

Embracing vulnerability in AI-mediated coaching: A developmental pathway for ethical HRD

Sebastian Kot

*The Management Faculty, Czestochowa University of Technology,
Poland;
Faculty of Economics and Management Sciences, North-West
University, Vaal Triangle Campus, Vanderbijlpark,
South Africa
sebastian.kot@pcz.pl
ORCID 000-0002-8272-6918*

Harmandeep Kaur

*Research and Development, Scientific Societal & Behavioral Research
Center (SSBRC),
Canada
harman@ssbr.com
ORCID 0009-0004-3876-2104*

Adnan ul Haque

*Business and Management, Yorkville University,
Canada;
Research and Development, Scientific Societal & Behavioral Research
Center (SSBRC),
Canada
ahaque@yorkvilleu.ca
ORCID 0000-0003-2051-8635*

Abstract. Artificial intelligence (AI) is increasingly being integrated into human resource development (HRD) practices, particularly in coaching, mentoring, and feedback systems. While AI enhances scalability and efficiency, it may also diminish the sensitivity required for meaningful learning, ethical reflection, and relational development. This study explores how individuals experience vulnerability in AI-mediated coaching. Drawing on semi-structured interviews with 34 HRD practitioners, coaches, and professionals across Canadian organizations, we identify three patterns of vulnerability enactment (1) algorithmic suppression, where AI systems inadvertently signal psychological risk; (2) mediated vulnerability, where individuals navigate partial openness in digital contexts; and (3) cultivated courage, where intentional HRD practices allow vulnerability to flourish despite technological constraints. Our findings indicate that AI systems often reduce relational cues and prioritize efficiency over developmental depth. However, when HRD practitioners deliberately design for

Received:

July, 2025

1st Revision:

March, 2026

Accepted:

June, 2026

DOI:

10.14254/2071-
8330.2026/19-2/2

vulnerability through facilitator presence, reflective pauses, and ethical framing: AI-mediated environments can support, rather than hinder, developmental growth. We contribute to HRD theory by extending vulnerability as a developmental construct into AI-mediated contexts and refining the automation–augmentation paradox to include vulnerability as a critical moderating mechanism. Practically, we offer guiding principles for designing AI-enabled HRD tools that preserve human connection, ethical reflection, and social justice.

Keywords: Artificial intelligence, HRD, vulnerability, coaching, ethical development, social justice, Canada.

JEL Classification: J24, O33, M53

1. INTRODUCTION

Artificial intelligence has moved from the periphery to the core of organizational life. In human resource development (HRD), AI is now shaping how individuals learn, receive feedback, and progress professionally (Yoo et al., 2024). From AI-powered coaching platforms that deliver personalized development plans to algorithmic feedback systems that analyze communication patterns, technology is increasingly mediating developmental relationships. This transformation has significant implications for HRD's core mission: fostering human growth, reflective learning, and ethical practice (Vaijyanthi & Dhanushree, 2026; Dhakal et al., 2023; Hernik et al., 2025). However, as algorithmic systems increasingly drive developmental processes, HRD faces a fundamental paradox. While AI offers efficiency, scalability, and data-driven precision, it may inadvertently constrain the very human capacities essential for development, including empathy, moral reasoning, and the willingness to be vulnerable (Raisch & Krakowski, 2021). Vulnerability involves the courage to acknowledge uncertainty, receive feedback openly, and engage in the relational exposure required for transformative learning (Edmondson & Bransby, 2023). When AI systems reduce developmental interactions to algorithmic nudges and automated feedback, they risk suppressing the vulnerability necessary for genuine development.

We argue that vulnerability offers a critical lens for addressing this question. Rather than framing vulnerability as a risk to be managed in AI implementation, we position it as a developmental resource that HRD practitioners must intentionally cultivate in technology-mediated environments.

Despite growing interest in AI and HRD (Ghosh & Lee, 2024; Mendy et al., 2025), empirical research on how vulnerability operates in AI-mediated developmental contexts remains limited. This gap is significant because, without such understanding, organizations risk implementing AI in ways that undermine the relational and ethical foundations of HRD.

This study addresses this gap through research questions:

1. "How do individuals experience vulnerability in AI-mediated coaching and developmental relationships?"
2. "What technological and organizational conditions help or restrict vulnerability in AI-facilitated HRD practices?"
3. "How can HRD practitioners incorporate vulnerability into the design and implementation of AI-enabled developmental tools to assist ethical and relational learning?"

To address these questions, we conducted a qualitative study using semi-structured interviews with 34 HRD practitioners, coaches, and professionals across Canadian organizations who have experience with AI-enabled developmental tools. Our findings identify three distinct patterns of vulnerability enactment in AI-mediated environments: algorithmic suppression, where AI systems inadvertently signal psychological

risk; mediated vulnerability, where individuals navigate partial openness in digital contexts; and cultivated courage, where intentional HRD practices enable vulnerability to flourish despite technological constraints.

Our study makes three key contributions. First, we extend HRD theory by reconceptualizing vulnerability as a developmental resource in AI-mediated contexts, moving beyond traditional framings of vulnerability as risk or deficit (Sepehr et al., 2025). Second, we refine the automation-augmentation paradox (Raisch & Krakowski, 2021) by demonstrating how vulnerability functions as a critical moderating mechanism between technological efficiency and developmental depth. Third, we offer empirically grounded principles for designing AI-enabled HRD practices that preserve ethical reflection, relational learning, and social justice.

2. LITERATURE REVIEW

2.1. Vulnerability as a developmental construct

In organizational scholarship, vulnerability is often framed as a condition to be minimized or guarded against. However, a growing body of research positions vulnerability not as weakness but as a foundational condition for learning, trust, and authentic relationships (Edmondson & Bransby, 2023; Kahn, 2018). In HRD, vulnerability reflects the willingness to expose oneself to uncertainty, feedback, and emotional risk in the pursuit of growth (Edmondson & Bransby, 2023).

