AI and Ethics (2026) 6:12
https://doi.org/10.1007/s43681-025-00903-5

RESEARCH

Ethical AI in the workplace: ensuring fairness and transparency

Mariitta Rauhala[1] · **Merja Drake**[1] · **Pirjo Saaranen**[1]

Received: 21 May 2025 / Accepted: 29 August 2025 / Published online: 1 December 2025

Abstract
This study investigates how transparency, fairness, and employee participation influence the ethical use and development of artificial intelligence (AI) in the workplace, with a focus on knowledge workers in Finland. Drawing on a survey of 474 respondents, the research explores how these ethical principles contribute to employee engagement and the development of digital and developer agency. The study employs confirmatory factor analysis and path analysis to validate a four-factor model comprising sense of fairness, transparency and involvement, participation and engagement, and development of digital and developer agency. The results show that transparency and fairness significantly enhance employee participation and engagement, which in turn fosters the development of digital and developer agency. The findings highlight the importance of inclusive and transparent AI practices in promoting ethical AI adoption and strengthening professional agency in evolving work environments. The study contributes to the growing body of research on responsible AI by offering empirical evidence on the social and organisational dimensions of AI ethics.

Keywords Artificial intelligence · Fairness · Transparency · Ethics

1 Introduction

The ethical use of artificial intelligence (AI) involves developing and applying AI technologies in a manner that respects human rights, fairness, transparency and accountability. In practice, this means that developers must identify and minimise potential biases and errors that could affect AI decisions. Additionally, developers and users must understand how AI makes decisions and the impacts of its decisions, such as on privacy. The development and use of AI must also consider the perspectives and needs of various stakeholders [1, 2]. User feedback and subsequent iterative improvement help ensure the fairness of AI systems [3].

In this article, the term "artificial intelligence" specifically refers to generative AI. Generative AI is a form of AI that creates new content, such as text, music, and images, with the user given a text-based prompt [4]. The increasing use of generative AI in workplaces and the rapid development of more advanced AI significantly affect the nature of work [5].

The ethical minimum for using AI should therefore include transparency, fairness, non-maleficence, accountability, privacy and responsibility [6, 7]. Additionally, transparency should be coupled with accountability, meaning that developers and users of AI systems are responsible for the operation, decisions and consequences of such systems [8].

Fairness in the use and development of AI means that AI systems make decisions without discrimination or bias and treat all users and groups equitably. This can be achieved by ensuring that different groups receive equal treatment (group fairness) and that individuals with similar characteristics receive similar treatment (individual fairness). Achieving fairness requires AI systems to be transparent, accountable and explainable [9].

Transparency is a key part of responsible AI development and use, promoting user trust and enabling effective oversight and error correction. This means that the operation and decision-making processes of as well as data used

✉ Merja Drake
merja.drake@haaga-helia.fi

Mariitta Rauhala
mariitta.rauhala@haaga-helia.fi

Pirjo Saaranen
Pirjo.Saaranen@haaga-helia.fi

1 Haaga-Helia University of Applied Sciences, Helsinki, Finland

by AI systems are clearly understandable and visible to all stakeholders [10].

Employee engagement means an emotional commitment and dedication to their work and to advancing the company's goals and objectives. It is a key factor in organisational success, since it directly affects employee satisfaction, productivity and commitment as well as overall business performance [11–13]. Engagement helps ensure that AI systems are designed and implemented fairly and transparently, thereby increasing employee trust and acceptance [14].

While engagement signifies that employees are emotionally invested in their work and organisation, participation refers to their active involvement in various activities and processes. This can occur in teamwork, projects or daily tasks [12]. It is emphasised that to ensure the ethicality of AI, it is important that all stakeholders, including employees, participate in its implementation process [14]. Participation is reflected in how stakeholders share their views and develop a common understanding of fair decision-making through reflection and discussions [15].

Involvement can be defined as the process by which employees participate in decision-making autonomously to plan their work in the best possible way, thereby impacting organisational productivity and employee well-being [16, 17]. The concept of involvement also includes teamwork within organisations, transparent communication practices and an atmosphere that encourages employee participation [18–20].

This article focuses on engaging the work community in the ethical use and development of AI to ensure fairness and transparency, particularly from the perspective of knowledge workers. The research was conducted based on a survey of knowledge workers in Finland in the spring of 2024 (n = 474). Our study shows that open discussion about the ethical use of AI and its potential benefits is desired in work communities. Issues include considering data security at work, developing work processes and identifying and addressing ethical problems. Knowledge workers seek ethical guidelines on what can be done and how to act in different situations. Similarly, the impacts of AI on one's job description and on the future of work are topics of discussion.

2 What is ethical use of AI in the workplace?

The rapid adoption of AI in organisational operations raises ethical questions that require careful consideration. While researchers have discussed AI ethics, they have usually referred to the ethical competence of technology developers and their understanding of how to develop AI ethically and impartially [21] or to the ethics of AI users [8], since technology is always intertwined with human actions [22].

Ethical AI use means using AI in a way that supports employee well-being and meaningful work. The adoption of AI in workplaces significantly affects employees' experiences of meaningful work, which is important for their well-being and autonomy. AI adoption can both enhance and diminish experiences of meaningful work [5]. This supports the idea that inclusive practices are important in considering the social impact of AI.

Human decisions are often not optimal but subjectively biased and irrational [23]. AI's decision-making capability is recognised as one of its most significant features [24–27]. One of the primary ethical considerations is transparency in AI decision-making processes [28–31]. Organisations should ensure that AI systems' decision-making processes are clear and understandable for all stakeholders [32]. This also applies to employees using AI in their tasks. Transparency allows individuals to understand how AI systems make decisions that affect them [30].

However, research has suggested that even if the factors of algorithmic decision-making are transparent, individuals may still perceive the results as unfair [33] or even biased [34–36]. Researchers have emphasised the importance of fairness and have proposed methods for detecting and reducing biases in AI systems to help prevent discrimination and promote inclusion [37]. Inclusive practices that consider the social impact of AI adoption are crucial.

To make AI systems fair, just and inclusive, organisations should regularly interact with AI users and developers, gather feedback from them, assess the impact of the feedback and ensure that the AI systems adapt to ethical norms [32, 38]. Forming diverse teams in AI development helps in identifying and correcting biases, while collecting user feedback improves system functionality and fairness [9].

Developing explainable and interpretable AI models increases transparency and helps users understand how an AI system arrives at certain decisions [39]. It is important to consider fairness metrics in employee training processes so that AI users and developers can identify discriminatory and unfair AI models [40].

