SA Journal of Industrial Psychology
ISSN: (Online) 2071-0763, (Print) 0258-5200

Artificial intelligence and employee assistance programmes in South Africa: A systematic review of ethical, cultural and digital transformation requirements for workforce well-being

Check for updates

Author:
Nthabeleng I. Mdhluli[1]

Affiliation:
[1]Department of Industrial and Organisational Psychology, College of Economic and Management Sciences, University of South Africa, Pretoria, South Africa

Corresponding author:
Nthabeleng Mdhluli,
mdhluni@unisa.ac.za

Dates:
Received: 24 July 2025
Accepted: 28 Jan. 2026
Published: 22 Apr. 2026

How to cite this article:
Mdhluli, N.I. (2026). Artificial intelligence and employee assistance programmes in South Africa: A systematic review of ethical, cultural and digital transformation requirements for workforce well-being. *SA Journal of Industrial Psychology/SA Tydskrif vir Bedryfsielkunde, 52*(0), a2354. https://doi.org/10.4102/sajip.v52i0.2354

Copyright:
© 2026. The Authors. Licensee: AOSIS. This work is licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0) license (https://creativecommons.org/licenses/by/4.0/).

Read online:
Scan this QR code with your smart phone or mobile device to read online.

Orientation: Artificial intelligence is increasingly transforming employee assistance programmes (EAPs) by enabling proactive, data-driven and context-sensitive approaches to employee well-being. In South African context, however, socio-economic inequalities, pronounced digital divides and cultural and linguistic diversity continue to constrain the adoption of artificial intelligence (AI)-enhanced EAPs, leaving ethical, cultural and structural considerations insufficiently examined.

Research purpose: To address this gap, this study employed a systematic review of peer-reviewed literature from 2012 to 2024 to examine and synthesise the ethical, cultural and digital transformation requirements influencing the adoption of AI-driven EAPs in South African workplaces.

Motivation for the study: The integration of AI in EAPs has not been fully explored in the South African context, where challenges of diversity, ethics and digital equity remain prominent.

Research approach/design and method: In accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 guidelines, a review of the Scopus, Web of Science, PubMed and EBSCOhost databases resulted in the identification of 50 relevant articles, which were analysed thematically.

Main findings: The synthesis produced six interconnected themes: (1) adoption and implementation dynamics (2) infrastructure and digital competency barriers (3) cultural and linguistic appropriateness (4) ethical governance and data privacy (5) economic and structural constraints and (6) employee well-being outcomes.

Practical/managerial implications: To unlock the potential of AI in EAPs, it is essential to create strategies that are culturally responsive, ethically governed and technologically inclusive.

Contribution/value-add: This review offers insights to assist policymakers, Human Resource (HR) professionals and EAP designers in developing fair and contextually relevant digital well-being interventions within South African workplaces.

Keywords: artificial intelligence; cultural diversity; digital transformation; employee assistance programmes; ethics in technology; leadership; well-being.

Introduction

The integration of artificial intelligence (AI) with employee assistance programmes (EAPs) marks a major development of workplace well-being. Traditionally designed as reactive, counsellor-led services, EAPs are now being transformed through AI platforms that offer adaptable and context-sensitive psychosocial support (Dutta & Mishra, 2024). This transformation is driven by mental health challenges and the structural shifts that followed the coronavirus disease 2019 (COVID-19) pandemic. Technologies such as machine learning, natural language processing and digital conversational agents enable real-time, data-driven engagement with employees, redefining the role of EAPs within the broader digital health ecosystem (Dutta & Mishra, 2024).

Despite advances in high-income countries, where AI adoption is supported by strong digital infrastructures and strict data governance frameworks (Aruleba & Jere, 2022), empirical evidence on AI implementation in African organisational contexts remains uneven and sector specific. Existing studies are predominantly concentrated in public-facing and digitally prioritised sectors, such as health (e.g. diagnostic support, telehealth, and public health surveillance), education (e.g. learning analytics and adaptive platforms) and financial services (e.g. credit scoring, fraud detection and customer service automation), where AI adoption has been driven by national digital strategies and service delivery imperatives (Aruleba & Jere, 2022; Bushe, 2019). In contrast, the application of AI within workplace wellbeing systems, particularly EAPs remain limited and underexplored. Within South Africa, this gap is amplified by a complex socio-technical landscape characterised by inequality, digital divides, linguistic diversity and persistent stigma around mental health (Aruleba & Jere, 2022; Bushe, 2019). These contextual conditions complicate the ethical design and equitable deployment of AI-enhanced EAPs, raising critical concerns related to digital exclusion, institutional trust, culturally situated understandings of psychological support, algorithmic transparency and data sovereignty (Hussey, 2012; Matsepe & Van der Lingen, 2022).

The limited empirical attention to AI-driven EAPs in South Africa, compared to more established application of AI in sectors, such as healthcare, education and financial services, constrains theoretical advancement within Industrial and Organisational Psychology and digital ethics. It also constrains the development of inclusive well-being systems that are in line with national regulatory frameworks, including the *Protection of Personal Information Act* (*POPIA*). Moreover, global debates on algorithmic bias, ethical AI design and human oversight are rarely contextualised within South Africa's organisational realities or transformation agenda (Matsepe & Van der Lingen, 2022). To address these gaps, this study identifies and explains the ethical, cultural and digital-transformation requirements needed for the responsible use of AI-enabled EAPs in South African workplaces. These requirements refer to the contextual, governance and technological conditions that make AI-driven well-being trustworthy, inclusive and sustainable. While AI-augmented EAPs are recognised worldwide as cost-effective tools for promoting well-being, local research remains scarce and tends to focus on technological efficiency rather than employee support (Aruleba & Jere, 2022; Ledden, 2022).

The existing literature seldom examines how socio-cultural, ethical and infrastructural factors influence the adoption of digital well-being technologies in the Global South. As a result, concerns about algorithmic accountability, informed consent, cultural sensitivity and compliance with privacy legislation remain underexplored.

This review proposes a multidimensional understanding of workforce well-being in the digital era grounded in ethical governance, cultural inclusivity and organisational readiness. It provides empirically informed insights to guide local policy, organisational practice and future research within the South African context.

By synthesising existing literature across these dimensions, the review clarifies how contextual, ethical and structural considerations shape the design and adoption of AI-enabled EAPs. On this basis, the study is guided by the following research question: What are the ethical, cultural and digital transformation requirements for designing AI-enabled EAPs that promote trust, transparency and well-being among employees in South African workplaces? This question reflects insight drawn from the reviewed literature and emphases the need for digital well-being interventions that are culturally responsive, ethically governed and psychologically safe within diverse organisational settings.

Literature review

Ethical requirements for artificial intelligence-enabled employee assistance programmes

The integration of AI into EAPs offers transformative possibilities for workplace well-being but introduces complex socio-technical and ethical challenges. These challenges highlight the need for a human-centred design (HCD) approach that embeds ethical awareness, organisational culture and employee engagement at the core of digital innovation. Human-centred design ensures that technology not only fulfils functional requirements but also aligns with users' lived experiences, norms and values within their organisational environments (Bednar & Welch, 2020; Trist & Bamforth, 1951). The sensitivity of EAPs, particularly in the domains of mental health and well-being, necessitates trust, confidentiality and moral accountability. Ethical AI design therefore requires developers and organisations to prioritise empathy, fairness and discretion. The Unified Theory of Acceptance and Use of Technology (UTAUT) (Williams et al., 2015) strengthens this argument by demonstrating that, social influence and facilitating conditions shape employee trust and willingness to adopt AI tools. In South African workplaces, infrastructural inequality and varying levels of digital literacy remain decisive factors influencing both accessibility and technical implementation (Aruleba & Jere, 2022; Bushe, 2019).