Edmondson's (1999) seminal work on psychological safety highlights that individuals must feel safe to engage in interpersonal risk-taking such as asking questions, admitting mistakes, or sharing ideas without fear of humiliation. At its core, such risk-taking constitutes an act of vulnerability. More recent work by Sepehr et al. (2025) suggests that vulnerability extends beyond an individual trait and should be understood as a relational and organizational phenomenon. Organizations that normalize vulnerability tend to foster cultures of trust, innovation, and adaptive learning.

Within leadership and HRD scholarship, vulnerability has been closely linked to authenticity, ethical courage, and transformative learning (Edmondson & Bransby, 2023; Lemoine et al., 2019; Mishchuk et al., 2025). Leaders who demonstrate vulnerability signal that imperfection and learning are valued, thereby encouraging others to engage in developmental risk-taking (Cunha et al., 2019), supporting the participative style of HRM (Belás et al., 2024). In HRD practice, vulnerability supports openness to feedback, self-reflection, and relational engagement that are key elements of effective coaching and mentoring (Boyatzis, 2025).

Importantly, vulnerability is not merely the absence of psychological safety; it is an active developmental resource. It involves the capacity to remain open to uncertainty and emotional exposure throughout the process of growth. For HRD, this implies that cultivating vulnerability is not about eliminating risk but about creating conditions in which individuals can engage with risk constructively. The recent work of Rickert et al. (2024) found that technology introduces new forms of vulnerability in workplace.

2.2. The AI-HRD paradox

The integration of AI into HRD presents both opportunities and challenges. On one hand, AI offers unprecedented scalability, personalization, and data-driven insights into learning and development (Yoo et al., 2024). AI-powered coaching platforms can deliver real-time feedback, identify skill gaps, and provide tailored developmental content at scale. These capabilities have the potential to democratize access to developmental resources and enhance organizational effectiveness.

On the other hand, AI risks constraining the human dimensions of development. Raisch and Krakowski (2021) describe the automation–augmentation paradox: while AI automates routine tasks, it can also enhance human capabilities but only when organizations intentionally design for augmentation rather than mere automation. In HRD contexts, this paradox becomes particularly pronounced. When AI mediates coaching, feedback, or mentoring, it may prioritize efficiency over empathy, data over relational depth, and standardization over personal meaning (Ghosh & Lee, 2024).

Scholars have raised concerns about AI's potential to erode reflective practice, emotional intelligence, and moral reasoning in developmental contexts (Li, 2024; Mendy et al., 2025; Zsigmond & Mura, 2023). Additionally, the work of Diller et al. (2024) revealed that threat-related affective responses are a result of AI coaching, which affirms concerns about algorithmic suppression of vulnerability. By their nature, algorithmic systems rely on patterns and predictions, making it difficult to accommodate the ambiguity, emotion, and ethical complexity inherent in human development. Moreover, AI-mediated interactions may reduce relational cues such as tone, body language, and shared vulnerability that signal safety and trust (Ardichvili, 2022).

This tension is not inevitable; outcomes depend on how AI is designed and implemented. When HRD practitioners approach AI as a tool to augment rather than replace human connection, it can support rather than undermine development. However, achieving this requires deliberate attention to the conditions that enable vulnerability in digital environments.

2.3. Ethical HRD and social justice

HRD has increasingly been conceptualized as a moral and ethical practice. Beyond improving performance, its mission encompasses fostering human dignity, reflective learning, and socially responsible organizational practices (Vaijayanthi & Dhanushree, 2026). This ethical dimension becomes especially significant in the context of AI, where algorithmic systems may reinforce bias, reduce transparency, and marginalize already vulnerable populations (Lima & Rahman, 2024).

The special issue's call explicitly states vulnerability at the intersection of HRD, ethics, and social justice. This alignment is critical. Social justice in HRD requires attention to how developmental opportunities are distributed, whose voices are heard, and who is included in learning processes (Ardichvili, & Harmon, 2022). Vulnerability is understood as the willingness to be open, uncertain, and exposed is essential for recognizing one's own biases, engaging with marginalized perspectives, and committing to equitable practice. Moreover, vulnerability supports the moral courage needed to question algorithmic decisions, advocate for fairness, and resist technological solutions that compromise human dignity (Hannah et al., 2011). In this sense, cultivating vulnerability in AI-mediated HRD is not merely a developmental preference but an ethical imperative. This links ethical leadership (which requires vulnerability) to stress and innovation (Haque & Yamoah, 2021; Haque et al., 2025; Knezevic, 2023).

2.4 Conceptual framework

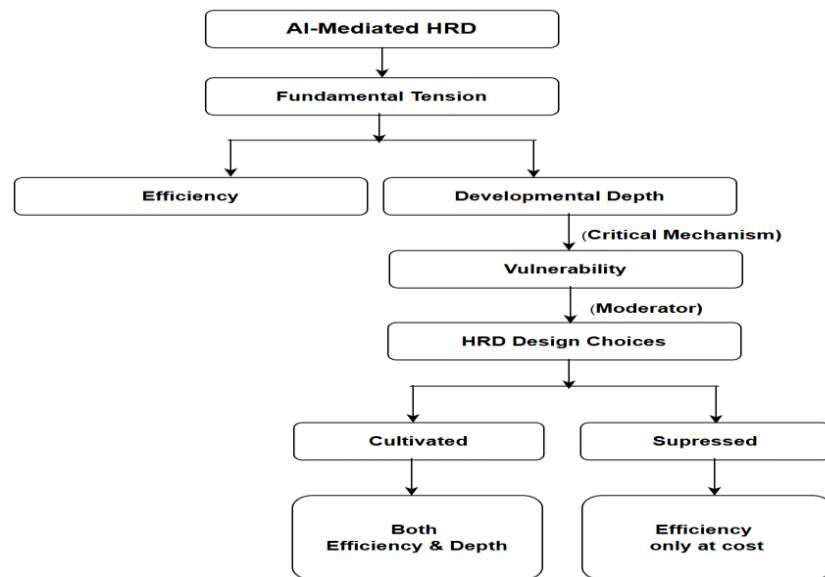


Figure 1. Conceptual Framework Linking AI Mediation, Vulnerability, and Developmental Outcomes in HRD

Source: author's own illustration

By synthesizing these streams of literature, we propose a conceptual framework that positions vulnerability as a bridging mechanism between technological capability and developmental integrity.

