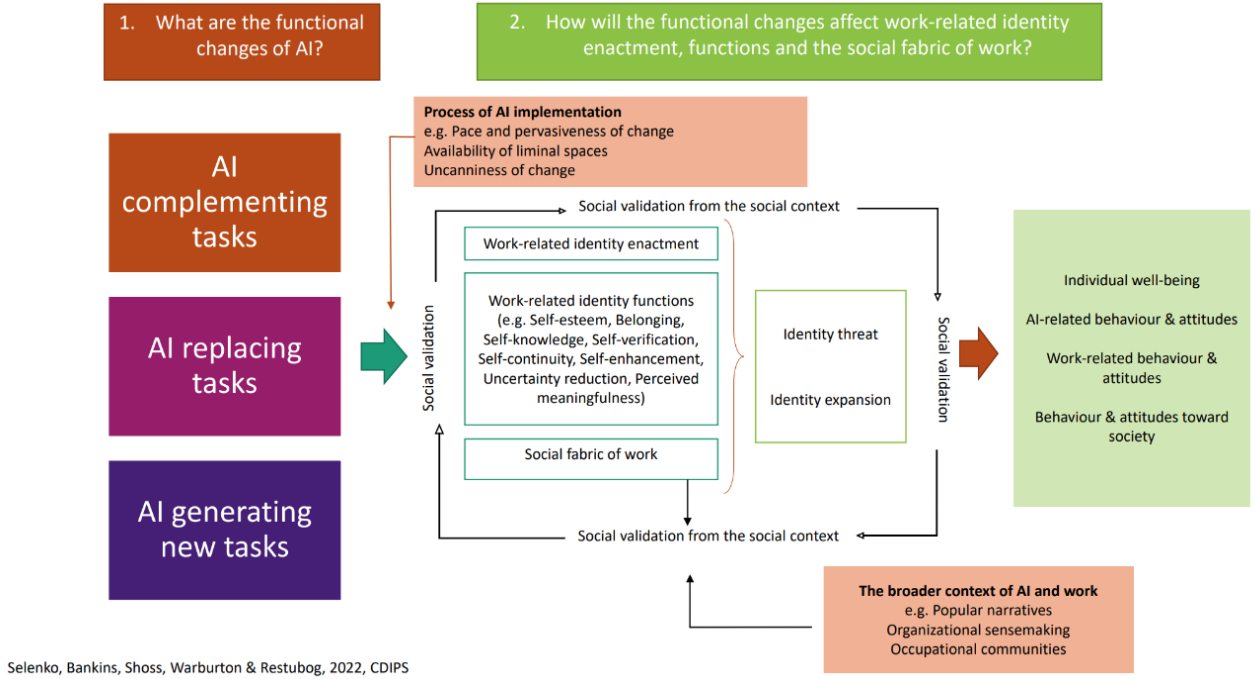
Social validation of the context – work identity

- Does it change worker communities, teams, professional bodies that provided identity functions guidance, meaning, belonging, appreciation?
- Does it change status structures within organisations, within projects?
- How welcoming are others in one's team, organisation, profession, toward the change – tech utopians vs doomsday-ers?

- Frequent interaction with AI systems leads to affiliation but also feelings of loneliness (Tang et al., 2023)

- Individual-level factors buffer the effect of AI experience on well-being outcomes: Self-affirmation strategies (Yam et al., 2023), attachment anxiety (Tang et al., 2023).