Ethical implementation is not confined to technical safeguards, it also requires an organisational infrastructure that embeds ethical decision-making within digital transformation strategies. Leadership commitment, transparent communication and continuous employee training form essential components of this systemic ethical culture (Vial, 2021). The socio-technical systems perspective highlights the dynamic interplay between, human agency and organisational culture. Effective AI design must therefore integrate co-evolutionary process that include employees in decision-making, reflecting context specific realities (Bednar & Welch, 2020; Trist & Bamforth, 1951). Neglecting socio-cultural context and employee participation

often results in diminished trust and resistance to adoption. Orlikowski's (1999) concept of 'technology-in-practice' further illustrates how organisational routines, power structures and interpretive flexibility continually shape the development of AI systems. This insight aligns with emerging evidence indicating that employee perceptions, organisational narratives and leadership commitment critically influence the ethical success of AI-enabled well-being programmes (Kruger, 2024).

In South Africa, ethical design must also be culturally responsive. Cultural identity, language diversity and stigma surrounding mental health significantly impact both participation and trust (Dergaa et al., 2023). Odero et al. (2024) indicate that the incorporation of socio-cultural distinctions, such as indigenous languages and culturally rooted coping strategies, significantly improves user acceptance and fosters meaningful engagement in digital EAPs. This implementation aligns with extensive global studies that demonstrate how participatory design, marked by the active engagement of stakeholders during the development process, promotes trust, relevance and empowerment (Van den Broek et al., 2024). In this context, the African philosophical paradigm of Ubuntu presents a compelling ethical and relational perspective, emphasising collective agency, interconnectedness and communal care (Ndlovu-Gatsheni, 2018). Incorporating Ubuntu principles into the design of AI and its organisational implementation can mitigate the risks of technocentrism and dehumanisation by promoting inclusivity, empathy and ethical stewardship. Moreover, research studies emphasise the necessity of ethical AI frameworks and human-centred governance models to address the risks linked to bias, surveillance and privacy in AI applications within the workplace (Hussey, 2012; Matsepe & Van der Lingen, 2022). These frameworks promote transparency, explainability and participatory oversight mechanisms that enhance employee agency and organisational accountability. Sustaining these ethical standards demands alignment between governance frameworks and broader digital-transformation initiatives, ensuring that organisational systems, culture and technology evolve coherently. Consequently, effective AI-enabled EAPs must include these governance principles to maintain legitimacy and social license across various workplace settings. Taken, these ethical and cultural considerations establish the foundational conditions under which AI-enabled EAPs can be trusted and legitimised, thereby shaping how technological design choices, human factors and governance mechanisms interact in organisational well-being systems.

Technological innovation and human factors in organisational well-being

Building on these ethical and cultural foundations, research on technological innovation in organisational well-being increasingly highlight the interdependence between AI capabilities, human factors and organisational context. Current discussions surrounding AI-driven well-being solutions highlight the need of incorporating ethical, human-focused design principles to guarantee effectiveness, acceptance and sustainability in organisational settings. The philosophical and practical imperatives of responsible AI emphasise that AI systems should not only operate efficiently but also reflect human values such as empathy, fairness and discretion. The significance of these attributes is particularly apparent in mental health applications, where the subtleties of emotional intelligence and confidentiality are fundamental to developing user trust and achieving positive therapeutic outcomes (Van den Broek et al., 2024). Research demonstrates that hybrid models, which combine AI tools with human practitioners, increase enhanced engagement and clinical outcomes by effectively balancing flexibility with personalised care (Khalifa et al., 2024).

In multicultural and socio-politically intricate organisational settings, like those found in South African workplaces, the socio-historical factors affecting mental health and help-seeking behaviours significantly impact the acceptance and use of digital well-being technologies (Parry et al., 2023). These contexts require AI interventions that are not only technically sound but also culturally sensitive and integrated within supportive organisational cultures that promote psychological safety. Dergaa et al. (2023) research on organisational culture highlights the importance of psychological safety as a critical prerequisite for the adoption of innovation, allowing employees to interact with new technologies without the apprehension of stigma or retaliation. This claim is supported by recent empirical findings from Dutta and Mishra (2024), which illustrate that clear communication about AI's capabilities and limitations, paired with strong human fallback mechanisms, is essential for promoting employee trust in AI-enabled EAPs.

Digital transformation extends beyond the deployment of AI tools, it represents a systemic and organisational reconfiguration that influences strategy, leadership, workflows and employee experience (Vial, 2021).

Within this process, human factors such as trust, competence and psychological safety become decisive enablers of transformation success. Organisations that cultivate a learning-oriented culture and adaptive leadership are better positioned to sustain digital change while protecting employee well-being (Kane et al., 2021).

The design and implementation of AI well-being solutions greatly benefit from iterative co-design methodologies, which incorporate ongoing stakeholder feedback throughout the development process. This participatory approach reduces the risks associated with algorithmic bias, user disengagement and ethical violations by promoting shared ownership and reflexivity (Van den Broek et al., 2024). The systemic, socio-technical perspectives can be traced back to the foundational contributions of Trist and Bamforth (1951), who articulated the idea that technology is fundamentally intertwined with social and organisational subsystems.

The theory of contemporary socio-technical systems emphasises that sustainable innovation necessitates the alignment of technical artefacts with the social structures of organisations, which includes power dynamics, communication networks and cultural norms (Savaget et al., 2019). This alignment lies at the heart of digital transformation, where technical, cultural and structural subsystems evolve together to enhance organisational resilience and adaptability.

Numerous research studies clarify the intricate relationship between technological advancement and human elements in promoting organisational well-being. For instance, the incorporation of ethical AI governance frameworks that emphasise transparency, accountability and inclusiveness has demonstrated an ability to enhance employee perceptions of fairness and organisational justice (Hussey, 2012; Matsepe & Van der Lingen, 2022). This, in turn, contributes to improved well-being and reduces resistance to technological change. Insights gained from this research are particularly advantageous for the South African context, where the legacies of systemic inequalities and cultural diversity necessitate that AI interventions not only address digital divides but also actively promote equity and dignity.

Ultimately, the integration of artificial intelligence, human factors and organisational well-being necessitates a comprehensive set of evaluation metrics that go beyond traditional productivity measures to include psychological, social and ethical aspects. Embedding these frameworks within a broader digital-transformation strategy ensures that well-being remains central to organisational modernisation, reinforcing that technological progress must coincide with human development and ethical governance. These insights illustrate that technological innovation, ethical design and human experience are mutually reinforcing, demonstrating the need to examine transparency and employee engagement as central mechanisms through which AI-enabled EAPs are enacted in practice.

Trust, transparency and employee engagement in artificial intelligence-driven employee assistance programmes

Within this interconnected ethical, cultural and technological landscape, trust and transparency emerge as a crucial relational mechanism that mediates employee engagement within AI-driven EAPs. Trust serves as a crucial factor in the acceptance and ongoing use of AI-enabled interventions in workplace mental health systems. The seminal research conducted by Mayer et al. (1995) presents a comprehensive model of organisational trust, outlining three fundamental components: ability (competence), benevolence (intent) and integrity (adherence to principles). The aforementioned dimensions integrate effortlessly into the digital landscape, supporting essential elements of trust in AI systems, including algorithmic fairness, data privacy and operational. In the evolving landscape of AI-driven EAPs, perceived transparency stands out as a significant predictor of employee engagement, influencing how users understand and react to the complexities often linked with intricate algorithms (Rane et al., 2024). This relationship is consistent with existing literature indicating that factors that enhance trust have a direct impact on the adoption rates and effective use of digital mental health tools (Langlieb et al., 2021; Sutherland, 2020).