Figure 1 represent the framework visually. The framework posits four key relationships:

- First, AI mediation in HRD creates a fundamental tension between efficiency (scalability, data-driven insight) and developmental depth (relational learning, ethical reflection, authentic growth).
- Second vulnerability is defined as the willingness to engage with uncertainty and emotional exposure that acts as a critical mechanism that enables developmental depth to be sustained even within technologically mediated environments. Developmental depth becomes possible due to vulnerability while without vulnerability, efficiency dominates.
- Third, HRD design choices (e.g., facilitator presence, reflective pauses, ethical framing, transparency regarding AI's role) moderate whether vulnerability is suppressed or cultivated. Key moderators identified by our literature synthesis include ethical framing (transparency about AI 's role and limitation), facilitator presence (human coaches bridging AI tools), and data governance (clear policies on data usage and confidentiality).
- Fourth, when vulnerability is cultivated, AI-mediated HRD can achieve both efficiency and developmental depth, thereby supporting ethical learning and social justice. When vulnerability is suppressed, gains in efficiency come at the expense of relational and ethical integrity.

This framework guides our empirical investigation. Rather than assuming that AI inherently suppresses vulnerability or conversely that all AI is concern/problematic, we examine how individuals experience vulnerability in AI-mediated environments and identify the conditions that enable it to flourish. Hence, this framework informs both our research questions as well as our analytical approach.

3. METHODOLOGY

3.1. Research paradigm

This study adopts an interpretivist, qualitative research paradigm. Given the exploratory nature of our research questions and the need to understand participants lived experiences of vulnerability in AI-mediated contexts, a qualitative approach is most appropriate (Haque & Yamoah, 2021).

3.2. Ethics approval and compliance

This study was conducted in accordance with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2), which governs research involving human participants in Canada. This study adhered to all three principles throughout the research.

Ethics approval was obtained from the lead author's institutional Research Ethics Board (Protocol #2026-08EP) prior to data collection. This research was funded by the Scientific Societal & Behavioral Research Center (SSBRC; 2026).

Key ethical considerations included:

- All participants provided written informed consent prior to participation.
- Identifying information (name, job title, organization) was removed, and participants were assigned pseudonyms (P1, P2, P3, etc.).
- Data was stored on encrypted, password-protected institutional servers.
- Participants were provided with the list of professional support resources (mental health counselling services and employee assistance program). A debriefing procedure was conducted following each interview, asking participants to feel free to share if they have any concerns, special request, or wish to withdraw. No participant chose to withdraw.

These procedures ensured that the study met or exceeded Canadian research ethics standards.

3.3. Sampling strategy

Participants were recruited using a combination of purposive and snowball sampling strategies (Faizan et al, 2019). Eligible participants met the following criteria:

- Currently employed in Canada
- Hold a role in HRD, coaching, talent development, or people management
- Have direct experience with AI-enabled developmental tools (e.g., AI coaching platforms, automated feedback systems, AI-assisted performance tools)
- Possess at least two years of professional experience in their field

Recruitment channels included professional networks operating in Canada, LinkedIn outreach, and referrals from initial participants. Recruitment continued until theoretical saturation was reached the point at which additional interviews yielded no new conceptual insight.

Table 1 summarizes the demographic and professional characteristics of the 34 participants.

Table 1

Participant demographic and professional characteristics			
Characteristics	Category	Count (n=34)	Percentage
Gender	Male	11	32.4%
	Female	22	64.7%
	Non-binary	1	2.9%
Professional Role	Executive Coach	9	26.5%
	HRD Practitioner	14	41.2%
	People Manager	6	17.6%
	Talent Development Lead	5	14.7%
Total Experience	2-5 Years	7	20.6%
	6-10 Years	15	44.1%
	11-15 Years	8	23.5%
	15 or above	4	11.8%
Sector/Industry	Education	7	20.6%
	Professional Services	2	5.9%
	Financial Services	4	11.8%
	Technology	12	35.3%
	Healthcare	3	8.8%
	Manufacturing	2	5.9%
	Not-For-Profit	4	11.8%
AI Tool Experience	AI Learning Platform	21	61.8%
	AI-Assisted Performance Tool	19	55.9%
	AI Coaching	16	47.1%
	Multiple Tools	28	82.4%
	Automated Feedback System	24	70.6%

Source: Field data

3.4. Data collection

Data were collected through semi-structured interviews conducted between January and March 2026. Interviews were conducted via secure video conferencing (Zoom).

The interview protocol was developed based on the study's conceptual framework and research questions. Key topic areas included:

- Participants' experiences with AI-mediated coaching and developmental tools
- Perceptions of vulnerability in digital versus face-to-face developmental contexts
- Conditions that enabled or constrained openness and authenticity in AI-mediated environments
- Strategies used to cultivate or protect vulnerability
- Ethical concerns and social justice considerations

Interviews ranged from 45 to 90 minutes in length. With participants' consent, all interviews were audio-recorded and transcribed verbatim.

3.5. Data analysis

Data analysis followed the six phases of thematic analysis as outlined by Braun and Clarke (2021):

Familiarization: All authors read and re-read the transcripts to become deeply immersed in the data. Initial impressions and observations were recorded in analytic memos.