Identity, behaviour and well-being



Selenko, Bankins, Shoss, Warburton & Restubog, 2022, CDIPS

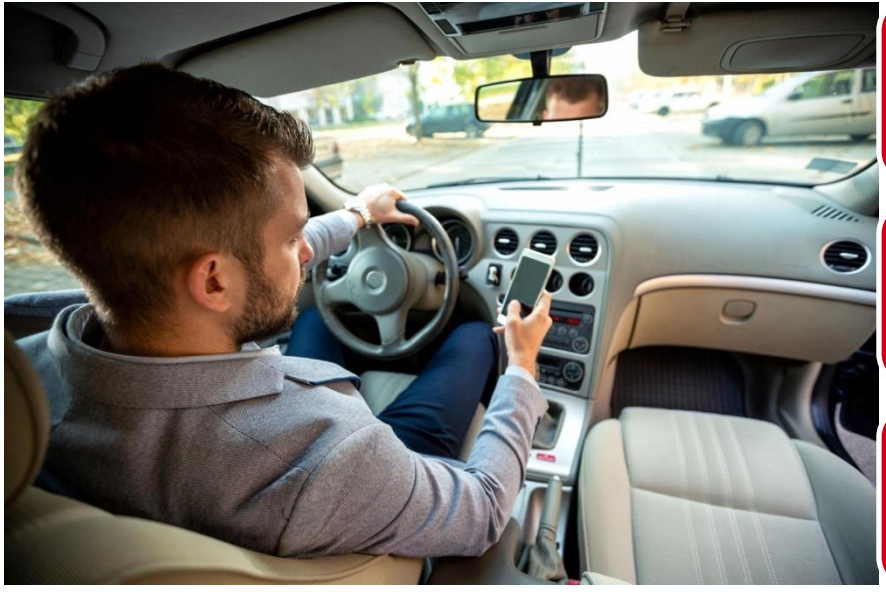
Psychosocial risks: three failure modes

1. We stop thinking

2. Work gets harder, not easier

3. No one knows who's in charge

1. We stop thinking



- Situation awareness
- Automation bias
- Trust

Situation awareness

We stop thinking - Situation awareness

AI can either sharpen or blunt situation awareness depending on how information is filtered, timed, and displayed.

What is situation awareness?

- Perceive
- Understand
- Anticipate

People maintain basic awareness (Level 1 “what’s there”)

But lose deeper understanding (Level 2: “what it means”)

And struggle with projection (Level 3: “what will happen next”)

We stop thinking – Automation bias

People systematically defer to automated systems - especially under time pressure, uncertainty, and high workload.

Errors of omission
Errors of commission

Why?

- Cognitive offloading (we stop checking)
- Authority bias (“the system knows better”)
- Reduced vigilance under automation (Parasuraman)

Psychosocial risks:

- Reduced confidence in own judgement
- Learned helplessness over time
- Blame shifts to individuals when systems fail (“should’ve known better”)

We stop thinking – Trust in AI

Right trust. Right time. Right reason

Aotearoa New Zealand has one of the lowest levels of reported trust in AI (KPMG, 2025).

This matters because:

- Trust is a gatekeeper for adoption
- Low trust can lead to anxiety and lowered wellbeing
- Trust is socially constructed - contagious
- Trust can be mis-calibrated
- Algorithm aversion leads to a lack of transparency

Trust calibration

Over trusting is just as limiting as under trusting (Lee and See, 2025).