Within the South African context, the process of building trust is deeply embedded into a complex socio-political and historical landscape, marked by persistent legacies of institutional mistrust and systemic inequalities (Bazana, 2024). The legacies in question exacerbate the difficulties associated with implementing AI in sensitive areas like mental health, where concerns regarding privacy and scepticism about organisational motives are significantly intensified. Zidaru et al. (2021) emphasise the significance of culturally attuned mental health initiatives, noting that culturally responsive programmes not only promote psychological safety but are also essential for authentic engagement with AI systems. Psychological safety, characterised as the collective conviction that the workplace is conducive to interpersonal risk-taking (Edmondson, 1999), serves as a fundamental element in cultivating an atmosphere where employees are comfortable revealing vulnerabilities and utilising AI-enabled support.

Trust and transparency are not only ethical imperatives but also strategic components of digital transformation. Organisational change efforts that aim to embed AI into daily operations depend on open communication, participatory governance and leadership credibility (Kane et al., 2021). These factors influence the degree to which employees perceive transformation as inclusive rather that imposed, reinforcing engagement and long-term commitment to new digital systems. In this sense, digital transformation must be as a cultural and relation process, where trust functions as the social infrastructure enabling sustainable technological integration.

Algorithmic transparency and explainability serve as crucial tools to address 'algorithmic alienation', a state in which employees feel estranged and disempowered because of the perceived complexity and lack of control over AI decision-making processes (Engel, 2019). Transparency clarifies the workings of AI, enhancing the interpretability of system outputs and allowing employees to develop informed mental models regarding the technology's capabilities and limitations (Doshi-Velez & Kim, 2017). This cultivates a renewed social agreement between employees and organisations, articulated through Rousseau's (1995) interpretation of psychological contracts, implicit mutual expectations that regulate workplace dynamics. The perception of fairness and transparency in AI systems among employees enhances the psychological contract, thereby improving organisational trust and commitment to ethical practices (Kim et al., 2024). Furthermore, empirical evidence from numerous studies indicates that transparency is not a

singular concept but rather a multifaceted one (Engel, 2019; Hussey, 2012; Larsson & Heintz, 2020). It includes informational transparency (clarity regarding AI functions), procedural transparency (openness of processes) and interpersonal transparency (communication style and tone). Organisations that strategically implement these aspects promote increased employee autonomy and empowerment, which are essential for addressing resistance to AI integration (Kim et al., 2024). This is particularly crucial in EAPs, as trust affects both initial acceptance and continued engagement with recommended interventions (Langlieb et al., 2021).

The combination of cultural context, organisational culture and technological design necessitates a comprehensive and inclusive approach to AI-driven well-being solutions. Utilising principles of inclusive design and co-creation guarantees that AI tools align with the varied values and expectations of employees, thereby reducing the likelihood of reinforcing systemic biases or marginalisation (Hussey, 2012). In South Africa, this involves acknowledging and incorporating indigenous epistemologies and communal values like Ubuntu, which highlights relationality, dignity and collective well-being (Ndlovu-Gatsheni, 2018).

Incorporating these philosophies into transparency protocols and trust-building strategies fosters an ethical foundation that goes beyond just technical compliance, embracing essential humanistic organisational principles. Accordingly, trust and transparency function not only as ethical principles but as organisational capabilities that link technological design, cultural legitimacy and employee well-being outcomes within broader digital transformation processes.

Data privacy, consent and legal frameworks in South Africa

Trust dynamics are shaped by legal and regulatory frameworks that govern data use, consent and accountability in AI-enabled workplace interventions. The implementation of AI-enabled EAPs in South Africa requires strict compliance with the *POPIA* (Republic of South Africa, 2013), which establishes extensive data privacy protections in accordance with global data protection standards, including the European Union's General Data Protection Regulation (GDPR). *Protection of Personal Information Act* clearly mandates the lawful, fair and transparent processing of personal information, providing enhanced protections for sensitive health-related data, which includes mental health metrics gathered by AI-driven systems (Republic of South Africa, 2013). Adherence to *POPIA* is essential, serving as the foundational legal framework for the development of AI applications aimed at enhancing workplace well-being.

In addition to fulfilling legal requirements privacy concerns are a key factor influencing user trust and engagement in digital health platforms. This study highlights that the perceived vulnerabilities related to personal data usage, such as concerns about unauthorised access, data breaches and misuse, considerably hinder both the adoption rates and the continued use of AI-enabled mental health interventions. The concerns are particularly relevant in organisational environments marked by historical patterns of monitoring and distrust, where ethical data governance is central not only to legal compliance but also to organisational transformation. In the context of digital-transformation, establishing transparent and accountable data management practices signifies a shift in organisational culture, from reactive compliance to proactive ethical stewardship of information.

The implementation of AI-driven predictive analytics in EAPs enhances these privacy concerns by incorporating advanced methods of profiling and the use of secondary data. These technologies possess the capability to identify sensitive behavioural and psychological characteristics, which may extend beyond the initially agreed-upon parameters of consent, consequently increasing ethical and legal concerns (Donnelly, 2022). The identified risks necessitate the implementation of proactive governance frameworks that integrate technical safety features, including differential privacy, federated learning and secure multiparty computation.

Additionally, digital transformation in this domain requires organisations to institutionalise ethical data governance as a continuous, system process supported by cross functional teams, policy audits and digital literacy initiatives. This ensures that AI integration an opportunity to build ethical infrastructure rather than an exercise in technological expansion.

Sebunya and Gichuki (2024) present a convincing argument for participatory governance models that incorporate transparency, digital literacy and cultural responsiveness into data practices. In the diverse South African workforce, characterised by linguistic variety, cultural multiplicity and differing socio-economic backgrounds, participatory models are crucial to mitigate exclusionary impacts and to enable employees to become active participants in data governance instead of remaining passive data subjects. This significant transformation places employees at the forefront of data ethics, enhancing organisational accountability and promoting democratic oversight of AI-enabled systems.

Dynamic consent models are emerging as a best practice within this ethical landscape, facilitating ongoing, detailed and user-focused control over personal data sharing (Prictor et al., 2020). In contrast to static, one-time consent, dynamic consent mechanisms enable employees to engage in ongoing negotiations regarding the scope and extent of data disclosure, adjust their preferences in real time, and obtain clear feedback on the utilisation of their data. This method aligns with the constitutional principles of South Africa that emphasise individual dignity, autonomy and the right to control personal information (Republic of South Africa, 1996). Dynamic consent supports ethical AI lifecycles by embedding respect for employee agency, promotion trust

and mitigating power asymmetries inherent in employer-employee data relationships (Prictor et al., 2020). Moreover, the incorporation of these legal and ethical frameworks requires collaborative efforts across various disciplines, including technologists, legal experts, organisational psychologists and ethical scholars. This collaboration guarantees that AI-enabled EAPs are developed and managed within a socio-legal framework that is attuned to local contexts, regulatory requirements and advancing technological capabilities. This comprehensive governance framework not only ensures compliance but also encourages a resilient organisational culture in which data privacy is embraced as a core value rather than an afterthought in regulation.

Comprehensive approaches to ethical artificial intelligence in employee support systems

Integrating ethical design, human factors, trust mechanisms and legal compliance necessitates comprehensive AI governance frameworks that align technological innovation with organisational transformation and employee well-being. The discussion surrounding ethical AI in employee support systems increasingly acknowledges the necessity of comprehensive frameworks that go beyond simple technical aspects to incorporate ethical, psychological and socio-cultural factors in the design and implementation of AI.