Initial coding: Two authors independently coded three transcripts using open coding, generating preliminary codes that captured meaningful units of data. Codes were then compared, discussed, and refined to ensure consistency.

Generating themes: Codes were grouped into candidate themes through an iterative process of reviewing, combining, and refining. Visual mapping techniques (thematic maps) were used to explore relationships among themes.

Reviewing themes: Themes were evaluated against both coded extracts and the full dataset to ensure coherence and distinctiveness. Themes that were overly broad or insufficiently supported were refined or discarded.

Defining and naming themes: Each theme was clearly defined, with its scope and boundaries explicitly articulated. Theme names were selected to capture the essence of each pattern.

Producing the report: Findings were presented using representative quotes and interpreted in relation to the research questions and theoretical framework.

NVivo 14 qualitative data analysis software was used to support the analysis. To enhance rigor, we employed strategies recommended by Faizan et al. (2019):

- Investigator triangulation: Multiple authors were involved in coding and theme development.
- Member checking: Five participants reviewed a summary of the findings to ensure alignment with their experiences.
- Peer debriefing: Regular meetings with a research group not involved in the study provided critical feedback on the analysis.
- Reflexivity: Authors maintained reflexive journals documenting their assumptions, biases, and evolving interpretations.

3.6 Methodological limitations

Several methodological limitations should be acknowledged. First, although the sample was diverse across sectors and roles, it may not be representative of all HRD professionals in Canada. Second, participants self-selected into the study. Third, the cross-sectional design captures experiences at a single point in time. Finally, social desirability bias may have influenced participants' willingness to disclose experiences related to vulnerability.

4. EMPIRICAL RESULTS AND DISCUSSION

The analysis identified three overarching themes that characterize participants' experiences of vulnerability in AI-mediated coaching and developmental contexts. Each theme is presented below, supported by representative quotes and interpretive analysis.

4.1. Theme 1: Algorithmic suppression

The first theme illustrates how the design and functionality of AI systems can unintentionally signal psychological risk, thereby discouraging vulnerability. Participants reported that algorithmic mediation often created conditions perceived as incompatible with openness, authenticity, and developmental risk-taking.

Subtheme 4.1.1: The surveillance gaze

Many participants expressed concerns that AI-mediated tools introduced a sense of surveillance, making vulnerability feel unsafe. Unlike human coaches, who ensure confidentiality, AI systems often operate within organizational ecosystems where data may be accessible to managers or aggregated for organizational analytics.

“When I learned that my AI coaching conversations were being aggregated and shared with my manager as part of a team ‘engagement dashboard,’ I completely changed how I used it. I stopped discussing real challenges, the issues I actually needed help with. Instead, I began giving safe responses, treating it like a performance review. That’s not coaching; it’s surveillance disguised as development.” (Participant 12, HRD Practitioner, Technology Sector)

While human coaches provide confidentiality within professional ethical boundaries, AI systems introduce ambiguity around data governance, thereby increasing perceived risk (Kellogg et al., 2020).

Subtheme 4.1.2: Algorithmic judgment

Participants also described how the perception of being “evaluated” by algorithms diminished their willingness to be vulnerable. Even when AI tools were explicitly positioned as developmental rather than evaluative, participants experienced them as judgmental.

“The AI feedback tool gives you a score on your communication style. It says it’s for development, but it’s still a score, a number. You’re being evaluated. And when you’re being evaluated, you don’t experiment or take risks where you might fail. You perform. The vulnerability that enables real growth disappears.” (Participant 18, Executive Coach, Healthcare Sector)

Kim & Kim (2024) found that psychological safety is directly reduced by job insecurity, leading to knowledge hiding behavior. Our study also found the connection between AI and psychological safety, hence, reflecting algorithmic judgement expressed by study participants.

The very features that make AI valuable because of its ability to identify patterns and generate insights that can also make it feel judgmental, thereby suppressing the openness that development requires (Diller et al., 2024).

Subtheme 4.1.3: The absence of relational cues

A third dimension of algorithmic suppression relates to the absence of relational cues that typically signal safety and invite vulnerability. Participants described AI-mediated interactions as lacking the embodied and relational elements necessary for vulnerability to emerge.

“When I’m with a human coach, I can see in their eyes that they are present with me. I can sense whether they are judging me or holding space for me. With the AI platform, there is nothing, it’s just text on a screen. I don’t know how to be vulnerable with something that has no face, no body, and no capacity to care.” (Participant 7, Executive Coach, Professional Services)

These accounts highlight a key limitation of current AI systems: their inability to demonstrate responsive presence which is the capacity to be affected by another’s vulnerability and respond in ways that convey genuine engagement (Turkle, 2021). In the absence of such responsiveness, vulnerability becomes performative rather than transformative.

4.2 Theme 2: Mediated vulnerability

The second theme illustrates how participants navigated the tension between openness and protection in AI-mediated environments, often engaging in partial or strategic forms of vulnerability. Rather than fully suppressing or fully embracing vulnerability, participants described mediated vulnerability.

Subtheme 4.2.1: Strategic self-editing

Participants described carefully selecting what they shared with AI tools disclosing enough to derive value while protecting themselves from potential harm.

“I have a rule: I’ll share work-related challenges with the AI coach, but never personal ones. I’ll talk about project management struggles, but not about feeling overwhelmed or questioning my career. The work-related issues feel safer because they are professional. The personal side that’s where real vulnerability lies, and I’m not willing to share that with an algorithm.” (Participant 14, Talent Development Lead, Education Sector)

This strategic self-editing reflects a form of calculated vulnerability, openness shaped by ongoing risk assessment. While this approach allows participants to gain some benefit from AI tools, it also constrains the depth of developmental engagement (Petriglieri & Petriglieri, 2020).