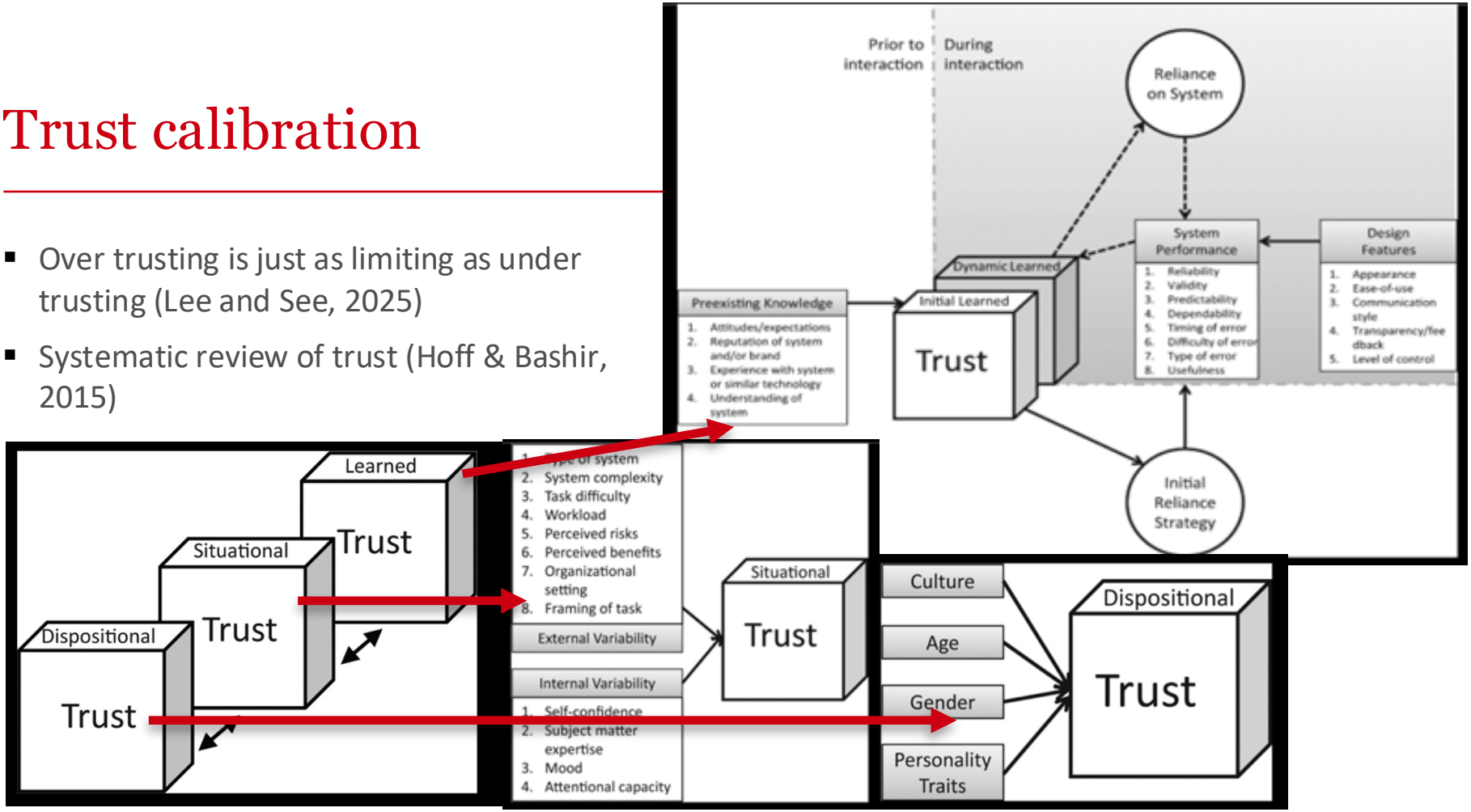
Systematic review of trust (Hoff & Bashir, 2015).

6-D model of national culture (Hofstede 2010) NZ scores low on power distance, and high on individualism.

We can measure this using signal detection theory.

Trust calibration

- Over trusting is just as limiting as under trusting (Lee and See, 2025)
- Systematic review of trust (Hoff & Bashir, 2015)



2. Work gets harder, not easier

- Workload
- Monitoring/vigilance
- Emotional load

Work gets harder, not easier - Workload

AI Shifts Workload — It Doesn't Just Reduce It

Workload impacts wellbeing through:

- Cognitive limitations
- Not enough cognitive rest
- Dual tasks, and competing demands
- Automation complacency
- Emotional load

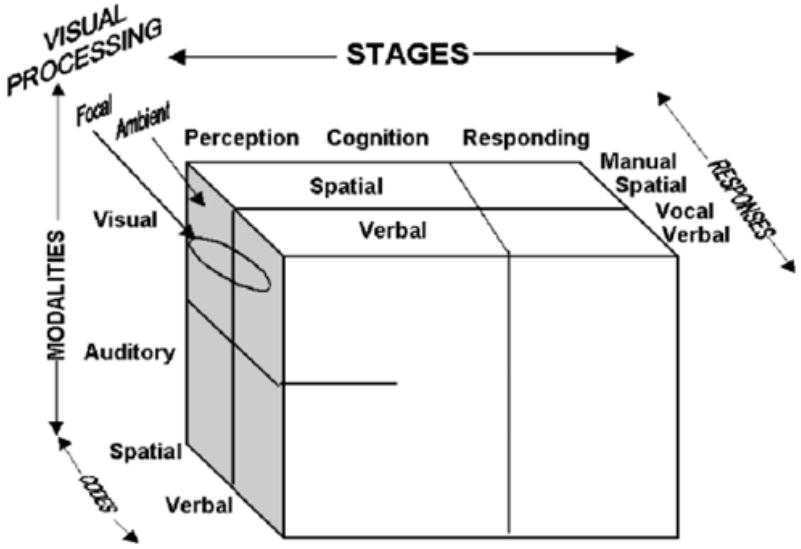


Figure 1. The 4-D multiple resource model.