Organisations should create proactive and inclusive measures to ensure that the benefits of AI are distributed equitably and do not widen existing socioeconomic gaps [41]. Organisations that prioritise responsible data practices promote the construction of reliable AI systems that advance transparency and privacy and align with ethical principles [42].

Researchers have found that weaknesses related to AI ethics include a lack of transparency and ethical guidelines and practices [6, 36, 43]. As AI systems become increasingly complex, a lack of transparency can lead to ambiguity in decision-making, which jeopardises trust among stakeholders. The importance of transparency is not only in building trust but also in ensuring accountability and understanding of AI-based decisions and actions that affect individuals [44, 45].

Hypothesis 1 *Transparency in decision-making and inclusive practices promotes participation and engagement in the use of AI.*

2.1 Transparency and fairness in developing and using AI at work

When studying organisations, it is important to use a multi-level approach to understanding the impacts of AI at different levels and their interconnections [5]. The impact and utilisation opportunities of AI vary at different levels, so a multi-level approach helps in identifying and optimising the benefits of AI and managing its risks. At the same time, multi-levelness enables effective decision-making, collaboration and the promotion of innovation at all levels of organisations [10]. The impacts of AI should be studied at different levels to develop comprehensive strategies for utilising AI in organisations [5]. To make AI systems fair, just and inclusive, organisations should regularly interact with AI users and developers, gather feedback from them, assess the impact of the feedback and ensure that AI systems adapt to ethical norms [32, 38].

Organisational researchers are interested in how the sense of fairness relates to the use of AI, since a sense of injustice can negatively affect, among other things, employee productivity, the treatment of other employees and it may ultimately result in the complete loss of employment [45]. Marikyan et al. [46] studied factors that can lead to individuals' satisfaction with the use of AI-based digital assistants and examined the impact of satisfaction on productivity and work engagement. Their results showed that expected performance, social presence and trust were positively associated with job satisfaction, which in turn correlated with productivity and engagement.

At the individual level, knowledge workers' use of AI and opportunities to participate in AI development have been examined from the perspectives of professional agency [47, 48], digital agency [45, 49, 50, 66], developmental agency [51] and transformative agency [52, 53].

Agency can appear as an individual or collective characteristic of the work community [50]. It relates to the choices made by individuals and their consequences [54]. Professional agency in working life encompasses power, competence, professional development and influence over one's work [47, 48, 55].

Digital agency consists of digital competence, trust and accountability [45]. It refers to the ability to control and adapt to the digital world and to use digital technologies ethically and safely [49]. A digital agent has a strong understanding of how and where technology can be utilised in work tasks but also perceives the ethical challenges of digitalisation [56]. The role of digital agency is significant in promoting equality, equity, democracy and well-being [50].

An employee with developer agency is proactive and raises everyday ideas. Developer agency also involves collaboration with other developmental agents [51, 56]. It encompasses motivation for the adoption of technologies and the design of new applications [56]. Developmental agency is shaped by changes in communities, especially in conflict situations [51, 57].

As professional agency develops into digital and developer agency, the need for individual autonomy increases, and one's ability to understand the impacts, benefits and risks of technology expands. Ideas and development proposals must be put into practice, which requires transformative agency. A transformative agent takes care of shared negotiations within a team and initiates changes in operations [52, 53]. Agency types, their definitions, examples in AI context, and their relation to other agency types are show in the Table 1.

When individuals have achieved developmental and transformative agency, team-level development, shared experiences and building a common understanding are important for the ethical development and use of AI. The adoption of AI can lead to the emergence of new job roles that combine technology and business needs. These roles

Table 1 The agency types and their example in AI context

Type of agency	Definition	Example in AI context	Relation to other types
Professional agency [47, 48]	The capacity of individuals to act purposefully and reflectively in their work roles	Making informed decisions about integrating AI into one's workflow	Forms the foundation for other agency types; evolves through digital and developer agency
Digital agency [45, 49, 50]	The ability to critically and creatively engage with digital technologies	Using AI tools to enhance productivity or solve problems	Builds on professional agency; enables deeper interaction with technological systems
Developer agency [51]	The capacity to co-create, adapt, or influence the development of digital tools	Participating in the design or customization of AI applications for workplace needs	Extends digital agency; involves active shaping of technological environments
Transformational agency [52, 53]	The ability to initiate change in practices, structures, or norms	Leading organizational shifts in AI adoption or ethical AI governance	Emerges from the interplay of all other agency types; reflects systemic impact

enable teams to better utilise AI and develop new innovations [58].

AI developers and users should understand how to mitigate the ethical challenges related to AI [59, 60, 63], such as by creating ethical guidelines and codes of conduct [43]. However, ethical guidelines have been criticised, with [61] finding that codes of conduct developed by companies had little impact on the decision-making of technology developers or remained too abstract [62]. Researchers have emphasised that AI ethical guidelines are largely useless and that we should instead talk about AI fairness [8].

If AI applications are developed to be explainable and interpretable, they increase transparency and help users understand how AI systems arrive at certain decisions [65]. Training processes should also consider fairness metrics so that users can identify discriminatory and unfair AI applications.

Organisations should create proactive and inclusive measures to ensure that the benefits of AI are distributed equitably and do not widen existing socioeconomic gaps. Organisations that prioritise responsible data practices promote the construction of reliable AI systems that increase transparency, ensure privacy protection and align with ethical principles.

The impact of algorithms on organisations is not solely technological but also involves significant social and organisational factors. The use of algorithms is shaped by organisational norms, values and individual interpretive frameworks, which together determine how algorithms reorganise organisational processes and practices. This also highlights the need to understand the use of algorithms in a broader social and organisational context for ethical AI use [9].

Hypothesis 2 *The sense of fairness plays a significant role in knowledge workers' participation in the development and use of AI and in developing their agency.*

Hypothesis 3 *The use of AI is significant in the development of knowledge workers' professional agency into digital and developmental agency.*

The conceptual model (Fig. 1 below) illustrates both direct and indirect pathways through which Sense of Fairness influences Digital and Developmental Agency. The direct effect refers to the immediate impact of Sense of Fairness on Digital and Developmental Agency. In contrast, the indirect effect is mediated by Participation and Engagement, suggesting that Sense of Fairness enhances individuals' participation and engagement, which in turn fosters their digital and developmental agency.