Subtheme 4.2.2: The human-AI hybrid

Several participants described using AI tools alongside human coaches or mentors, creating hybrid arrangements that allowed them to benefit from AI’s efficiency while maintaining the relational safety necessary for deeper vulnerability.

“I use the AI platform for low-stakes tasks like preparing for meetings, organizing my thoughts, and getting feedback on presentation drafts. For more complex issues career decisions, navigating difficult relationships, or dealing with burnout I turn to my human coach. The AI supports me, but the human guides me. I need human interaction for vulnerability.” (Participant 9, HRD Practitioner, Technology Sector)

This hybrid approach suggests that participants are actively constructing vulnerability portfolios strategic combinations of human and AI resources designed to balance efficiency with relational depth (see Figure 3). This pattern has important implications for how organizations design and implement AI-enabled HRD practices.

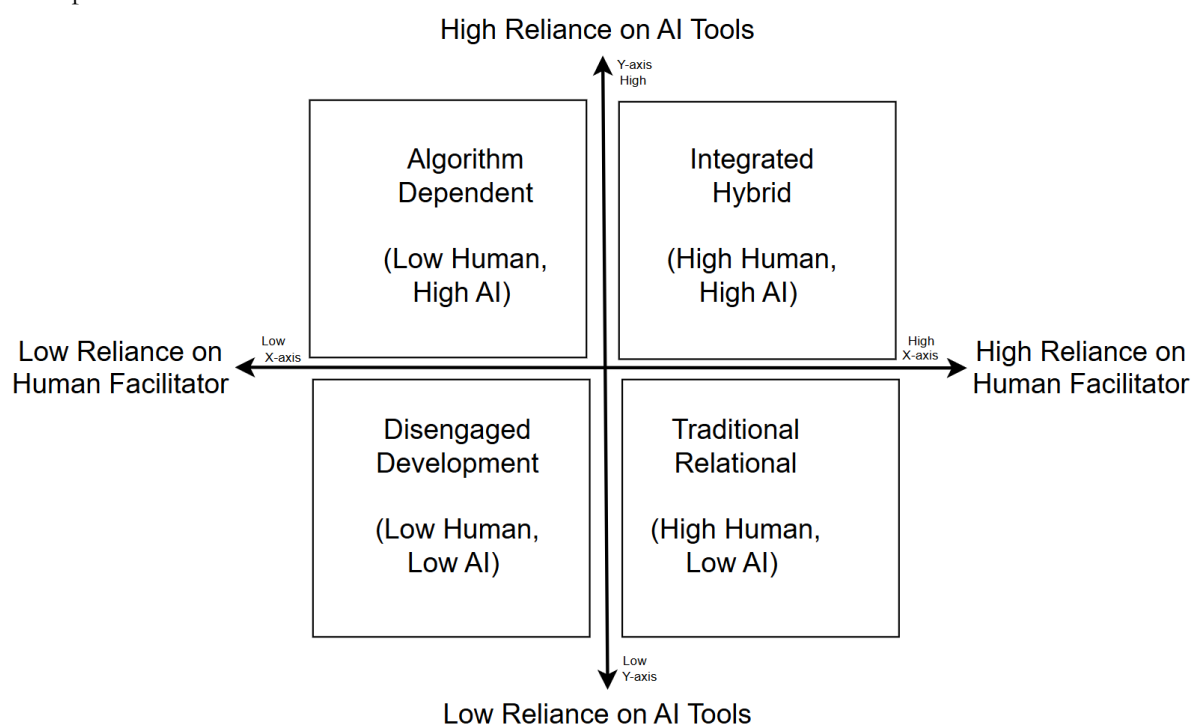


Figure 3. Vulnerability portfolios: Strategic combinations of human and AI resources

Source: Author's own illustration

Note: The horizontal axis represents Reliance on Human Facilitator (Low to High). The vertical axis represents Reliance on AI Tools (Low to High). The Integrated Hybrid quadrant (top-right) represents the optimal portfolio for cultivating vulnerability.

Subtheme 4.2.3: Contextual openness

Participants explained that their willingness to be vulnerable with AI varied significantly across contexts, influenced by factors such as the tool's purpose, data governance practices, and organizational culture.

"In my previous role, I didn't share anything meaningful with AI tools because the culture was punitive. In my current role, I'm more open because there is trust. However, that trust is not about the technology itself, it's about the culture within which the technology operates. If the culture is supportive, I am willing to be vulnerable even with AI. If it is not, no amount of a friendly chatbot interface will make me open up." (Participant 3, Talent Development Lead, Technology Sector)

This finding highlights the critical role of organizational context in shaping vulnerability. AI tools do not operate in isolation; their effects on vulnerability are influenced by the broader organizational culture, policies, and climate of trust (Edmondson, 2018).

4.3. Theme 3: Cultivated courage

The third theme highlights the conditions and practices that enabled participants to experience vulnerability as a developmental resource within AI-mediated contexts. Rather than viewing AI as inherently suppressing vulnerability, participants emphasized how intentional design and facilitation can create environments in which vulnerability is able to flourish.

Subtheme 4.3.1: Facilitator presence

A key enabling condition was the presence of a human facilitator who bridged the AI tool and the participant, providing a relational space in which vulnerability could safely emerge.

"Our organization uses an AI coaching platform, but it is paired with human coaches who review the conversations and provide additional support. That changes everything. Knowing that a human or someone I know and trust is reading my responses makes me willing to be more vulnerable. The AI serves as a tool for capturing my thoughts, but the relationship is with the human." (Participant 19, People Manager, Manufacturing)

This subtheme indicates that AI is most effective when deployed within a relational architecture, which is a structure of human relationships that provides the safety and trust necessary for vulnerability (Maak & Pless, 2006).

Subtheme 4.3.2: Reflective pauses

Participants highlighted that AI tools designed to create opportunities for reflection rather than simply delivering automated responses fostered conditions more conducive to vulnerability.