Work gets harder, not easier - Monitoring/vigilance

The job becomes: notice the one moment that matters — after hours of nothing

- AI shifts work to monitoring
- Humans are poor at sustained attention
- Low activity = low engagement
- Rare events are hardest to detect
- When AI fails, response must be immediate

AI shifts many roles to vigilance tasks. AI is not often wrong 2%, therefore a high amount of vigilance is required. We need to protect against mental passivity and be aware of how roles are changing and what skills will be needed.

Work gets harder, not easier – Emotional load

AI removes admin, not emotional responsibility

- Functional identity perspective – Selenko et al

Work gives people:

- competence
- meaning
- belonging
- self-esteem
- social recognition

3. No one knows who's in charge

AI teaming

Roles

Accountibility

Who's in charge - AI teaming

AI Is Not a Tool — It's a Teammate (Whether You Like It or Not)

Safety depends on coordination between human and AI roles.

What goes wrong:

- Role ambiguity ("who's responsible?")
- Poor handovers between human and AI
- No shared mental model - If the human and AI don't share the same picture, coordination breaks

Automation trade-offs - performance and workload depend on the type & level of automation

Who's in charge? - Roles

If roles aren't explicit, people hesitate — or act at the wrong time

- Role ambiguity increases error and delay
- Automation shifts roles from action → supervision
- Function allocation matters (not just capability)
- Authority and override must be defined
- Poor role design creates "out-of-the-loop" problems

Accountability

"US lawyer sanctioned after being caught using ChatGPT for court brief"

"AI keeps inventing fake cases. Lawyers keep citing them"

"Lawyer Fined for Using AI-Generated Legal Documents with Fake Citations"

"California prosecutors' office used AI to file inaccurate motion in criminal case"

Accountability

When AI gets it right, we call it innovation. When it gets it wrong… we call it human error.

- Responsibility is often unclear in AI systems
- Accountability does not disappear — it shifts
- Humans are held responsible without full control
- Blame follows outcome, not system design
  Good outcomes → system credited
  Bad outcomes → human blamed

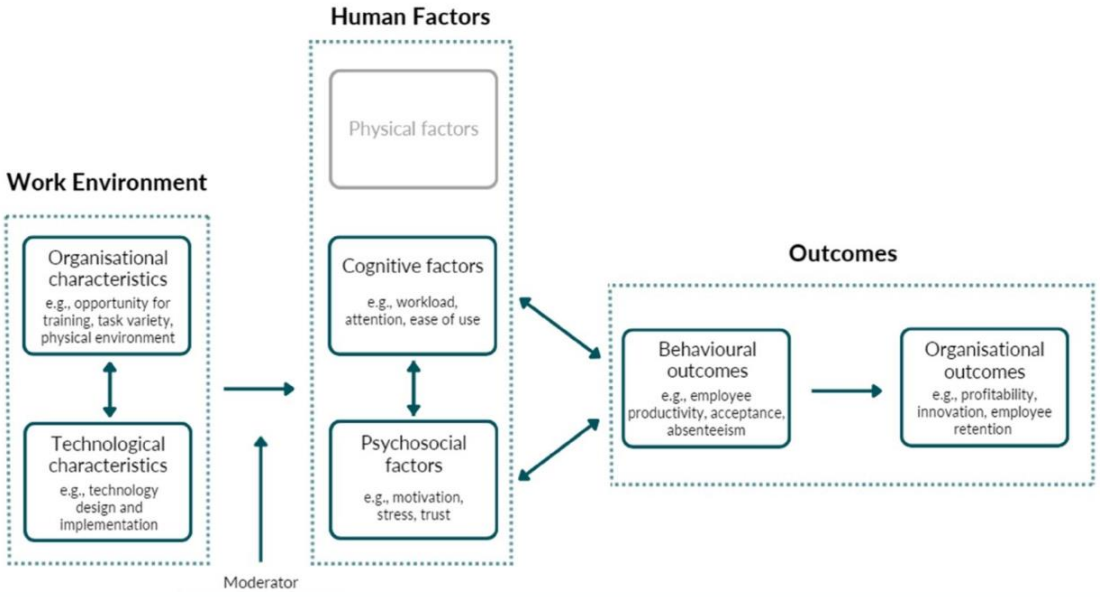
We automate decisions… but not accountability.
When we play the 'blame game', transparency fails.