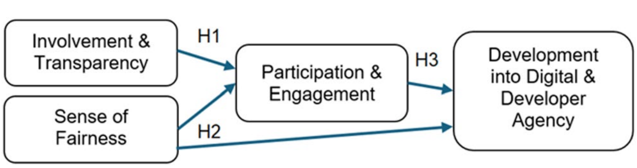

Fig. 1 Model illustrating the hypotheses and relationships between the factors

3 Methods

3.1 Data and variables

The research data were collected from a Webropol online survey from February 14 to April 22, 2024. The survey link was distributed to members of several professional unions employing knowledge workers in Finland as well as to employers known to employ knowledge workers. Additionally, the survey was sent to 28,724 representatives of private and public organisations obtained from the Selector company database, representing a wide range of companies of different sizes and industries. A total of 484 people responded to the survey, of whom 10 were excluded because they were retired, full-time students, on parental or care leave or unemployed. The remaining 474 respondents were included in the analysis because they were currently actively employed.

The survey contained 43 questions, including both quantitative and qualitative open-ended questions based on the literature review and agency theories. In the preliminary trial phase of the study, 21 Likert scale variables were selected for consideration in this article, of which 17 were included in the final model. The variables utilised in this study were measured using five-point (12 variables) and seven-point Likert scales (5 variables).

The dataset was described using seven demographic variables. The respondents included representatives from both companies and public administration. The industries represented included insurance and finance, information and communication, hotels and restaurants, education, industry, administrative and support services, national defence, real estate, technical, construction, wholesale and retail trade, health and social services, arts, entertainment and recreation and other services. Slightly more than half of the respondents (51.5%) worked in large organisations with over 249 employees, and less than a quarter (23.4%) worked in medium-sized organisations with 50–249 employees. Small organisations with 10–49 employees employed 11.8% of the respondents, and 13.3% of the respondents worked in organisations with fewer than 10 employees. The majority of respondents (92.0%) worked full time and 81.6% had completed a higher education degree. Of the respondents, 41.8% were aged over 51 years, 35.4% were in the 41–50 age group, 18.4% were in the 31–40 age group, and only 4.4% were aged under 30 years. Women made up 64.4% of the respondents and men made up 31.6%. The majority of respondents had a career length of 20–29 years (37.3%). In terms of workplace position, the largest group consisted of professionals, comprising 41.6% of the respondents (Table 2).

3.2 Measurement model

The data obtained were analysed using SPSS 29.0 and AMOS 29.0 statistical data analysis software. The selected Likert scale variables were first pretreated; four negatively phrased variables were transposed to positive, and the seven-point Likert scale variables were converted to five-point Likert scale variables to standardise the measurement scales for the analyses and the composite variables.

Table 2 Demographics of the respondents

Variable	Frequency	Percentage
Gender		
Male	150	31.6%
Female	**306**	**64.6%**
Other or prefer not to say	18	3.8%
Age		
≤30	21	4.4%
31–40	87	18.4%
41–50	168	35.4%
51–64	**193**	**40.7%**
≥65	5	1.1%
Education		
No vocational education	1	0.2%
Vocational education	86	18.1%
Higher university-level education	**387**	**81.6%**
Employment		
Full-time	**436**	**92.0%**
Part-time	15	3.2%
Entrepreneur/self-employed	23	4.9%
Number of employees		
< 10 (micro)	63	13.3%
10–49 (small)	56	11.8%
50–249 (medium-sized)	111	23.4%
> 249 (large)	**244**	**51.5%**
Length of career in years		
< 5	15	3.2%
5–9	34	7.2%
10–19	111	23.4%
20–29	**177**	**37.3%**
30–39	106	22.4%
> 39	31	6.5%
Position in the workplace		
Manager (supervisor)	145	30.6%
Specialist	75	15.8%
Professional	**197**	**41.6%**
Clerical support worker	42	8.9%
Service or sales worker	7	1.5%
Other category of worker	8	1.7%

The next step involved conducting confirmatory factor analyses (CFAs). The maximum likelihood (ML) estimation method was used in the CFA, as the sample size (N = 474) exceeds the commonly recommended minimum of 200 cases for ML estimation. CFA is utilised to evaluate the measurement model within equation modelling by testing the validity of hypothesised relationships between observed variables and latent constructs. It verifies whether the observed variables accurately represent the latent constructs they are intended to measure. CFA was used to examine the four latent factors: sense of fairness (SF), involvement and transparency (IT), participation and engagement (PE) and development of digital and developer agency (DDA). In the first round of analysis, 21 observed variables were selected for the latent factors, but the model fit indices were not valid. Due to low factor loadings, four observed variables were deleted from the model, leaving 17 variables. In addition, error covariances were added within the latent factors on high modification indices (MI > 10), suggesting shared measurement variance not captured by the latent construct. Figure 2 describes the revised measurement model as the result of CFA.

The goodness-of-fit values suggest an acceptable fit of the model. Chi-square divided by degrees of freedom (χ^2/df) was 2.464, indicating an acceptable fit relative to the degree of freedom. The comparative fit index (CFI) and normed fit index (NFI) values were 0.950 and 0.919, indicating an acceptable fit. The Tucker–Lewis index (TLI) was acceptable, with a value of 0.936. The root mean square error of approximation (RMSEA) value was 0.056, and the standardised root mean square residual (SRMR) value was 0.0562. Since these values were within good or acceptable limits the four-factor structure was validated. The goodness-of-fit index (GFI) value of 0.942 and the adjusted GFI (AGFI) value of 0.918 indicate an acceptable fit between the model and the observed data. Table 3 presents the fit indices for the original model and the modified model after the removal of four variables and added error covariances, along with the corresponding threshold values [63, 67, 68, 69].

All factor loadings were above 0.5, and the Cronbach's alphas were at a good level (> 0.8) in three factors and at an acceptable level (0.749) in one factor [69]. Based on the "Cronbach's Alpha if Item Deleted" values, removing any of the selected items would not improve the internal consistency of the factors. Despite some standardized residual covariances exceeding ± 2, the indicators were retained in the model due to their conceptual importance. For instance, the elevated residuals for PE2 and TI2 were not considered problematic from a theoretical standpoint. PE2 reflects the