"The best AI coaching tool I've used doesn't provide immediate answers. It asks a question and then pauses. It says, 'Take a moment to reflect before you respond.' That pause is crucial. It signals that this is about my reflection, not the AI's answer. It gives me the space to be honest with myself before replying." (Participant 11, HRD Practitioner, Education Sector)

This design principle involves temporal affordances, in other words features that provide time and space for reflection rather than optimizing for speed (Wajcman, 2016). Such affordances counter the efficiency-driven default of many AI systems.

Subtheme 4.3.3: Ethical framing and transparency

Finally, participants emphasized that transparency regarding AI's role, capabilities, and limitations enabled them to calibrate their vulnerability more effectively.

"My organization was very clear about what the AI coaching tool could and could not do, as well as how the data would be used. They explained: 'This AI is your assistant, not your judge. Your conversations are confidential. Your manager will not have access to them.' That clarity made me much more willing to be open. When I understand the boundaries, I can operate within them." (Participant 6, Executive Coach, Healthcare Sector)

These accounts highlight the importance of ethical infrastructure such as the policies, practices, and communication mechanisms that govern AI use in creating conditions that support vulnerability.

4.4. Cross-theme synthesis

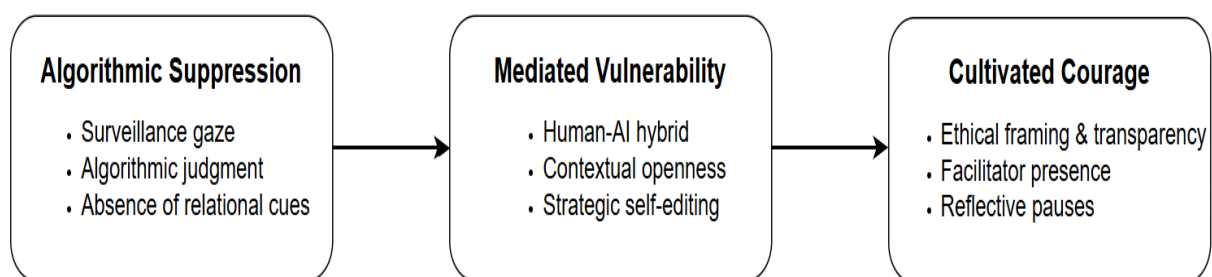


Figure 2. Major three themes and their interrelationships

Source: Authors' own illustration

Figure 2 presents a visual synthesis of the three themes and their interrelationships. The figure illustrates:

- Algorithmic Suppression as the baseline condition, wherein AI systems without intentional design tend to suppress vulnerability through surveillance, perceived judgment, and the absence of relational cues.
- Mediated Vulnerability reflects participants' adaptive responses, involving strategic self-editing, hybrid arrangements, and contextual calibration to balance openness and protection.
- Cultivated Courage highlights the conditions under which vulnerability can flourish, enabled by intentional design choices such as facilitator presence, reflective pauses, and ethical transparency.

The arrows in the figure indicate that movement from suppression to cultivated courage is not automatic; rather, it requires deliberate HRD intervention at the levels of technology design, organizational practice, and cultural infrastructure.

4.5. Theoretical contributions

We extend the concept of vulnerability as a developmental construct into AI-mediated contexts. Prior research has established vulnerability as fundamental to learning, trust, and authentic growth in human-to-human interactions (Edmondson & Bransby, 2023). Our findings demonstrate that vulnerability operates differently in AI-mediated environments. This extends existing theorizing by emphasizing the technological embeddedness of vulnerability, a dimension largely overlooked in prior HRD scholarship (Sepehr et al., 2025). We refine the automation–augmentation paradox by identifying vulnerability as a critical moderating mechanism. Raisch and Krakowski (2021) argue that AI can either automate or augment human capabilities depending on how it is designed. We introduce the concept of vulnerability portfolios to explain how

individuals strategically combine human and AI resources for development. Future HRD scholarship should examine how vulnerability portfolios are constructed and how they evolve as AI capabilities advance. We contribute to the emerging literature on ethical HRD and social justice by demonstrating how vulnerability intersects with fairness and inclusion. Organizations genuinely committed to social justice require attention not only to algorithmic bias but also to the conditions that enable vulnerability (Ardichvili & Harmon, 2022).

Similar tensions between technological opportunity and ethical risk have been documented through expert interviews in other AI-mediated professional domains, where transparency regarding data use and regulatory clarity emerged as central preconditions for stakeholders' willingness to engage openly with AI tools (Schinello, 2025).

4.6. Practical implications

Our findings have several implications for HRD practice, AI design, and organizational governance.

AI tools should be paired with human facilitators who can provide the relational safety necessary for vulnerability.

Clear policies regarding data use, confidentiality, and the limitations of AI tools are essential for enabling vulnerability.

The value of AI in development lies not in faster answers, but in better questions and the time to engage with them (Li, 2024).

Our findings suggest design principles that prioritize vulnerability rather than undermine it. These principles are summarized in Table 2.

Table 2

Design principles for vulnerability-cultivating AI in HRD

Principle	Implementation guidance
Ethical Transparency	Enable user control, disclose data use, publish accessible data policies
Inclusive Design	Include marginalized voices, conduct equity impact assessments
Relational Infrastructure	Ensure access to human coaches, pair AI with human facilitators
Accountability Frameworks	Create appeal mechanisms, establish human-in-the-loop requirements

Source: Authors' own illustration

This aligns with recent evidence-based HR analytics research showing that ethical governance, including privacy-by-design, role-based data access, and algorithmic fairness audits, functions most effectively not as optional post-implementation additions, but as mandatory operational constraints embedded from the earliest design stage (Boiaryn, 2026).