We're asking the wrong question

Not how do we stop AI causing psychological harm?

But

What kind of work system are we creating when we introduce AI?



Controversial(?) takeaway

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

AI failures are system failures, not human failures

- Stop blaming users, start blaming systems and design
- Don’t ask “who made the mistake” ask “what made the mistake possible”

Human-centred design solutions for a healthy wellbeing ecosystem

Classification: In-Confidence

How do we design with wellbeing in mind?

“Design isn’t about the thing. It’s about the *experience* the thing enables.” – Brenda Laurel
The entire field of human-centred design in one sentence.

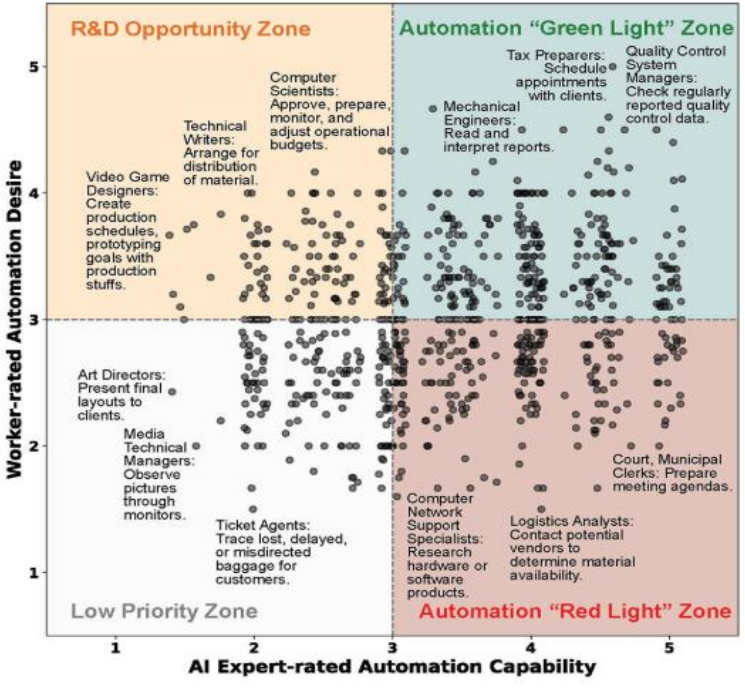
Know what is going on in your organisation – this isn’t a tech problem

- Values based decisions
- Trust calibration
- Workload measures
- Change models – which work best for AI and an ongoing change
- Track wellbeing
- Know what problems to solve and identify the data sets that might be useful
- Being able to ask the right questions – very important going forward

Design principles: how to know what is good?