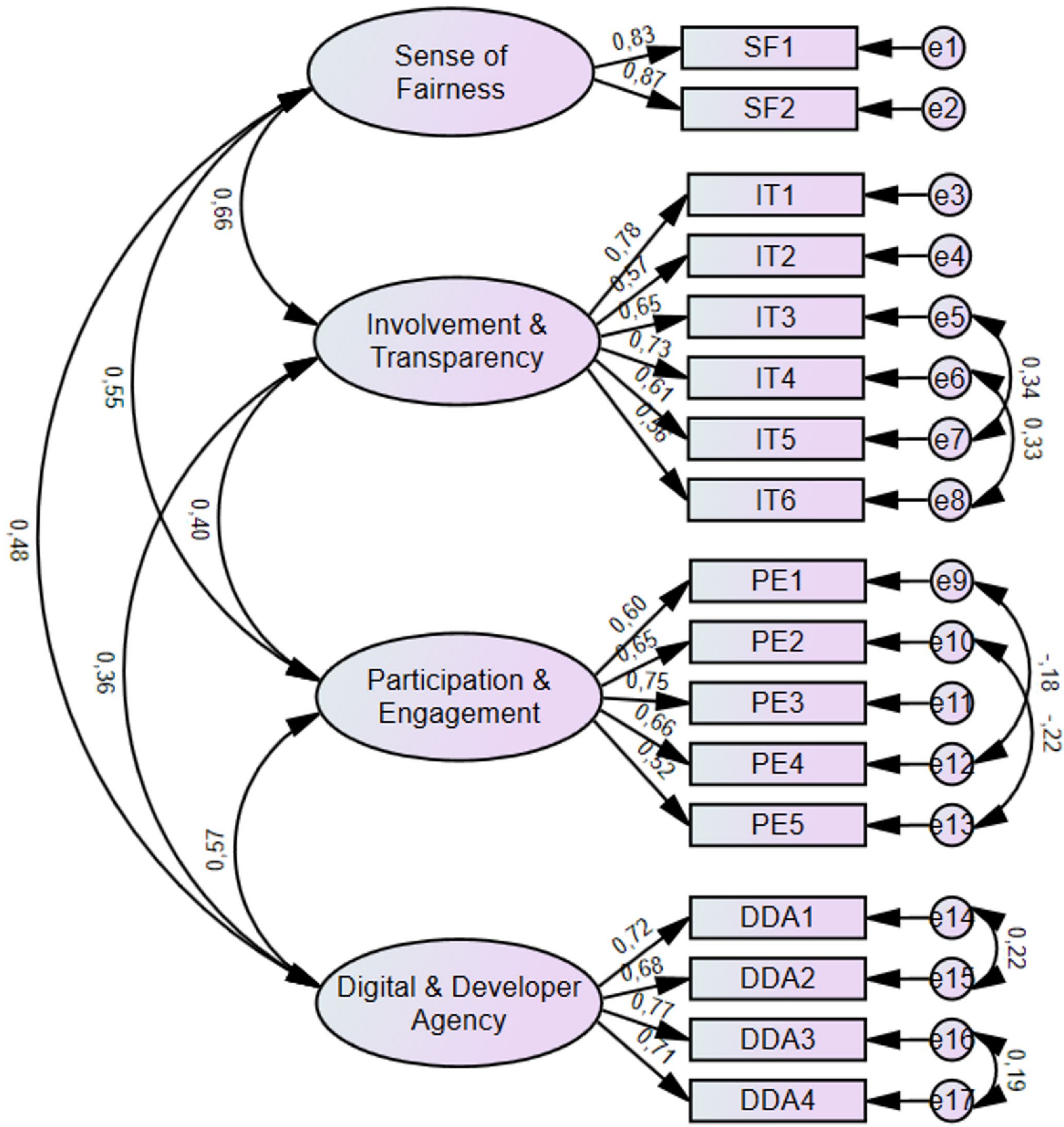

Fig. 2 Confirmatory factor analysis diagram

individual's perceived expertise to contribute to AI development, while TI2 captures the adequacy of instructions provided for utilizing AI in daily tasks. These items represent essential dimensions of organizational AI readiness—namely, competence and support—and offer perspectives not fully captured by other indicators in the model. Given their theoretical relevance and the overall satisfactory model fit, no modifications were made. The factors, factor loadings and Cronbach's alphas are described in Table 4.

3.3 Hypothesis testing using path analysis

Path analysis examines the structural relationships between composite variables. The results include standardised regression coefficients, standard errors, t-values and *p*-values for

Table 3 Model fit indices

Model fit index	Original model (21 variables)	Revised model (17 variables)	Threshold values
χ^2/df	3.148	2.464	<3 acceptable, <2 good
CFI	0.897	0.950	>0.90 acceptable, >0.95 good
TLI	0.876	0.936	>0.90 acceptable, >0.95 good
RMSEA	0.067	0.056	<0.08 acceptable, <0.05 good
SRMR	0.0744	0.0562	<0.08 good
NFI	0.857	0.919	>0.90 good
GFI (AGFI)	0.905 (0.874)	0.942 (0.918)	>0.90 good

each path in the model, indicating whether each path is statistically significant.

For the path analysis, composite variables were created for sense of fairness, involvement and transparency, participation and engagement and development of digital and developer agency by calculating the mean of the observed indicators for each latent variable. The path analysis examined the relationships between these composite variables (Fig. 3). The p-value of the chi-square (χ^2) test was 0.231 (>0.05), with the nonsignificant value indicating that the model fitted the data. The values for CFI (0.999), NFI (0.997), TLI (0.994) and GFI (0.998), were all above 0.95. Additionally, the RMSEA (0.030) and SRMR (0.0127) were both below 0.08. These model fit indices indicated a good fit.

The path analysis indicated that the impact of involvement and transparency on participation and engagement was positive and significant ($p=0.004$), supporting Hypothesis 1. Similarly, the impact of sense of fairness on participation and engagement was positive and significant ($p<0.001$), supporting Hypothesis 2 Furthermore, the impact of participation and engagement on how employees develop as digital and developer agents was positive and significant ($p<0.001$), supporting Hypothesis 3 Sense of fairness also had a direct impact on how employees develop as digital and developer agents (Table 5).

The standardised total effect of sense of fairness to the development of digital and developer agency was 0.365 ($0.377 \cdot 0.337+0.238$). Squared multiple correlations (Table 6) indicated that the model explained 21.8% of the variance in the level of participation and engagement and 24.3% in the level of development of digital and developer agency. The results suggest that sense of fairness and involvement and transparency play a moderate role in enhancing participation and engagement, in turn contributing to development of digital and developer agency.

4 Discussion

The results indicate that the measurement model used was generally reliable, as most factor values were above 0.5 and the Cronbach's alpha values were at a good level (>0.8) for three factors (sense of fairness, transparency and involvement, development of digital and developer agency) and at an acceptable level (0.749) for one factor (participation and engagement). This suggests that the observed variables can measure latent factors, such as sense of fairness, participation and transparency, engagement and development of digital and developer agency. In particular, the sense of fairness and development of digital and developer agency factors showed high reliability, highlighting their importance in the study.