Prominent frameworks like the Institute of Electrical and Electronics Engineers' Ethically Aligned Design as stated in Barrows et al. (2019), outline essential principles such as human well-being, accountability, transparency and contextual responsiveness. These principles require careful adaptation to the unique socio-economic environment and persistent historical inequities present in South Africa. This adaptation requires an intricate understanding that AI interventions in organisational well-being are intrinsically connected to wider socio-political contexts, where systemic inequality and resource disparities influence both access and outcomes.

The incorporation of Afrocentric epistemologies provides a transformative ontological and ethical basis that enhances these global frameworks by prioritising relationality, communal values and social justice as fundamental elements of AI ethics (Ndlovu-Gatsheni, 2018; Ngcece & Mkhize, 2023). Afrocentricity highlights the principle of Ubuntu, which embodies an ethic of shared humanity and interconnectedness (Ndlovu-Gatsheni, 2018). This perspective enhances AI governance by fostering collective agency and accountability, aligning with digital transformation principles that emphasise participatory change and socially embedded innovation rather than isolated technological deployment. This method emphasises the cultural consistency of AI-driven EAPs, ensuring they genuinely connect within African organisational environments and promote inclusivity that goes beyond basic compliance. Technical methodologies such as Explainable AI (Doshi-Velez & Kim, 2017) and participatory co-design (Van den Broek et al., 2024) implement these ethical imperatives by improving system transparency, interpretability and stakeholder engagement. The ability of Explainable Artificial Intelligence (XAI) to make AI decision-making processes understandable is crucial for building trust and addressing the 'black-box' issue, especially in sensitive areas such as mental health, where unclear algorithms can undermine psychological safety. Participatory co-design enhances the decentralisation of AI development by integrating a variety of employee perspectives and lived experiences into iterative design processes, thus guaranteeing contextual relevance and cultural sensitivity (Van den Broek et al., 2024). These approaches are consistent with modern socio-technical systems theory, which highlights the interrelationship between technological artefacts and organisational social structures (Appelbaum, 1997). This interrelationship reinforces that digital transformation is not a purely technological act but a process of socio-organisational restructuring that demands adaptive leadership, ongoing learning and ethical foresight.

The establishment of ethics within organisations is essential for the continuation of these integrative practices. Ngcece and Mkhize (2023) are among researchers who propose the formation of specialised ethics committees, ongoing ethical impact assessments, and comprehensive AI governance frameworks that tie together productivity demands with transformation objectives and psychological safety. Such structures represent the organisational backbone of digital transformation, institutionalising ethical reasoning and shared accountability as core features of strategic change. These governance mechanisms help organisations anticipate and manage ethical dilemmas, power imbalances and unintended consequences that may arise from AI adoption.

Research design

This study employs a systematic literature review methodology, carefully adhering to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines (Page et al., 2021). The systematic literature review was selected to guarantee a transparent, replicable and thorough synthesis of the existing body of knowledge regarding AI-enhanced EAPs, specifically emphasising their applicability and ethical considerations within the South African socio-economic and cultural landscape. This methodology is in accordance with the rigorous interdisciplinary inquiry that connects technical innovation with societal impact.

A comprehensive literature search was performed across six prominent academic databases to guarantee thorough multidisciplinary coverage. The sources included Scopus, Web of Science, EBSCOhost, ScienceDirect, PubMed and Sabinet, with the latter specifically included to capture South African and African scholarship. The search period extended from 2012 to 2024. The year 2012 was selected as a theoretical and empirical starting point, as it coincides with the resurgence of machine learning. The starting point was selected as the period marked a shift from the rule-based

systems to data-driven and predictive applications, enabling the subsequent development of conversational agents, predictive analytics and adaptive digital health tools (Krizhevsky et al., 2012).

A systematic and constant search strategy was established utilising a blend of controlled vocabulary and free-text keywords. Boolean operators ('AND', 'OR', 'NOT') were utilised strategically to enhance both the sensitivity and specificity of search results. Essential terms such as 'artificial intelligence', 'machine learning', 'chatbots', 'predictive analytics', 'employee assistance programme', 'workplace well-being' and 'occupational mental health' were enhanced with contextual modifiers including 'South Africa', 'African workplace' and 'cultural relevance'. This method effectively facilitated the collection of relevant literature while reducing the inclusion of irrelevant content.

The inclusion criteria required that studies be peer-reviewed, empirical and theoretical, published in English, and specifically focused on AI applications within EAPs or workplace mental health interventions, particularly in relation to South African workplaces or similar organisational contexts. Non-peer-reviewed was considered where it offered essential regulatory or contextual insight, such as government or institutional reports. Quality thresholds were established using the Mixed Methods Appraisal Tool (MMAT, 2018) for quantitative and mixed-method studies, and Guba and Lincoln's (1994) trustworthiness framework for qualitative research.

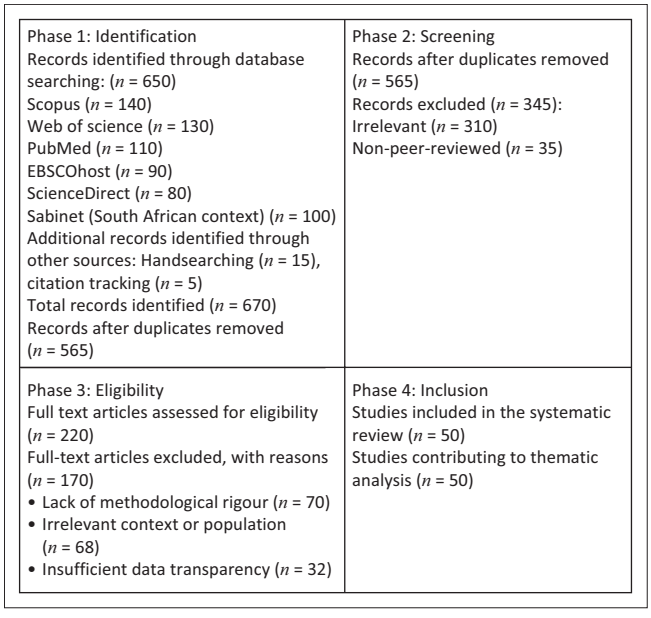
Studies scoring less than 50% on the MMAT criteria or demonstrating low credibility, dependability or transferability were excluded. Studies with minor methodological limitations were retained but explicitly qualified during synthesis to preserve contextual breath while maintaining analytical integrity. Of the 220 full-text articles screened, 30 were excluded because of quality scores and 20 were included but qualified in the synthesis to highlight methodological limitations.

To ensure methodological rigour, the selection and screening processes followed a structured and auditable protocol. All phases of record management were carried out independently by the researcher, including identification, duplicate removal, title and abstract screening, and full-text eligibility review. Duplicate records were identified using automated tools within each database, and then manually verified to ensure accuracy. To increase transparency and reduce bias, an external academic peer with expertise in systematic review methods reviewed a random sample (20%) of screening and eligibility decisions. There were no conflicts or discrepancies discovered during verification, and the consistency of inclusion and exclusion criteria was confirmed. A detailed log of all screening decisions and reasons for exclusion was kept, resulting in a complete audit trail for reproducibility.

In accordance with PRISMA protocols by Page et al. (2021), the preliminary search resulted in a total of 670 records, which included 650 sourced from databases and 20 obtained through handsearching and citation tracking. Following the elimination of 105 duplicates, 565 distinct records advanced to the title and abstract screening phase, leading to the exclusion of 345 articles, primarily because of irrelevance or absence of peer review. A comprehensive evaluation of 220 articles was conducted, applying the inclusion and exclusion criteria rigorously, resulting in a final selection of 50 studies considered appropriate for detailed analysis. The selection process is thoroughly documented in the PRISMA flow diagram (Figure 1), emphasising the methodological rigour and reproducibility of this review.