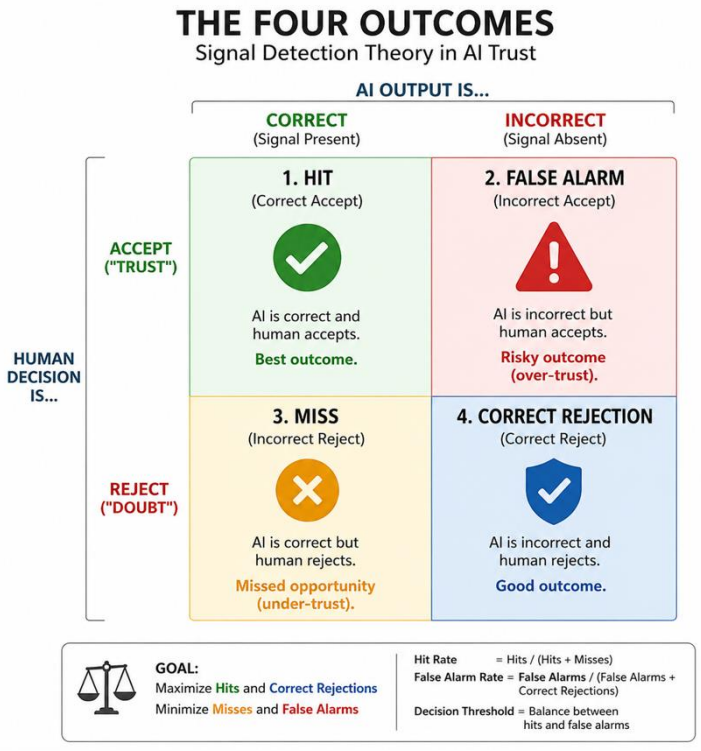
- Green Light Zone: Tasks with high automation desire and high capability
- Red Light Zone: Tasks with low desire and high capability
- R&D Opportunity Zone: Tasks with high desire but low capability
- Low Priority Zone: Tasks with low desire and low capability

Erik Brynjolfsson, Diyi Yang



Dividing worker and AI expert perspectives among four task zones showed mismatches in what workers want and what is technically feasible. | Stanford Social and Language Technologies (SALT) Lab

Signal detection theory and trust



Consideration of Maori culture

The five tests as a framework of assessment

“a method or methods for assessing a situation or event that challenges our thinking and our values.”

Mead stresses that this process results in *a* position, not *the* definitive position.

Test 1: Tapu

Test 2: Mauri

Test 3: Take-utu-ea

Test 4: Precedent

https://www.taiuru.co.nz/maori-data-governance-in-2025/ - Governance focus

What's missing

- Not enough women, DEI, indigenous ethnicities or other minorities.
- Different Value Lens (efficiency, profit vs trust, equity and wellbeing).
- Emerging Unknowns: AI risks and benefits continue to evolve - issues such as model drift, long-term bias, or social ripple effects are not yet fully understood.
- Variable Readiness (Digital maturity, data governance, and leadership capability).
- Stable regulation

What's missing from the research

- Vigilance and sustained attention – functional vigilance
- Cognitive skill degradation
- Human-AI trust calibration
- Teammate cognition
- Identity and meaningful work
- Safety and psychosocial risks
- Human design of AI – practical frameworks

Closing

Practical takeaways:

1. When do people stop thinking
2. When is work harder, not easier
3. Who's in charge, and does everyone know?

If you don't design for people:

- Workload spikes
- Trust mis-calibrates
- Skills decay
- Safety and psychosocial risks increase
- The wellbeing ecosystem fails

Patai?
Questions?

Classification: In-Confidence

Ka kite
Thanks

Classification: In-Confidence

References

Alami, J., El Iskandarani, M., & Riggs, S. L. (2025). The Effect of Workload and Task Priority on Multitasking Performance and Reliance on Level 1 Explainable AI (XAI) Use. *Human Factors*. https://doi.org/10.1177/00187208251323478

Alarcon, G. M., Capiola, A., Lee, M. A., Willis, S., Hamdan, I. A., Jessup, S. A., & Harris, K. N. (2024). Development and Validation of the System Trustworthiness Scale. *Human Factors*, *66*(7), 1893–1913. https://doi.org/10.1177/00187208231189000

Australia, K. (2025). *Trust, attitudes and use of artificial intelligence A global study 2025*. https://doi.org/10.26188/28822919

Bai, L., Liu, X., & Su, J. (2023). ChatGPT: The cognitive effects on learning and memory. *Brain-X*, *1*(3). https://doi.org/10.1002/brx2.30

Bowdler, M., Lahti, H., Jelenko, M., Buresti, G., & Valtonen, T. (2026). Algorithmic management and psychosocial risks at work: An emerging occupational safety and health challenge. *Scandinavian Journal of Work, Environment & Health*, *52*(1), 1–5. https://doi.org/10.5271/sjweh.4270