We further examined the results according to three hypotheses. For Hypothesis 1 (transparency in decision-making and inclusive practices promote participation and engagement in the use of AI), the factor values for transparency and involvement ranged from 0.557 to 0.785 and the Cronbach's alpha was 0.830 (see Table 4). The path analysis (Table 3) indicated that transparency and involvement is a significant factor influencing employee engagement and commitment to AI use. This is supported by the theoretical framework, as transparency in decision-making can ensure employee involvement [16–19].

For Hypothesis 2 (the sense of fairness plays a significant role in knowledge workers' participation in the use and development of AI and developing their agency), the factor values for sense of fairness were high (0.832 and 0.873) and the Cronbach's alpha was 0.841 (see Table 4). The path analysis (Table 5) indicated that sense of fairness is strongly associated with employees' engagement in AI use and development. Marikyan [46] and Dietz [64] also emphasised that a sense of fairness impacts a person's proactivity and work engagement.

For Hypothesis 3 (the use of AI is significant in the development of knowledge workers' professional agency into digital and developmental agency), the factor values for development of digital and developer agency ranged from 0.678 to 0.771 and the Cronbach's alpha was 0.829. The path analysis (Table 5) indicated that AI use is a significant factor in employees' professional agency development. The high factor values and good reliability supported the hypothesis that AI use promotes employees' professional agency development towards digital and developer agency. Previous agency research has focused more on the general impact of technology on agency development [45, 51, 56], although there remains little research specifically on AI-related agency.

Although there were some lower factor values in the transparency and involvement and participation and engagement

Table 4 Factors, factor loadings, and Cronbach's alphas

Factor	Variable (scale from 1 to 5)	Factor loadings	Cronbach's alpha
Sense of fairness	SF1: My ability to utilise AI in my work is recognised as part of my professional development	0.832	0.841
	SF2: My contribution to AI projects or the development of AI is appreciated and acknowledged	0.873	
Transparency and involvement	TI1: I feel that I receive enough information and/or resources from my employer to be proactive and to take the initiative in using AI in my work	0.785	0.830
	TI2: I feel that I have enough instructions for utilising AI in my work tasks	0.557	
	IT3: Our work community has clear plans for utilising artificial intelligence	0.640	
	IT4: My work community (foreman, management, communications) keeps staff up to date on the AI applications under development	0.737	
	IT5: We develop AI collaboratively	0.613	
	IT6: In my work community, all employees are equally involved in the development of artificial intelligence	0.562	
Participation and engagement	PE1: I recognise opportunities to utilise AI in the work and processes shared by the work community	0.602	0.749
	PE2: I have enough expertise to be involved in the development of artificial intelligence	0.648	
	PE3: I have ideas on how to develop the work using artificial intelligence	0.753	
	PE4: I would like to participate in the development of artificial intelligence to learn new things and skills	0.657	
	PE5: I am interested in developing my work with AI	0.517	
Development of digital and developer agency	DDA1: How do you feel that the use of AI has affected your self-efficacy in your work?	0.721	0.829
	DDA2: The use of AI has changed my ability to cope with challenging tasks in my work	0.678	
	DDA3: The use of AI has changed my opportunities to tailor my work to my own vision	0.771	
	DDA4: The use of AI has changed the proactive and initiative-taking approach in my work	0.710	

factors, the overall reliability was still at an acceptable level. On the other hand, the lower level of the participation and engagement factor may be explained by the fact that during the data collection phase (spring 2024), AI adoption was still in the experimental phase in many organisations.

In summary, the path analysis results (Table 5) supported all the presented hypotheses. Participation and transparency were shown to significantly promote engagement and commitment to AI use ($p = 0.004$), confirming Hypothesis 1. Sense of fairness was shown to significantly affect both engagement and commitment ($p < 0.001$) and the development of digital and developer agency ($p < 0.001$), confirming Hypothesis 2 Engagement and commitment were shown to significantly promote the development of digital and developer agency ($p < 0.001$), confirming Hypothesis 3 The results of this study highlight the importance of transparency, fairness and participation in developing employee agency in the context of AI use.

According to Table 6, the model explained 21.8% of the variation in engagement and commitment and 24.3% of the variation in digital and developer agency. In relation to the previously presented research results, this means that while the model explained a significant part of the variation in these factors, there are still other factors that influence engagement, commitment and professional agency. Transparency, fairness and participation are important, but other factors should also be studied more closely in the future. Based on the factor analysis, a low but moderate Cronbach's alpha (0.749) was obtained for the factors of participation and engagement, but this can be explained by the fact that in the spring of 2024, the survey participants still had only limited experience with the use of AI.

It is acknowledged that elevated residual covariances may affect local validity; however, the global fit indices and conceptual coherence of the model support its use for the study's objectives. The theoretically grounded choices strengthen the relevance of the model.

The results of the study may have been distorted somewhat by the fact that the respondents were almost all knowledge workers. The respondents were also likely individuals who were interested in AI. The results of the study may be somewhat distorted by the fact that the respondents were almost all knowledge workers. The respondents were also likely those individuals who are interested in AI. Notably, the majority of respondents in this study were women, which contrasts with previous findings suggesting that men are generally more inclined to adopt and use AI technologies. However, the present results indicate that female participants demonstrated a high level of interest not only in utilizing AI applications but also in actively contributing to their development within their respective workplaces. This suggests a potential shift in gendered engagement with AI,

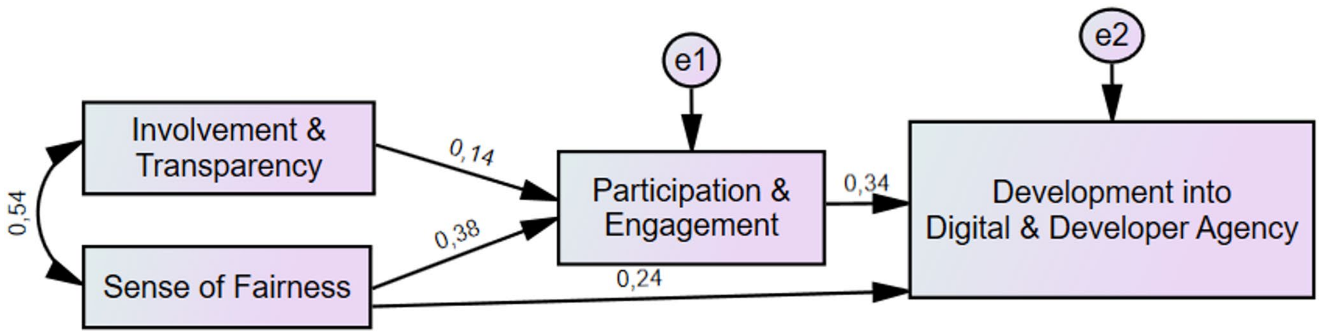

Fig. 3 Path analysis diagram

Table 5 Path analysis results

Hypothesis	Standardised estimate	t-value	*p*-value	Decision
H1: Involvement and transparency > participation and engagement	0.139	2.895	0.004	H1 is supported
H2: Sense of fairness > participation and engagement	0.377	7.834	< 0.001	H2 is supported
H2: Sense of fairness > development of digital and developer agency	0.238	5.299	< 0.001	H2 is supported
H3: Participation and engagement > development into digital and developer agency	0.337	7.524	< 0.001	H3 is supported

Table 6 Squared multiple correlations

	Estimate
PEmean	0.218
DDAmean	0.243

highlighting the importance of inclusive design and participatory development strategies.