Our findings suggest three key considerations for organizational leaders and governance bodies:

- Advanced AI cannot foster vulnerability in an environment where the culture is punitive or unsupportive. Leaders need to ensure that psychological safety and trust are firmly established before introducing AI tools into developmental processes (Edmondson, 2018).
- As AI becomes more independent in its functions, organizations must define clear lines of responsibility. Who is answerable when AI interventions fail to promote vulnerability or cause harm?

- Experiences of vulnerability differed among participants, highlighting the need for inclusive design. Involving diverse stakeholders ensure that tools support all employees.

4.7. Social justice implications

Vulnerability is unevenly experienced. Participants who held more organizational influence, enjoyed job security, or had positive prior developmental experiences were more open to engaging vulnerably with AI tools. This pattern suggests that AI-mediated HRD could unintentionally reinforce existing inequalities (Lima & Rahman, 2024). Kaur et al. (2025) directly speak of the vulnerability and marginalization resulting from absence of proper social justice.

The surveillance effects we observed have important implications for equity-seeking groups. Prior research indicates that AI monitoring disproportionately affects women, racialized employees, and other marginalized populations (Kellogg et al., 2020).

This concern resonates with broader debates in the AI-labour literature, where scholars caution that AI deployed primarily for surveillance or opaque decision-making risks becoming, in effect, an "invisible dictator" over workers, intensifying frictions for already vulnerable groups unless governance choices explicitly safeguard human dignity and accountability (Kuzior et al., 2025).

Participants' experiences of algorithmic judgment highlight concerns around labeling and scoring. AI systems that categorize or rate employees can have long-term consequences for career progression. Ethical frameworks for AI in HRD should embed mechanisms for accountability, contestation, and remediation.

4.8. Limitations and future research

This study has several limitations that highlight opportunities for future research.

Although our participants represented diverse sectors and roles, they were all based in Canada. Future studies should examine vulnerability in AI-mediated HRD across different cultural contexts.

Our findings reflect experiences at a single point in time. Longitudinal research is needed.

Future research could complement this approach with observational methods.

While participants used a range of AI tools, our study did not systematically compare the effects of different tool types. Future research should investigate how various AI architecture differentially shape experiences of vulnerability.

Future studies should test these principles through intervention research that manipulates features such as transparency, reflective pauses, and human facilitation, and evaluates their impact on vulnerability, learning outcomes, and equity.

5. CONCLUSION

The aim of this study was to explore how vulnerability operates in AI-mediated coaching and developmental contexts. Our findings reveal a nuanced reality: AI systems, by design, introduce monitoring, algorithmic judgment, and the absence of relational cues, all of which can suppress vulnerability. Yet when thoughtfully designed with human facilitation, reflective pauses, and ethical transparency AI can become part of a developmental ecosystem that strengthens, rather than diminishes, vulnerability. As AI increasingly shapes how we learn, develop, and lead, HRD has an ethical responsibility to ensure that technology serves human growth. We have argued that vulnerability is a source of strength rather than weakness. Perhaps the most critical mission for HRD in the AI era is to cultivate vulnerability.

ACKNOWLEDGEMENT

Authors acknowledge the support of SSBRC in the research process. This research was funded by Scientific Societal & Behavioral Research Center (SSBRC), (SSBRC; 2026). Ethical approval was obtained from the lead author's institutional Research Ethics Board (Protocol #2026-08EP) prior to data collection.

REFERENCES

- Ardichvili, A. (2022). The impact of artificial intelligence on expertise development: Implications for HRD. *Advances in Developing Human Resources*, 24(2), 78-98.
- Ardichvili, A., & Harmon, B. (2022). Toward a new political economy of human resource development. *Human Resource Development Review*, 21(3), 352-366.
- Belás, J., Jr., Petráková, Z., Streimikis, J., & Kozová, K. (2024). Participative management style in SMEs: Influence of CSR factors. Empirical evidence from the Visegrad Group. *Journal of International Studies*, 17(3), 133-147. doi:10.14254/2071-8330.2024/17-3/7
- Boiaryn, T. (2026). HR analytics for evidence-based people management: A structured synthesis of methods, decision interfaces, and governance principles for sustainable organizations. *Economics, Management and Sustainability*, 11(1), 163-172. <https://doi.org/10.14254/jems.2026.11-1.9>
- Boyatzis, R. E. (2025) The Democratization of Coaching and Development Through Peer Coaching in Groups, *Engaged Management Review*, 8(2), Article 2.
- Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE Publications.
- Cunha, M. P. e, Simpson, A. V., Clegg, S. R., & Rego, A. (2019). Speak! Paradoxical effects of a managerial culture of “speaking up”. *British Journal of Management*, 30(4), 829-846.
- Dhakal, K., Baczar, P., Devkota, N., Paudel, U. R., & Szabó, S. (2023). Institutional perception of e-recruitment adoption in Nepal: Evidence from structural equation modeling. *Economics and Sociology*, 16(3), 227-248. doi:10.14254/2071-789X.2023/16-3/12
- Diller, S. J., Stenzel, L. C., & Passmore, J. (2024). The coach bots are coming: exploring global coaches' attitudes and responses to the threat of AI coaching. *Human Resource Development International*, 27(4), 597-621. <https://doi.org/10.1080/13678868.2024.2375934>
- Edmondson, A. C. (1999). Psychological safety and learning behaviour in work teams. *Administrative Science Quarterly*, 44(2), 350-383.
- Edmondson, A. C. (2018). *The fearless organization: Creating psychological safety in the workplace for learning, innovation, and growth*. John Wiley & Sons.
- Edmondson, A. C., & Bransby, D. P. (2023). Psychological safety comes of age: Observed themes and future directions. *Annual Review of Organizational Psychology and Organizational Behavior*, 10, 123-148.
- Faizan, R., Haque, A.U., Cockrill, A., & Aston, J. (2019). Females at Strategic Level affecting Logistics Firms' Competitiveness: Qualitative Comparative Analysis. *Forum Scientiae Oeconomia*, 7(1), 57-71. https://doi.org/10.23762/FSO_VOL7_NO1_4
- Ghosh, R., & Lee, Y. (2024). Human-centred HRD in the digital era: From efficiency to empathy. *Human Resource Development International*, 27(3), 211-228.
- Hannah, S. T., Uhl-Bien, M., Avolio, B. J., & Cavarretta, F. L. (2011). A framework for examining leadership in extreme contexts. *The Leadership Quarterly*, 20(6), 897-919.
- Haque, A. U., & Yamoah, F. A. (2021). The Role of Ethical Leadership in Managing Occupational Stress to Promote Innovative Work Behaviour: A Cross-Cultural Management Perspective. *Sustainability*, 13(17), 9608, <https://doi.org/10.3390/su13179608>