A detailed quality appraisal was conducted to ensure methodological rigour and reliability of the included studies. The MMAT was systematically applied to all qualitative, quantitative and mixed-methods studies (Hong et al., 2018). Each study was evaluated against the five MMAT criteria relevant to its design, including clarity of research questions, adequacy of data collection methods, appropriateness of analysis, interpretation of results and coherence between data and inclusions. Each criterion was rated as 'Yes', 'No' or 'Can't tell', and a composite quality score was subsequently assigned.

For qualitative studies, the Guba and Lincoln (1994) criteria were used as a benchmark to evaluate reported methodological quality. This involved examining whether each study demonstrated evidence of credibility (e.g. triangulation and member validation), dependability (clear and auditable research process), confirmability (reflexivity and neutrality) and transferability (rich contextual description). These indicators were not applied directly by the reviewer but assessed as reported in each study.



Source: Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

FIGURE 1: Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of the systematic literature review process.

In order to reduce the possibility of bias, the processes of study screening and data extraction were conducted independently by the researcher. Where uncertainty arose in the appraisal process, consultation was undertaken with an external methodological advisor to confirm consistent interpretation of criteria. Studies that failed to meet three or more MMAT criteria or showed low methodological transparency against trustworthiness indicators were excluded. Those with minor methodological limitations were retained and their methodological constraints clearly acknowledged during the synthesis to preserve the breadth and contextual richness of the evidence base.

The data extraction process adhered to a standardised protocol that carefully captured essential study characteristics. This included details such as authorship, publication year, research design, sample demographics, types of AI technology utilised and various outcome measures, including efficacy, engagement, ethical considerations and cultural contextualisation. All extracted data were compiled into a structured data matrix in Microsoft Excel to ensure traceability and consistency throughout the review. The qualitative data underwent coding and thematic analysis through the use of NVivo 12 software, applying both inductive and deductive coding approaches. The analytical process followed Braun and Clarke's (2006) six-phase thematic analysis, comprising familiarisation with the data, initial coding, theme development, theme review, theme definition and reporting. Within this framework, line-by-line coding and the progression from descriptive to higher-order analytical themes were informed by thematic synthesis as proposed by Thomas and Harden (2008). Both inductive and deductive coding strategies were employed to ensure that themes reflected both the empirical evidence and the theoretical perspectives underpinning ethical AI, cultural responsiveness and organisational psychology. The process was iterative and reflexive, allowing for continuous comparison and refinement of themes until theoretical saturation was achieved. This approach highlighted the interaction between technological innovation and social justice, diversity and power dynamics within South African workplaces, thereby aligning with the focus on socio-technical systems and societal transformation.

Ethical reflexivity was diligently upheld, acknowledging the secondary nature of the data. The researcher engaged in a thorough examination of positionality and interpretative actions to maintain objectivity and cultural sensitivity. The review adhered completely to South Africa's POPIA and associated ethical regulations, ensuring thorough attention to data privacy, informed consent and equity concerns related to AI implementation.

The limitations associated with systematic literature reviews are recognised, including reliance on the availability and quality of primary research, the potential for publication bias, and constraints related to language, specifically English publications. The incorporation of reliable grey literature has slightly minimised these concerns. The systematic management of heterogeneity in study designs and outcomes was achieved through a rigorous quality appraisal process and a thoughtfully developed thematic framework, thereby ensuring that the findings are both valid and contextually relevant.

Ethical considerations

Ethical clearance to conduct this study was obtained from the University of South Africa College of Economic and Management Sciences ERC Industrial and Organisational Psychology (No. 8589).

Results

A total of 50 peer-reviewed studies published between 2012 and 2024 were analysed through a thematic analysis.

Both deductive coding, guided by the review question on ethical, cultural and digital transformation requirements, and inductive analysis of emergent concepts were employed. Codes were initially grouped into 24 preliminary categories and then synthesised into six overarching themes to capture recurrent patterns across organisational, ethical, and socio-technical contexts. Thematic development was verified through constant comparison and re-examination of coding memos to ensure analytical consistency.

Organisational readiness and leadership commitment

Across most studies, AI-EAP success depends on strategic readiness and ethical stewardship at leadership level. Large, digitally mature organisations demonstrated structured implementation processes, while small and medium enterprises (SMEs) faced infrastructure and governance constraints (Bushe, 2019; Kruger, 2024; Lange et al., 2022). Evidence from Booyse and Scheepers (2024) showed that limited executive awareness of data-ethics obligations hindered compliance with *POPIA*. Together these findings illustrate that organisational readiness represents an ethical prerequisite for digital transformation.

Infrastructure and workforce digital competency

A consistent pattern revealed across the reviewed studies was that digital divides between urban and rural workplaces and generational cohorts limit equitable adaptation (Aruleba & Jere, 2022; Lange et al., 2022). Low digital literacy among older employees and the absence of ongoing digital-skills training restricted sustained engagement.

Insufficient investment in digital training by organisations further restricts sustainability (Zou et al., 2020).

Studies describing pilot programmes with continuous learning interventions found that AI-EAP platforms were more widely accepted (Dergaa et al., 2023; Ledden, 2022). As a result, digital transformation must be viewed as an organisational change process that necessitates capability development rather than simply acquiring new technology.

Appropriateness in culture and language

Evidence demonstrates that cultural dissonance undermines employee trust in AI-mediated care. Findings indicate that employees found automated counselling impersonal and detached from communal norms (Langlieb et al., 2021; Sutherland, 2020). Inclusion of indigenous languages and Ubuntu-based relational principles enhanced user identification and perceived authenticity, but exclusivity marginalises non-English speakers (Dergaa et al., 2023; Odero et al., 2024). This theme highlights that culturally grounded design constitutes a foundational requirement for AI-EAP legitimacy in South Africa.

Ethical governance and data integrity

Findings highlighted a weak governance mechanism surrounding data privacy and algorithmic fairness. Frequent concerns included poor consent procedures and opaque decision-making (Donnelly, 2022; Ngcece & Mkhize, 2023). Algorithmic prejudice disproportionately impacts marginalised groups (Hussey, 2012), and insufficient transparency in AI decision-making raises ethical challenges (Donnelly, 2022). The existing regulatory oversight is still insufficient (Dutta & Mishra, 2024; Folorunso et al., 2024).

Economic and structural constraints

Cost intensity and unequal infrastructure investment were recurring constraints. Small and medium enterprises cited unsustainable licencing costs and inadequate connectivity (Bushe, 2019), Broader structural inequities, such as socioeconomic disparities and limited policy incentives further exacerbate digital exclusion (Aruleba & Jere, 2022; Folorunso et al., 2024). These findings position financial and structural barriers as critical inhibitors to equitable digital transformation.

Employee well-being and psychological safety

Across studies, AI-EAPs enhanced early help-seeking and stress literacy (Ferrari et al., 2022; Kim & Lee, 2024) yet participants also expressed scepticism towards algorithmic empathy and loss of human connection. Self-efficacy and stress literacy have improved (Kim & Lee, 2024), but perceptions of AI inauthenticity (Fiske et al., 2019), reduced team cohesion (Langlieb et al., 2021), and unequal access persist. Programmes that combined human follow-up with AI triage yielded stronger perceptions of care quality and authenticity. These results indicate that psychological safety and trust depend on balancing technological efficiency with human presence.

Discussion

This section interprets the six thematic findings in relation to the research goal of exploring the ethical, cultural and digital transformation requirements for the implementation of AI-driven EAPs in South African workplaces. Drawing on socio-technical, organisational and Afrocentric theoretical frameworks, the discussion links the empirical synthesis to larger debates about equity, inclusivity and responsible digital transformation.