The findings underscore the pivotal role of perceived fairness and participatory decision-making in shaping employees' acceptance of artificial intelligence in the workplace. Specifically, the opportunity for employees to engage in decisions regarding the use and development of AI technologies emerges as a critical determinant of trust and commitment. This reinforces the notion that AI implementation is not solely a technical endeavor but inherently involves ethical and inclusive leadership practices [44].

While prior research has predominantly emphasized the economic and efficiency-related benefits of AI [6], this study contributes a novel perspective by examining its implications for professional agency. The results suggest that AI can function as a catalyst for the development of digital and developer agency among knowledge workers—a dimension that has received limited systematic attention in existing literature [46].

Transparency is identified as a key enabler of both fairness, participation and engagement, serving as a foundational element in fostering employee engagement. These factors not only enhance organizational commitment but also support the evolution of professional identity in increasingly digital work environments.

From a practical standpoint, organizations must critically reflect on how AI adoption and development are governed. Ethical and transparent leadership, coupled with participatory decision-making processes, can significantly strengthen employees' trust and proactive engagement with AI systems. This calls for a shift from top-down implementation models toward participatory frameworks that align technological innovation with human-centered values. The complexity of knowledge work and constant change require leadership that supports co-creation, a culture of experimentation, and pluralistic dialogue within companies and networks.

The open-ended responses complemented the quantitative findings by highlighting that transparency and employee involvement contribute to the enhancement of digital and developmental agency. Furthermore, the qualitative data revealed that female respondents and those aged between 41 and 50 were particularly eager to engage in the adoption and development of artificial intelligence. These participants referred to themselves as 'AI research explorers' and 'pathfinders' in their open responses, indicating a proactive and pioneering attitude toward AI integration.

5 Conclusion

The study provides compelling evidence that AI is profoundly transforming working life and simultaneously influencing the development of agency at the individual, team and organisational levels. It holds significant potential to strengthen employees' professional agency—particularly towards digital and developmental agency—which highlights the need for further research into AI's role in this evolution.

This study clearly demonstrated that transparency and a sense of fairness are central to the use of AI, especially from the perspectives of participation and work engagement. For employees to commit to their work and feel that they are working in a fair environment, they must have a clear understanding of what their organisation expects from them and the vision of management towards the use of AI.

Therefore, it is crucial for organisational leadership to promptly define clear guidelines, boundaries and a vision for the use and development of AI and other emerging technologies. These must be communicated openly, transparent and comprehensibly to the entire workforce.

Future research should explore the role of leadership in the adoption and integration of artificial intelligence within organizational contexts. Additionally, further studies are warranted to investigate the factors underlying gender-based differences in attitudes toward AI, as well as the distinctions observed across various age groups.

Author contributions M.R. All authors have contributed equally to the preparation and writing of this manuscript. All authors reviewed the manuscript M.D. All authors have contributed equally to the preparation and writing of this manuscript. All authors reviewed the manuscript P.S. All authors have contributed equally to the preparation and writing of this manuscript. All authors reviewed the manuscript.

Funding This research was partially funded by the Finnish Work Environment Fund (Työsuojelurahasto).

Data availability The metadata of the research data and the research data in CSV format is published on Qvain Fairdata repository: https://doi.org/10.23729/fd-7d07bd8c-840c-3f00-8806-bc7f84c179b0

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Resnik, D.B., Hosseini, M.: The ethics of using artificial intelligence in scientific research: new guidance needed for a new tool. AI Ethics (2024). https://doi.org/10.1007/s43681-024-00493-8
2. UNESCO: Recommendation on the ethics of artificial intelligence. https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence (2022). Accessed 15.9.2024
3. Akinrinola, O., Okoye, C.C., Ofodile, O.C., Ugochukwu, C.E.: Navigating and reviewing ethical dilemmas in AI development: strategies for transparency, fairness, and accountability. GSC Adv. Res. Rev. (2024). https://doi.org/10.30574/gscarr.2024.18.3.0088
4. Stanford University: Generative AI: perspectives from Stanford HAI. Stanford University Human-Centered Artificial Intelligence. https://hai.stanford.edu/sites/default/files/2023-03/Generative_AI_HAI_Perspectives.pdf (2023). Accessed 15.9.2024
5. Bankins, S., Formosa, P.: The ethical implications of artificial intelligence (AI) for meaningful work. J. Bus. Ethics (2023). https://doi.org/10.1007/s10551-023-05339-7
6. Hagendorff, T.: The ethics of AI ethics: an evaluation of guidelines. Mind. Mach. (2020). https://doi.org/10.1007/s11023-020-09517-8
7. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nat. Mach. Intell. (2019). https://doi.org/10.1038/s42256-019-0088-2
8. Munn, L.: The uselessness of AI ethics. AI Ethics (2023). https://doi.org/10.1007/s43681-022-00209-w
9. Ferrara, E.: Fairness and bias in artificial intelligence: a brief survey of sources, impacts and mitigation strategies. Sci (2023). https://doi.org/10.48550/arXiv.2304.07683
10. Liao, Q.V., Wortman Vaughan, J.: AI transparency in the age of LLMs: a human-centered research roadmap. Harv. Data Sci. Rev. (2024). https://doi.org/10.1162/99608f92.8036d03b
11. Mittal, P., Jora, R.B., Sodhi, K.K., Saxena, P.: A review of the role of artificial intelligence in employee engagement. In: 2022 8th Int. Conf. Adv. Comput. Commun. Syst. (2023). https://doi.org/10.1109/icaccs57279.2023.10112957
12. Rožman, M., Oreški, D., Tominc, P.: Integrating artificial intelligence into a talent management model to increase the work engagement and performance of enterprises. Front. Psychol. (2022). https://doi.org/10.3389/fpsyg.2022.1014434
13. Schaufeli, W.B., Salanova, M., González-Romá, V., Bakker, A.B.: The measurement of engagement and burnout: a two sample confirmatory factor analytic approach. J. Happiness Stud. (2002). https://doi.org/10.1023/A:1015630930326
14. Ayling, J., Chapman, A.: Putting AI ethics to work: are the tools fit for purpose? AI Ethics (2021). https://doi.org/10.1007/s43681-021-00084-x
15. Zhang, A., Walker, O., Nguyen, K., Dai, J., Chen, A., Lee, M.K.: Deliberating with AI: improving decision-making for the future through participatory AI design and stakeholder deliberation. Proc. ACM Hum. Comput. Interact. (2023). https://doi.org/10.1145/3579601
16. Amah, E., Ahiautzu, A.: Employee involvement and organizational effectiveness. J. Manag. Dev. (2013). https://doi.org/10.1108/JMD-09-2010-0064
17. Eliyana, A., Buchdadi, A.D., Hamidah, Sariwulan, T., Muhaziroh, K.: The effect of employee involvement on job satisfaction. Sys. Rev. Pharm. (2020). https://doi.org/10.31838/srp.2020.7.72
18. Abildgaard, J.S., Hasson, H., von Thiele Schwarz, U., Løvseth, L.T., Ala-Laurinaho, A., Nielsen, K.: Forms of participation: the development and application of a conceptual model of participation in work environment interventions. Econ. Ind. Democr. (2020). https://doi.org/10.1177/0143831X17743576