- Haque, A. U. Yamoah, F. A., & Nweke, U. (2025). The stress-innovation link: leadership and strategies of female entrepreneurs in diverse economies. *Strategy & Leadership*, 53(5), 504-520. <https://doi.org/10.1108/SL-12-2024-0144>
- Hernik, J., Sagan, A., Jarecki, W., & Grinberga-Zalite, G. (2025). Digital transformation in business process management: The role of employee engagement. *Human Technology*, 21(1), 203–221. <https://doi.org/10.14254/1795-6889.2025.21-1.10>
- Kahn, W. A. (2018). The vulnerable leader. *Harvard Business Review*, 96(5), 98-105.
- Kaur, H., Faizan, R., & Haque, A. U. (2025). From Margins to Markets: Exploring Barriers and Pathways for Transgender Entrepreneurs. In A. U. Haque (Ed.), *Entrepreneurship: A contemporary perspective* (pp. 133-139). Routledge. <https://doi.org/10.4324/9781003504351-11>
- Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, 14(1), 366-410.
- Kim, B.-J., & Kim, M.-J. (2024). How artificial intelligence-induced job insecurity shapes knowledge dynamics: The mitigating role of artificial intelligence self-efficacy. *Journal of Innovation & Knowledge*, 9(4), 100590.
- Knezevic, D. (2023). The impact of technologies on leadership styles in Southeast European counties. *Journal of International Studies*, 16(2), 127-139. doi:10.14254/2071-8330.2023/16-2/8
- Kuzior, A., Marszałek-Kotzur, I., Sanszybayeva, K. N., & Lukács, E. (2025). From AI vibrancy to labour market outcomes: Testing displacement across education groups. *Economics & Sociology*, 18(4), 131-159. <https://doi.org/10.14254/2071789X.2025/18-4/7>
- Lemoine, G. J., Hartnell, C. A., & Leroy, H. (2019). Taking stock of moral approaches to leadership: An integrative review of ethical, authentic, and servant leadership. *Academy of Management Annals*, 13(1), 148–187.
- Li, H. (2024). Developing AI literacy in HRD: Competencies, approaches, and implications. *Human Resource Development International*, 27(4), 312-329.
- Lima, S. A., & Rahman, M. M. (2024). Algorithmic fairness in HRM: Balancing AI-driven decision making with inclusive workforce practices. *Journal of Information Systems Engineering & Management*, 9(4s). <https://doi.org/10.52783/jisem.v9i4s.13251>
- Maak, T., & Pless, N. M. (2006). Responsible leadership in a stakeholder society: A relational perspective. *Journal of Business Ethics*, 66(1), 99-115.
- Mendy, J., Jain, A., & Thomas, A. (2025). Artificial intelligence in the workplace—challenges, opportunities and HRM framework: A critical review and research agenda for change. *Journal of Managerial Psychology*, 40(5), 517-538.
- Mishchuk, H., Oliynyk, O., Bilan, Y., & Skare, M. (2025). The Role of Leadership in Attracting Intellectual Migrants and Boosting Innovative Development of Countries. *Contemporary Economics*, 19(3), 339-354. DOI:10.5709/ce.1897-9254.570
- Petriglieri, G., & Petriglieri, J. L. (2020). The return of the oppressed: A systems psychodynamic approach to organization studies. *Academy of Management Annals*, 14(1), 411–449.
- Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, 46(1), 192-210.
- Rickert, A., Frau, D., & Deuss, Y. (2024, September). The age of datafication: Balancing employee-leader vulnerability, trust, and responsibility in datafied pay for performance systems. *Conference Proceeding 38th British Academy of Management (BAM) Conference, Nottingham, United Kingdom*. British Academy of Management. <https://digitalcollection.zhaw.ch/handle/11475/32620>

- Schinello, S. (2025). Challenges and opportunities in the use of artificial intelligence in creative economy: Insights from expert interviews. *Economics and Sociology*, 18(1), 199-216. <https://doi.org/10.14254/2071-789X.2025/18-1/10>
- Sepehr, S., Voola, R., Ray, S., & Waller, D. S. (2025). Observing the “observed vulnerability”: A critical discourse analysis. *Journal of Macromarketing*. <https://doi.org/10.1177/02761467251366570>
- Turkle, S. (2021). *The empathy diaries: A memoir*. Penguin Press.
- Vaijayanthi, P., & Dhanushree, B. S. (2026). Reimagining human resource management in the age of technological disruption: A digital HR framework. *International Journal of Research in Engineering, Science and Management*, 9(4), 23-26.
- Wajcman, J. (2016). *Pressed for time: The acceleration of life in digital capitalism*. University of Chicago Press.
- Yoo, S., Nimon, K., & Patole, S. R. (2024). Artificial intelligence in human resource development: An umbrella review protocol. *PloS one*, 19(9), e0310125.
- Zsigmond, T., & Mura, L. (2023). Emotional intelligence and knowledge sharing as key factors in business management – evidence from Slovak SMEs. *Economics and Sociology*, 16(2), 248-264. <https://doi.org/10.14254/2071-789X.2023/16-2/15>