Aligning technological potential with industry realities

The findings revealed that organisational readiness and leadership commitment are stronger determinants of AI-EAP success rather than mere availability of technology. This highlights that digital transformation is not a technical event but an organisational and behavioural process (Matsepe & Van der Lingen, 2022). South African workplaces with visionary leadership, participatory governance and adaptive change management strategies demonstrated greater success in integrating AI-EAPs (Kruger, 2024). This aligns with the socio-technical theory, which emphasises the co-evolution of technology and human systems. The evidence also extends Bushe's (2019) argument that leadership in African organisations must prioritise ethical innovation and inclusivity to avoid reinforcing workplace inequalities. Thus, sustainable digital transformation requires cultivating a culture of trust, psychological safety and shared accountability.

Infrastructural inequity and digital competence

Ongoing disparities in internet access between urban and rural areas, as well as differing degrees of digital literacy, are important barriers to attaining equitable access for all. The insufficient infrastructure highlighted by restricted data access, unreliable connectivity and poor device availability demonstrates broader geographical disparities that exist within South African society (Aruleba & Jere, 2022; Bushe, 2019). From a systemic standpoint, digital transformation cannot progress without foundational infrastructure and capacity development. Organisations that invest in workforce upskilling reported improved engagement with AI-enabled support platforms (Zou et al., 2020). These findings support international evidence that digital literacy enhances employee autonomy and technology acceptance (Kim & Lee, 2024). Consequently, South African organisations must embed digital competency into human resource development and national policy frameworks to ensure that digital well-being tools reach all segments of the workforce.

Appropriate culture and language

The findings confirmed that many AI-EAPs fail to align with the cultural and linguistic realities of South African employees. The dominance of Western, English-language models limits accessibility and perceived authenticity (Langlieb et al., 2021; Sutherland, 2020). Incorporating Ubuntu as a philosophy of interconnectedness and collective care offers a way to reclaim cultural congruence in digital well-being innovation (Ndlovu-Gatsheni, 2018). This discussion builds on Odero et al.'s (2024) argument that culturally responsive and multilingual design increases user trust and engagement. Embedding indigenous epistemologies into AI-EAP frameworks shifts the conversation from inclusion to

epistemic justice, ensuring that digital transformation reflects African values rather than replicating technocentric paradigms.

Ethical governance and the emerging normativity of artificial intelligence in psychological care

The results indicate prominent gaps in regulations concerning the ethical implementation of AI-EAPs.

Although South Africa has established a strong data protection framework through *POPIA*, there are ongoing inconsistencies in its implementation, particularly around consent, algorithmic transparency and equitable design (Ngcece & Mkhize, 2023). This is consistent with broader critiques that legal compliance alone cannot replace ethical governance. The combination of Explainable AI (Doshi-Velez & Kim, 2017) and participatory oversight mechanisms could improve transparency and accountability. Furthermore, Ngcece and Mkhize's (2023) proposal for cross-disciplinary ethics committees within organisations can institutionalise ethical reflection as a core organisational practice rather than a peripheral requirement.

Economic and structural constraints

The ability of AI-EAPs to provide mental health support easily accessible is hindered by financial disparities and a lack of policy incentives (Folorunso et al., 2024). The significant costs associated with implementation and the limited support from the government hinder broad adoption, particularly in sectors that are already facing economic challenges (Sutherland, 2020). Moreover, the lack of innovation subsidies or frameworks for public-private partnerships indicates a gap between digital health policy and reality of the labour market.

In the absence of strategic investments and supportive policy frameworks, AI-EAPs may evolve into resources that primarily benefit the elite, thereby deepening the dual economy that defines South African society.

Employee well-being and psychological safety

The findings revealed that, while AI-EAPs improve early intervention and self-efficacy, they lack the emotional resonance required for long-term well-being. This is consistent with Fiske et al. (2019), who claim that algorithmic empathy cannot fully replicate human relational care. The discussion places this within the socio-technical systems framework, emphasising that well-being emerges from synergistic human-technology collaboration rather than automation alone. Hybrid approaches, which combine digital tools with professional counselling, show greater promise for improving trust and psychological safety (Langlieb et al., 2021).

However, unless these interventions are embedded in supportive organisational cultures that address workload, inclusion and job security, they risk remaining superficial. To achieve true well-being transformation, AI-EAPs must be integrated into holistic psychosocial ecosystems that are responsive to employees' lived experiences. These findings demonstrate that AI-EAPs' ethical, cultural and digital transformation requirements are mutually reinforcing. Ethical governance builds trust, cultural responsiveness increases engagement, and digital capacity ensures accessibility. They lay the groundwork for South Africa's digital well-being ecosystems to be long-lasting and inclusive.

This review advances the discussion of responsible AI by combining socio-technical systems theory, Afrocentric ethics and digital humanism. Practically, it provides a road map for leaders, policymakers and researchers looking to implement AI-powered employee support systems that are contextually grounded, ethically robust and socially equitable.

Recommendations and future research

Theoretical implication

This study significantly strengthens the theoretical framework of technology-mediated employee well-being by examining the interaction between AI innovation and culturally specific mental health paradigms in the Global South. By emphasising the South African context, it contests the universality of Western-centric models of EAPs and a call within decolonial and Afrocentric research to incorporate includes indigenous perspectives, such as Ubuntu, and relational ontologies into organisational theory (Ndlovu-Gatsheni, 2018; Ngcece & Mkhize, 2023). The results highlight that AI-enabled EAPs should be understood not only as technological tools but as socio-technical systems integrated within historically specific and culturally diverse contexts (Matsepe & Van der Lingen, 2022; Vial, 2021). This reframing encourages scholars in HRM, Information Systems, and AI ethics to rethink theoretical models of AI deployment through a decolonial perspective that recognises epistemic diversity, sociotechnical embeddedness and localised constructs of psychological support.

Practical implications

The findings provide a practical guide for human resource professionals, organisational psychologists and AI-EAP developers in the culturally informed design, implementation and expansion of digital mental health interventions. The study highlights the importance of developing interfaces that are both linguistically inclusive and culturally relevant, particularly within multilingual and collectivist communities. Localisation should include more than just translation; it additionally needs to incorporate idiomatic, contextual and emotional aspects of the user experience. In addition, focused initiatives aimed at enhancing digital literacy particularly for rural, older and blue-collar workers are crucial for promoting equitable participation. It is highly advisable to implement hybrid service models that combine AI capabilities with the oversight of human counsellors in order to maintain emotional authenticity and foster user trust. It is crucial to

establish organisational support structures, such as ongoing training, change management and psychological safety protocols, to guarantee the consistency of adoption and the potential for meaningful impact.

Policy implementation

The review highlights a significant gap in the governance of AI-driven psychosocial support tools within the context of South Africa's labour and health policy framework. To address ethical risks and systemic inequities, it is essential to establish a specialised regulatory framework. This involves establishing a statutory entity possibly under the Health Professions Council of South Africa (HPCSA) or in partnership with the Department of Employment and Labour tasked with the accreditation of AI-EAPs, promoting algorithmic transparency, ensuring adherence to the *POPIA*, and protecting mental health outcomes. Policy interventions should clearly focus on the digitalisation of psychosocial care in labour legislation, including consent protocols, employee rights and accountability for algorithms. It is recommended to implement incentivised public-private partnerships and subsidisation mechanisms to promote equitable deployment across various sectors, particularly for under-resourced SMEs.

Direction for future research

Future research should use longitudinal and mixed-method designs to assess the long-term sustainability, cultural acceptability and psychosocial outcomes of AI-EAPs. Empirical studies conducted over 1–3 years would provide useful information on how user perceptions, engagement and trust evolve over time.