Calvo, R. A., Peters, D., Vold, K., & Ryan, R. M. (2020). Supporting Human Autonomy in AI Systems: A Framework for Ethical Enquiry. In *Philosophical Studies Series* (Vol. 140, pp. 31–54). Springer Nature. https://doi.org/10.1007/978-3-030-50585-1_2

Carayon, P. (2006). Human factors of complex sociotechnical systems. *Applied Ergonomics*, *37*(4 SPEC. ISS.), 525–535. https://doi.org/10.1016/j.apergo.2006.04.011

Cooke, N., Demir, M., & Huang, L. (2020). *A Framework for Human-Autonomy Team Research* (pp. 134–146). https://doi.org/10.1007/978-3-030-49183-3_11

Cooke, N. J., Cohen, M. C., Fazio, W. C., Inderberg, L. H., Johnson, C. J., Lematta, G. J., Peel, M., & Teo, A. (2024). From Teams to Teamness: Future Directions in the Science of Team Cognition. *Human Factors*, *66*(6), 1669–1680. https://doi.org/10.1177/00187208231162449

Dong, J., & Jeon, M. (2025). Happiness improves perceptions and game performance in an escape room, whereas anger motivates compliance with instructions from a robot agent. *International Journal of Human Computer Studies*, *202*. https://doi.org/10.1016/j.ijhcs.2025.103547

Fiegler-Rudol, J., Lau, K., Mroczek, A., & Kasperczyk, J. (2025). Exploring Human–AI Dynamics in Enhancing Workplace Health and Safety: A Narrative Review. In *International Journal of Environmental Research and Public Health* (Vol. 22, Number 2). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/ijerph22020199

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Gillespie, S. , W. T. , M. A. , & H. G. (2025). *Trust, attitudes and use of artificial intelligence A global study 2025*. https://doi.org/10.26188/28822919

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, *53*(5), 517–527. https://doi.org/10.1177/0018720811417254

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Jia, N., Luo, X., Fang, Z., & Liao, C. (2024). WHEN AND HOW ARTIFICIAL INTELLIGENCE AUGMENTS EMPLOYEE CREATIVITY. *Academy of Management Journal*, *67*(1), 5–32. https://doi.org/10.5465/amj.2022.0426

References

Jörs, J. M., & De Luca, E. W. (2024). *Through Eudaimonia's Lenses: EudAImonic Well-Being Variables As Guiding Principles For AI Design*. https://www.researchgate.net/publication/380938771

Kadoma, K., & Metaxa, D. (n.d.). *Generative AI and Perceptual Harms: Who's Suspected of using LLMs?* (Vol. 1).

Kim, B. J., & Lee, J. (2024). The mental health implications of artificial intelligence adoption: the crucial role of self-efficacy. *Humanities and Social Sciences Communications*, *11*(1). https://doi.org/10.1057/s41599-024-04018-w

Kim, J. Y., Lester, C., & Yang, X. J. (2025). Beyond Binary Decisions: Evaluating the Effects of AI Error Type on Trust and Performance in AI-Assisted Tasks. *Human Factors*. https://doi.org/10.1177/00187208251326795

Klaus Schwab. (2016). *The Fourth Industrial Revolution*. World Economic Forum .

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (n.d.). *Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task ⚠*.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lou, B., Lu, T., Raghu, T. S., & Zhang, Y. (2025). *Unraveling Human-AI Teaming: A Review and Outlook*. http://arxiv.org/abs/2504.05755

Lyons, J. B., Sycara, K., Lewis, M., & Capiola, A. (2021). Human–Autonomy Teaming: Definitions, Debates, and Directions. In *Frontiers in Psychology* (Vol. 12). Frontiers Media S.A. https://doi.org/10.3389/fpsyg.2021.589585

Mcknight, D. H., Choudhury, V., & Kacmar, C. (2002). *Developing and Validating Trust Measures for e-Commerce: An Integrative Typology*.