19. Atouba, Y.C.: Tackling the turnover challenge among IT workers: examining the role of internal communication adequacy, employee work participation, and organizational identification. Commun. Rep. (2018). https://doi.org/10.1080/08934215.2018.1497180
20. Shadur, M.A., Kienzle, R., Rodwell, J.J.: The relationship between organisational climate and employee perceptions of involvement: the importance of support. Group Organ. Manag. **24**, 479–503 (1999)
21. Borenstein, J., Howard, A.: Emerging challenges in AI and the need for AI ethics education. AI Ethics (2021). https://doi.org/10.1007/s43681-020-00002-7
22. Verbeek, P.P.: Designing the morality of things. The ethics of behaviour-guiding technology. In: van den Hovens, J., Miller, S. Pogge, T. (eds.) Designing in Ethics, pp. 78–94. Cambridge University Press, Chicago (2017)
23. Gil, D., Hobson, S., Mojsilović, A., Puri, R., Smith, J.R.: AI for management: an overview. J. Acad. Mark. Sci. (2020). https://doi.org/10.1007/978-3-030-20680-2_1
24. Brock, J.K.-U., von Wangenheim, F.: Demystifying AI: what digital transformation leaders can teach you about realistic artificial intelligence. Calif. Manage. Rev. (2019). https://doi.org/10.1177/1536504219865226
25. Edwards, J.S., Duan, Y., Robins, P.: An analysis of expert systems for business decision making at different levels and in different roles. Eur. J. Inf. Syst. (2000). https://doi.org/10.1057/palgrave.ejis.3000344
26. Gupta, M.: Implications of expert systems for the operations of financial institutions. Technovation (2000). https://doi.org/10.1016/S0166-4972(99)00165-0
27. Jarrahi, M.H.: Artificial intelligence and the future of work: human-AI symbiosis in organizational decision making. Bus. Horiz. (2018). https://doi.org/10.1016/j.bushor.2018.03.007
28. Du, S., Xie, C.: Paradoxes of artificial intelligence in consumer markets: ethical challenges and opportunities. J. Bus. Res. (2021). https://doi.org/10.1016/j.busres2020.08.024
29. Nassar, A., Kamal, M.: Ethical dilemmas in AI-powered decision-making: a deep dive into big data-driven ethical considerations. Int. J. Respons. Artif. Intell. **11**(8), 1–11 (2021)
30. Novelli, C., Taddeo, M., Floridi, L.: Accountability in artificial intelligence: what it is and how it works. AI & Soc. (2023). https://doi.org/10.1007/s00146-023-01635-y
31. Vesnic-Alujevic, L., Nascimento, S., Pólvora, A.: Societal and ethical impact of artificial intelligence: critical notes on European policy framework. Telecommun. Policy (2020). https://doi.org/10.1016/j.telpol.2020101961
32. Olatoye, F.O., Awonuga, K.F., Mhlongo, N.Z., Ibeh, C.V., Elufioye, O.A., Ndubuisi, N.L.: AI and ethics in business: a comprehensive review of responsible AI practices and corporate responsibility. Int. J. Sci. Res. Arch. (2024). https://doi.org/10.30574/ijsra.2024.11.1.0235
33. Newman, J., Mintrom, M., O'Neill, D.: Digital technologies, artificial intelligence, and bureaucratic transformation. Futures (2021). https://doi.org/10.1016/j.futures.2021.102886
34. Courtland, R.: Bias detectives: the research striving to make algorithms fair. Nature (2018). https://doi.org/10.1038/d41586-018-05469-3
35. O'Connor, S., Liu, H.: Gender bias perpetuation and mitigation in AI technologies: challenges and opportunities. AI Soc. (2023). https://doi.org/10.1007/s00146-023-01675-4
36. Varsha, P.S.: How can we manage biases in artificial systems—a systematic literature review. Int. J. Inf. Manag. Data Insights (2023). https://doi.org/10.1016/j.jjimei.2023.100165
37. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A.: A survey on bias and fairness in machine learning. ACM Comput. Surv. (2021). https://doi.org/10.1145/3457607
38. Rane, N.: ChatGPT and similar generative artificial intelligence (AI) for building and construction industry: contribution, opportunities and challenges of large language models for Industry 4.0, Industry 5.0, and Society 5.0. Opportunities and challenges of large language models for industry. SSRN. (2023). https://doi.org/10.2139/ssrn.4603221
39. Dietz, J., Robinson, S.L., Folger, R., Baron, R.A., Schulz, M.: The impact of community violence and an organization's procedural justice climate on workplace aggression. Acad. Manag. J. (2003). https://doi.org/10.2307/30040625
40. Sikstrom, L., Maslej, M.M., Hui, K., Findlay, Z., Buchman, D.Z., Hill, S.L.: Conceptualising fairness: three pillars for medical algorithms and health equity. BMJ Health Care Inform. (2022). https://doi.org/10.1136/bmjhci-2021-100459
41. Yu, P.K.: The algorithmic divide and equality in the age of artificial intelligence. Fla. L. Rev. (2020). https://ssrn.com/abstract=3455772
42. Rahman, A.: AI revolution: shaping industries through artificial intelligence and machine learning. J. Environ. Sci. Technol. (2023). https://doi.org/10.4236/jcc.2023.119006
43. Brendel, A.B., Mirababae, M., Lembcke, T.-B., Hofeditz, L.: Ethical management of artificial intelligence. Sustainability (2021). https://doi.org/10.3390/su13041974
44. Shin, D.: The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI. Int. J. Hum. Comput. Stud. (2021). https://doi.org/10.1016/j.ijhcs.2020.102551
45. Passey, D., Shonfeld, M., Appleby, L., Judge, M., Saito, T., Smits, A.: Digital agency: empowering equity in and through education. Technol. Knowl. Learn. (2018). https://doi.org/10.1007/s10758-018-9384x
46. Marikyan, D., Papagiannidis, S., Rana, O.F., Ranjan, R., Morgan, G.: Alexa, let's talk about my productivity": the impact of digital assistants on work productivity. J. Bus. Res. (2022). https://doi.org/10.1016/j.jbusres.2022.01.015
47. Ylisassi, H., Hasu, M., Heikkilä, H., Käpykangas, S., Saari, E., Seppänen, L., Valtanen, E.: Työntekijöiden kehittämistoimijuutta edistämässä. Kehittämismenetelmäkokeilujen tuloksia vanhuspalveluissa (2016)
48. Sannino, A., Engeström, Y.: Rational agency, double simulation, and the object of activity. An intervention study in a primary school. In Edwards, (ed.) Working Relationally In and Across Practices: A Cultural-Historical Approach to Collaboration, pp. 43–57. Cambridge University Press, New York (2017)
49. Eteläpelto, A., Vähäsantanen, K., Hökkä, P., ja Paloniemi, S.: Tutkimus- ja kehittämishankeen tausta ja lähtökohdat. In: Teoksessa K., Vähäsantanen, S., Paloniemi, P., Hökkä, P., Eteläpelto, A. (eds.) Ammatillinen toimijuus: rakenne, mittari ja tuki, pp. 5–13. Jyväskylän yliopisto, Jyväskylä (2017). http://urn.fi/URN:ISBN:978-951-39-6980-6
50. Vähäsantanen, K., Paloniemi, S., Räikkönen, E., Hökkä, P., Eteläpelto, A.: Ammatillisen toimijuuden moniulotteinen rakenne ja mittarikehittely. In: Teoksessa K., Vähäsantanen, S., Paloniemi, P., Hökkä, P., Eteläpelto, A. (eds.) Ammatillinen toimijuus: rakenne, mittari ja tuki, pp. 13–33. Jyväskylän yliopisto, Jyväskylä (2017). http://urn.fi/URN:ISBN:978-951-39-6980-6
51. Virkkunen, J.: Dilemmas in building transformative agency. Activ. Rev. Electronique (2006). https://doi.org/10.4000/activites.1850
52. Giddens, A.: The Constitution of Society: Outline of the Theory of Structuration. Polity Press, Berkeley and Los Angeles (1984)
53. Eteläpelto, A., Vähäsantanen, K., Hökkä, P., Paloniemi, S.: Miten käsitteellistää ammatillista toimijuutta työssä? Aikuiskasvatus (2014). https://doi.org/10.33336/aik.94100
54. Emirbayer, M., Mische, A.: What is agency? Am. J. Sociol. (1998). https://doi.org/10.1086/231294