Mixed-method approaches that combine quantitative indicators, such as usage rates and well-being outcomes, with qualitative accounts of user experience and cultural resonance will be required to fully comprehend these dynamics. Further research should look into the impact of language diversity, cultural identity and social stigma on the adoption and effectiveness of AI systems, particularly among historically marginalised or linguistically diverse employees. The development and validation of ethical frameworks that incorporate African philosophical perspectives, particularly Ubuntu, would provide locally based guidance for responsible AI design and deployment.

Future research should also look into sustainable funding models for AI-EAPs, with an emphasis on equitable financing mechanisms that ensure access across all economic sectors. Empirical analysis of public-private partnerships and tiered subscription models would help policymakers identify viable approaches to scaling AI-based well-being interventions while minimising inequality. Interdisciplinary collaboration remains critical for achieving these goals. Future research should bring together organisational psychologists, AI developers, ethicists, legal scholars and community mental health practitioners to co-create frameworks that balance technological innovation and human-centred values. Such collaboration would enhance the integration of ethical governance, cultural sensitivity and digital transformation into South Africa's overall well-being landscape.

Limitations

This review, while contributing valuable insights, is limited by various methodological and contextual constraints. The existing literature is notably lacking in longitudinal studies, which restricts our understanding of the long-term effectiveness of AI-EAP interventions and the evolution of user trust and therapeutic engagement over time. The effectiveness over time has not been adequately theorised or empirically examined. Furthermore, the deliberate omission of grey literature including industry reports, organisational evaluations and government white papers can lead to the exclusion of valuable, practice-based innovations and sector-specific insights that are frequently unavailable through conventional academic publishing avenues. This exclusion may limit the relevance of the findings to practical organisational settings.

An important limitation is the linguistic consistency of the data, characterised by a predominant dependence on publications in the English language. This creates a systemic bias that marginalises research conducted in indigenous South African languages, consequently neglecting culturally embedded knowledge and employee experiences that could greatly improve the adaptation of AI-EAP solutions.

Conclusion

This review provides a theoretically informed and contextually advanced analysis of AI-enhanced EAPs in South Africa, presenting critical insights into their transformative potential and inherent limitations. The success of these technologies in enhancing workplace mental health support depends on strong ethical frameworks, cultural and linguistic awareness, and equitable access. The findings confirm that AI-EAPs should not be viewed solely as technological solutions, but rather as culturally integrated support systems that align with various epistemologies and lived experiences. Future advancement depends on interdisciplinary collaboration, regulatory adaptability and strategic investment in digital capacity development. Only through comprehensive efforts can AI-driven mental health interventions develop into tools for psychosocial equity and organisational resilience throughout South Africa.

Acknowledgements

The author wishes to acknowledge the support of the University of South Africa in providing access to research resources and academic platforms that contributed to the development of this article. No external assistance or funding influenced the content or findings of this study.

Competing interests

The author declares that no financial or personal relationships inappropriately influenced the writing of this article.

CRediT authorship contribution

Nthabeleng I. Mdhluli: Conceptualisation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Writing – original draft, Writing – review & editing. The author confirms that this work is entirely their own, has reviewed the article, approved the final version for submission and publication, and takes full responsibility for the integrity of its findings.

Funding information

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Data availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Disclaimer

The views and opinions expressed in this article are those of the author and are the product of professional research. They do not necessarily reflect the official policy or position of any affiliated institution, funder, agency or that of the publisher. The author is responsible for this article's results, findings and content.

References

Appelbaum, S.H. (1997). Socio-technical systems theory: An intervention strategy for organisational development. *Management decision, 35*(6), 452–463. https://doi.org/10.1108/00251749710173823

Aruleba, K., & Jere, N. (2022). Exploring digital transforming challenges in rural areas of South Africa through a systematic review of empirical studies. *Scientific African, 16*, e01190. https://doi.org/10.1016/j.sciaf.2022.e01190

Barrows, C., Bloom, A., Ehlen, A., Ikäheimo, J., Jorgenson, J., Krishnamurthy, D., Lau, J., McBennett, B., O'Connell, M., Preston, E., Staid, A., Stephen, G., & Watson, J.-P. (2019). The IEEE reliability test system: A proposed 2019 update. *IEEE Transactions on Power Systems, 35*(1), 119–127. https://doi.org/10.1109/TPWRS.2019.2925557

Bazana, S. (2024). Exploring worker subjectivity: Shaping industrial and organisational psychology in post-apartheid South Africa. *South African Journal of Psychology, 54*(4), 567–580. https://doi.org/10.1177/00812463241288544

Bednar, P.M., & Welch, C. (2020). Socio-technical perspectives on smart working: Creating meaningful and sustainable systems. *Information Systems Frontiers, 22*(2), 281–298. https://doi.org/10.1007/s10796-019-09921-1

Booyse, D., & Scheepers, C.B. (2024). Barriers to adopting automated organisational decision-making through the use of artificial intelligence. *Management Research Review, 47*(1), 64–85. https://doi.org/10.1108/MRR-09-2021-0701

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bushe, B. (2019). The causes and impact of business failure among small to micro and medium enterprises in South Africa. *Africa's Public Service Delivery and Performance Review, 7*(1), 1–26. https://doi.org/10.4102/apsdpr.v7i1.210

Dergaa, I., Fekih-Romdhane, F., Glenn, J.M., Saifeddin Fessi, M., Chamari, K., Dhahbi, W., Makram, Z., Bragazzi, N.L., Aissa, M.B., Guelmemi, N., El Omri, A., Swed, S., Weiss, K., Knechtle, B., & Ben Saad, H. (2023). Moving beyond the stigma: Understanding and overcoming the resistance to the acceptance and adoption of artificial intelligence chatbots. *New Asian Journal of Medicine, 1*(2), 29–36. https://doi.org/10.61838/kman.najm.1.2.4

Donnelly, D.L. (2022). First do no harm: Legal principles regulating the future of artificial intelligence in health care in South Africa. *Potchefstroom Electronic Law Journal/Potchefstroomse Elektroniese Regsblad, 25*(1), 1–43. https://doi.org/10.17159/1727-3781/2022/v25ia11118

Doshi-Velez, F., & Kim, B. (2017). *Towards A Rigorous Science of Interpretable Machine Learning*. arXiv. https://doi.org/10.48550/arXiv.1702.08608

Dutta, D., & Mishra, S.K. (2024). Bots for mental health: The boundaries of human and technology agencies for enabling mental well-being within organisations. *Personnel Review, 53*(5), 1129–1156. https://doi.org/10.1108/PR-11-2022-0832

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly, 44*(2), 350–383. https://doi.org/10.2307/2666999

Engel, S. (2019). Minding machines: A note on alienation. *Fast Capitalism, 16*(2), 129–139. https://doi.org/10.32855/fcapital.201902.012

Ferrari, M., Sabetti, J., McIlwaine, S.V., Fazeli, S., Sadati, S.H., Shah, J.L., Archie, S., Boydell, K.M., Lal, S., Henderson, J., Alvarez-Jimenez, M., Andersson, N., Nielsen, R.K.L., Reynolds, J.A., Iyer, S.N., & Iyer, S.N. (2022). Gaming my way to recovery: A systematic scoping review of digital game interventions for young people's mental health treatment and promotion. *Frontiers in Digital Health, 4*, 814248. https://doi.org/10.3389/fdgth.2022.814248

Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medical Internet Research, 21*(5), e13216. https://doi.org/10.2196/13216

Folorunso, A., Olanipekun, K., Adewumi, T., & Samuel, B. (2024). A policy framework on AI usage in developing countries and its impact. *Global Journal of Engineering and Technology Advances, 21*(1), 154–166. https://doi.org/10.30574/gjeta.2024.21.1.0192

Guba, E.G., & Lincoln, Y.S. (1994). Competing paradigms in qualitative research. In N.K. Denzin & Y.S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105–117). Sage Publications, Inc.