Morris, A. (2023). The Dilemma of Digital Colonialism unmasking facial recognition technology and data sovereignty in Aotearoa New Zealand. *Policy Quarterly*, *19*(1), 80–87. https://doi.org/10.26686/pq.v19i1.8109

Mosier, K. L., Skitka, L. J., Burdick, M. D., & Heers, S. T. (1996). Automation Bias, Accountability, and Verification Behaviors. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *40*(4), 204–208. https://doi.org/10.1177/154193129604000413

Nazareno, L., & Schiff, D. S. (2021). The impact of automation and artificial intelligence on worker well-being. *Technology in Society*, *67*, 101679. https://doi.org/10.1016/j.techsoc.2021.101679

Oksanen, A., Savela, N., Latikka, R., & Koivula, A. (2020). Trust Toward Robots and Artificial Intelligence: An Experimental Approach to Human–Technology Interactions Online. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.568256 Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

References

Parasuraman, R., & Wickens, C. D. (2008). Humans: Still Vital After All These Years of Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 511–520. https://doi.org/10.1518/001872008X312198

Parker, S. K., & Grote, G. (2022). Automation, Algorithms, and Beyond: Why Work Design Matters More Than Ever in a Digital World. *Applied Psychology*, *71*(4), 1171–1204. https://doi.org/10.1111/apps.12241

Passalacqua, M., Pellerin, R., Magnani, F., Doyon-Poulin, P., Del-Aguila, L., Boasen, J., & Léger, P. M. (2025). Human-centred AI in industry 5.0: a systematic review. In *International Journal of Production Research* (Vol. 63, Number 7, pp. 2638–2669). Taylor and Francis Ltd. https://doi.org/10.1080/00207543.2024.2406021

Read, G. J. M., Salmon, P. M., Lenné, M. G., & Stanton, N. A. (2015). Designing sociotechnical systems with cognitive work analysis: putting theory back into practice. *Ergonomics*, *58*(5), 822–851. https://doi.org/10.1080/00140139.2014.980335

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems. *Human Factors*, *58*(3), 377–400. https://doi.org/10.1177/0018720816634228

Singh, A., Taneja, K., Guan, Z., & Ghosh, A. (2025). *Protecting Human Cognition in the Age of AI*. http://arxiv.org/abs/2502.12447

Siu, H. C., Peña, J. D., Zhou, Y., & Allen, R. E. (2025). *In Pursuit of Predictive Models of Human Preferences Toward AI Teammates*. http://arxiv.org/abs/2503.15516

Squire, P. N., & Parasuraman, R. (2010). Effects of automation and task load on task switching during human supervision of multiple semi-autonomous robots in a dynamic environment. *Ergonomics*, *53*(8), 951–961. https://doi.org/10.1080/00140139.2010.489969

Strich, F., Mayer, A. S., & Fiedler, M. (2021). What do i do in a world of artificial intelligence? Investigating the impact of substitutive decision-making ai systems on employees' professional role identity. *Journal of the Association for Information Systems*, *22*(2), 304–324. https://doi.org/10.17705/1jais.00663

Valle-Cruz, D., García-Contreras, R., & Gil-Garcia, J. R. (2024). Exploring the negative impacts of artificial intelligence in government: the dark side of intelligent algorithms and cognitive machines. *International Review of Administrative Sciences*, *90*(2), 353–368. https://doi.org/10.1177/00208523231187051

Wickens, C. D. (2008). Multiple resources and mental workload. In *Human Factors* (Vol. 50, Number 3, pp. 449–455). https://doi.org/10.1518/001872008X288394

Xu, W., Gao, Z., & Dainoff, M. J. (2026). An HCAI Methodological Framework: Putting it Into Action to Enable Human-Centered AI. *IEEE Transactions on Human-Machine Systems*. https://doi.org/10.1109/THMS.2025.3631590

Yam, K. C., Tang, P. M., Jackson, J. C., Su, R., & Gray, K. (2023). The rise of robots increases job insecurity and maladaptive workplace behaviors: Multimethod evidence. *Journal of Applied Psychology*, *108*(5), 850–870. https://doi.org/10.1037/apl0001045

AI use

Gen AI was used in the making of these slides to summarise info for concise wording suitable for presentation.