55. Koivisto, T.: Digitoimijuus Terveydenhuollon ammattilaisen työssä. Tampereen yliopisto (2023)
56. Haapasaari, A., Engeström, Y., Kerosuo, H.: The emergence of learners' transformative agency in a change laboratory intervention. J. Educ. Work. (2016). https://doi.org/10.1080/13639080.2014.900168
57. Pemer, F., Werr, A.: Defusing digital disruption through creative accumulation: technology-induced innovation in professional service firms. J. Manage. Stud. (2023). https://doi.org/10.1111/joms.12972
58. Siau, K., Wang, W.: Artificial intelligence (AI) ethics: ethics of AI and ethical AI. J. Database Manag. (2020). https://doi.org/10.4018/JDM.2020040105
59. McNamara, A., Smith, J., Murphy-Hill, E.: Does ACM's code of ethics change ethical decision making in software development? In: Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, pp. 729–733 (2018). https://doi.org/10.1145/3236024.3264833
60. Morley, J., Kinsey, L., Elhalal, A., et al.: Operationalising AI ethics: barriers, enablers and next steps. AI Soc. (2023). https://doi.org/10.1007/s00146-021-01308-8
61. Liao, Q.V., Gruen, D., Miller, S.: Questioning the AI: informing design practices for explainable AI user experiences. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–15 (2020). https://doi.org/10.1145/3313831.3376590
62. Meijer, A., Lorenz, L., Wessels, M.: Algorithmization of bureaucratic organizations: using a practice lens to study how context shapes predictive policing systems. Public Admin. Rev. (2021). https://doi.org/10.1111/puar.13391
63. Hu, L., Bentler, P.M.: Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Modeling (1999). https://doi.org/10.1080/10705519909540118
64. Sathyanarayana, S., Mohanasundaram, T.: Fit indices in structural equation modeling and confirmatory factor analysis: reporting guidelines. Asian J. Econ. Bus. Account. (2024). https://doi.org/10.9734/ajeba/2024/v24i71430
65. Aagaard, T., Lund, A.: Digital Agency in Higher Education: Transforming Teaching and Learning. Routledge, London and New York (2020)
66. Alasoini, T., Ala-Laurinaho, A., Känsälä, M., Saari, E., Seppänen, L.: Työelämän digikuilujen yli: digitalisaatio kaikkien kaveriksi. Työterveyslaitos. https://www.julkari.fi/bitstream/handle/10024/143939/TTL-978-952-261-997-6.pdf?sequence=1&isAllowed=y (2022)
67. Raisch, S., Krakowski, S.: Artificial intelligence and management: the automation-augmentation paradox. Acad. Manag. Rev. (2021). https://doi.org/10.5455/amr.2018.0072
68. Hu, L., Bentler, P.M.: Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Modeling (1999). https://doi.org/10.1080/10705519909540118
69. Sathyanarayana S., Mohanasundaram, T.: Fit indices in structural equation modeling and confirmatory factor analysis: reporting guidelines. Asian J. Econ. Bus. Account. (2024). https://doi.org/10.9734/ajeba/2024/v24i71430