Hong, Q.N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., Vedel, I., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information, 34*(4), 285–291. https://doi.org/10.3233/EFI-180221

Hussey, N. (2012). The language barrier: The overlooked challenge to equitable health care. *South African Health Review, 2012*(1), 189–195.

Kane, G.C., Young, A.G., Majchrzak, A., & Ransbotham, S. (2021). Avoiding an oppressive future of machine learning: A design theory for emancipatory assistants. *MIS Quarterly, 45*(1), 371–396. https://doi.org/10.25300/MISQ/2021/1578

Khalifa, M., Albadawy, M., & Iqbal, U. (2024). Advancing clinical decision support: The role of artificial intelligence across six domains. *Computer Methods and Programs in Biomedicine Update, 5*, 100142. https://doi.org/10.1016/j.cmpbup.2024.100142

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), 1–15. https://doi.org/10.1057/s41599-024-04139-2

Kim, B.J., Kim, M.J., & Lee, J. (2024). Code green: Ethical leadership's role in reconciling AI-induced job insecurity with pro-environmental behavior in the digital workplace. *Humanities and Social Sciences Communications, 11*(1), 1–16.

Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25: Proceedings of the 25th International Conference on Neural Information Processing Systems* (pp. 1097–1105). Curran Associates, Inc.

Kruger, S. (2024, June). Transcultural leadership and sustainable development in the digital era: Navigating the 4IR in South Africa. In *International Conference on Society 5.0* (pp. 207–217). Springer Nature Switzerland.

Lange, S., Santarius, T., Dencik, L., Diez, T., Ferreboeuf, H., Hankey, S., Hilbeck, A., Hilty, L.M., Höjer, M., Kleine, D., Pohl, J., Reisch, L.A., Ryghaug, M., Schwanen, T., & Staab, P. (2022). *Digital reset: Redirecting technologies for the deep sustainability transformation*. Universitätsverlag der TU Berlin.

Langlieb, A.M., Langlieb, M.E., & Xiong, W. (2021). EAP 2.0: Reimagining the role of the employee assistance program in the new workplace. *International Review of Psychiatry, 33*(8), 699–710. https://doi.org/10.1080/09540261.2021.2013172

Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. *Internet Policy Review, 9*(2), 1–16. https://doi.org/10.14763/2020.2.1469

Ledden, V. (2022). *The efficacy of employee assistance programs: Connecting HR & EAP practitioners in the Irish context*. Doctoral dissertation, National College of Ireland.

Matsepe, N.T., & Van der Lingen, E. (2022). Determinants of emerging technologies adoption in the South African financial sector. *South African Journal of Business Management, 53*(1), 2493. https://doi.org/10.4102/sajbm.v53i1.2493

Mayer, R.C., Davis, J.H., & Schoorman, F.D. (1995). An integrative model of organisational trust. *Academy of Management Review, 20*(3), 709–734. https://doi.org/10.2307/258792

Ndlovu-Gatsheni, S. (2018). The dynamics of epistemological decolonisation in the 21st century: Towards epistemic freedom. *The Strategic Review for Southern Africa, 40*(1), 16–45. https://doi.org/10.35293/srsa.v40i1.268

Ngcece, S., & Mkhize, S.M. (2023). An exploratory study of the South African Police Services (SAPS) systems in combating cybercrime. In S.O. Ehiane, S.A. Olofinbiyi & S.M. Mkhize (Eds.), *Cybercrime and challenges in South Africa* (pp. 159–175). Springer Nature Singapore.

Odero, B., Nderitu, D., & Samuel, G. (2024). The Ubuntu way: Ensuring ethical AI integration in health research. *Wellcome Open Research, 9*, 625. https://doi.org/10.12688/wellcomeopenres.23021.1

Orlikowski, W.J. (1999). *Technologies-in-practice: An enacted lens for studying technology in organizations (Working Paper #4056-99)*. Sloan School of Management, Massachusetts Institute of Technology. http://hdl.handle.net/1721.1/2742

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Parry, D.A., Le Roux, D.B., Morton, J., Pons, R., Pretorius, R., & Schoeman, A. (2023). Digital well-being applications: Adoption, use and perceived effects. *Computers in Human Behavior, 139*, 107542. https://doi.org/10.1016/j.chb.2022.107542

Prictor, M., Lewis, M.A., Newson, A.J., Haas, M., Baba, S., Kim, H., Kokado, M., Minari, J., Molnár-Gábor, F., Yamamoto, B., Kaye, J., & Teare, H.J. (2020). Dynamic consent: An evaluation and reporting framework. *Journal of Empirical Research on Human Research Ethics, 15*(3), 175–186. https://doi.org/10.1177/1556264619887073

Rane, N., Choudhary, S.P., & Rane, J. (2024). Acceptance of artificial intelligence: Key factors, challenges, and implementation strategies. *Journal of Applied Artificial Intelligence, 5*(2), 50–70. https://doi.org/10.48185/jaai.v5i2.1017

Republic of South Africa. (1996). *Constitution of the Republic of South Africa, 1996* (Government Gazette No. 17678). https://www.gov.za/sites/default/files/images/a108-96.pdf

Republic of South Africa. (2013). *Protection of Personal Information Act No. 4 of 2013*. Government Gazette, 578(37067).

Rousseau, D. (1995). *Psychological contracts in organisations: Understanding written and unwritten agreements*. Sage Publications.

Savaget, P., Geissdoerfer, M., Kharrazi, A., & Evans, S. (2019). The theoretical foundations of sociotechnical systems change for sustainability: A systematic literature review. *Journal of Cleaner Production, 206*, 878–892. https://doi.org/10.1016/j.jclepro.2018.09.208

Sebunya, J., & Gichuki, A. (2024). Digital tools and platforms for enhancing community participation: A review of global practices. *International Journal of Scholarly Practice, 4*(2), 54–67.

Sutherland, E. (2020). The fourth industrial revolution–the case of South Africa. *Politikon, 47*(2), 233–252. https://doi.org/10.1080/02589346.2019.1696003

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC medical research methodology, 8*(1), 45.

Trist, E.L., & Bamforth, K.W. (1951). Some social and psychological consequences of the longwall method of coal-getting: An examination of the psychological situation and defences of a work group in relation to the social structure and technological content of the work system. *Human Relations, 4*(1), 3–38. https://doi.org/10.1177/001872675100400101

Van den Broek, S., Sankaran, S., De Wit, J., & De Rooij, A. (2024, August). Exploring the supportive role of artificial intelligence in participatory design: A systematic review. In *Proceedings of the Participatory Design Conference 2024: Exploratory Papers and Workshops - Volume 2* (pp. 37–44). Association for Computing Machinery.

Vial, G. (2021). Understanding digital transformation: A review and a research agenda. In A. Hinterhuber, T. Vescovi, & F. Checchinato (Eds.), *Managing digital transformation* (pp. 13–66). Routledge.

Williams, M.D., Rana, N.P., & Dwivedi, Y.K. (2015). The unified theory of acceptance and use of technology (UTAUT): A literature review. *Journal of Enterprise Information Management, 28*(3), 443–488. https://doi.org/10.1108/JEIM-09-2014-0088

Zidaru, T., Morrow, E.M., & Stockley, R. (2021). Ensuring patient and public involvement in the transition to AI-assisted mental health care: A systematic scoping review and agenda for design justice. *Health Expectations, 24*(4), 1072–1124. https://doi.org/10.1111/hex.13299

Zou, B., Liviero, S., Hao, M., & Wei, C. (2020). Artificial intelligence technology for EAP speaking skills: Student perceptions of opportunities and challenges. In M.R. Freiermuth & N. Zarrinabadi (Eds.), *Technology and the psychology of second language learners and users* (pp. 433–463). Palgrave Macmillan.