

AI-Enhanced Career Guidance Systems as Intelligent Labour-Market Intermediaries for Reducing Skills Mismatch: A Critical Integrative Review and Conceptual Framework

Dr. Nosheen Malik^{1*}

¹ Assistant Professor, Department of Education, The Islamia University of Bahawalpur, Punjab Pakistan, nosheen.malik@iub.edu.pk, ORCID ID: <https://orcid.org/0009-0007-5278-415X>

*Corresponding Author Email: nosheen.malik@iub.edu.pk

DOI: <https://doi.org/10.70670/sra.v3i2.2468>

Abstract

The persistent skills mismatch (the gap between the skills of the workforce and the skills needed to fulfil job requirements) has adverse effects on productivity, job mobility and the inclusiveness of technological transition. While there is a conceptual disconnection and potential ethical issues in how AI can support career guidance, its potential benefits for timely and personalized career guidance are promising. This paper reviews the concept of AI enhanced career guidance as a smart labour-market intermediary and proposes an integrated framework to understand the interactions between the emergence of AI and the identification, anticipation, and mitigation of skills mismatch. A critical integrative review that combines literature published primarily from 2015 to 2025 from the literature sources of labour economics, career development, educational technology, information systems, and responsible AI, while maintaining the basic theoretical literature sources. Thematic synthesis and mechanism-oriented synthesis were used to analyse evidence. AI can boost competency profiling, skills intelligence, occupational scenario analysis, personalised learning pathways, and learner/counsellor/institution/employer/workforce agency coordination. It is most useful when it helps to lessen information asymmetry and can facilitate adaptive career agency instead of providing prescriptive job advice. Data quality, digital accessibility, AI literacy, institutional preparedness, labour-market structure and governance moderate effectiveness. Biases, privacy violations, hallucination, automation bias, opaque recommendations, and unequal infrastructure can perpetuate or exacerbate the existing disadvantage. The paper argues for a human-in-the-loop, governed environment where AI gets the analytical and informational tasks that scale through the number of cases, but the role of counsellors remains open to the contextual interpretation, relational support, ethical judgment and accountability. It provides principles for implementation, a matrix of human and machine roles to allocate, and a research agenda of priority emphasis on longitudinal, cross-national, and equity-sensitive evaluation.

Keywords: Artificial Intelligence, Career Guidance; Skills Mismatch, Labour-Market Intelligence, Workforce Development; Human–AI Collaboration

Introduction

Skills mismatch as a structural labour-market problem

Skills mismatch refers to an imperfect match between skills people have, qualifications they hold, the tasks they engage in and what is needed for the jobs on offer. It covers vertical mismatch (overqualification or underqualification), horizontal mismatch (working outside the field of education or training), skill shortages, skill gaps within jobs, skill underutilisation and skill

obsolescence. Do not consider these forms interchangeable—there are different causes, measures, and policy implications associated with each of them. However, they all have one thing in common: the inefficient utilization and development of human capability (ILO & OECD, 2023b; McGuinness, Pouliakas, & Redmond, 2018). Mismatch can also arise from factors influencing the demand side, such as a lack of workers or education systems, as well as geographic, financial, occupational licensing, recruitment conventions, discrimination, lack of labour-market information, and access to continuing education. Search-and-matching theory thus has a hot focus on information quality and transaction cost as key factors to shape labour-market allocation, while institutional perspectives focus on firms, education suppliers, public employment services and social protection systems (Bobba et al., 2021). Further evidence also suggests that mismatch is linked to lower productivity and lower skills utilisation, both of which persist across the board (Adalet McGowan & Andrews, 2015). Digital and green transitions exacerbate this issue as the pace of change within the technologies will impact task content before formal qualifications or curricula can be redesigned. A task-based approach reveals that technology can do multiple things, it can replace certain tasks, it can enhance other tasks, and it can create new tasks and occupations; the distributional effects of technology are determined by institutions, investments and the trajectory of innovation, not just technology (Acemoglu & Restrepo, 2019). Rather than a single matching between occupations and skills, recent international evaluations have focused on resilience, adult learning, transferable skills and skills intelligence (OECD, 2023b; Cedefop, 2023).

Why career guidance matters in a volatile economy

Career guidance is a process that equips and supports individuals to understand their skills and values, to explore further education and employment options, to make choices, and to plan for transitions. Its economic role is to enhance information and matching, and its developmental role is to build agency, self-efficacy, adaptability and the ability to build meaningful careers. However, traditional services are limited because of counsellor workload, inequitable access, patchy sources of information, and the speed at which labour-market information can be made redundant before reaching the users. Rich skills can be harvested from unstructured data, competency can be compared between taxonomies, skills can be identified in job posts as they evolve, learning resources can be recommended, alternative transition pathways can be suggested, and guidance can be provided through conversational interfaces, which are all difficult to deliver at scale. Previous research on career guidance has shown the potential for course and job recommendations, skill-profile development, service triage, and highlighted questions of agency, data ecosystems, professional roles, and ethics (Westman et al., 2021). A more comprehensive interdisciplinary review also finds that AI is starting to influence career exploration, planning, and transitions throughout the life span, though there is inconsistent evidence of sustainable and equitable results (Bankins et al., 2024). Rather, the critical question is: when, if at all, does AI content serve as a reliable facilitator for information sharing between people, education, employers, and labour market sectors? Choices for careers are consequential, identity-relevant and are shaped within the context of family, culture, gender, disability, geography, and economic need. Such a technically correct recommendation can be unsuitable if it fails to take these contexts into consideration, or if it limits instead of amplifies agency.

Research gap, purpose, and questions

While the study of AI and career is increasing, three gaps are still present. First, technical research on recommender systems usually assesses the predictive accuracy but fails to link it with career development theory and labour-market matching. Second, in career research, AI is often presented as a contextual shift rather than concrete details of the data, algorithms, and institutions involved

in AI-driven guidance systems. Third, responsible-AI scholarship focuses on issues of fairness, privacy, transparency and accountability, but it doesn't usually develop into a realistic governance framework for career services. This paper aims to tackle those gaps by theorizing career guidance that relies on artificial intelligence as a governed labour-market intermediary. It combines labour economics, Social Cognitive Career Theory (SCCT), career construction theory, human-centred AI, and responsible-AI governance to capture opportunity and risk. There are 4 questions that guide the synthesis:

- RQ1. How can AI-enabled career advice help to mitigate, identify, and predict skills mismatch?
- RQ2. What are the technological, organisational, institutional and contextual factors that shape the effectiveness of the system?
- RQ3. What should be outsourced to AI and what should be done by human career professionals?
- RQ4. What governance safe guards are necessary to promote equity, respect rights and maintain user agency?

Theoretical Foundations

Labour economics: human capital, task change, and matching frictions

Human capital theory describes education/training as investments that are supposed to boost productivity and future earnings. If we assume the simple equilibrium notion, the price signals would be expected to affect the supply of labor and the demand for labor should respond to match supply and need. This account is inadequate, because there are significant and persistent mismatches: incomplete information, credit constraints, mobility constraints, uncertain returns to training, institutional segmentation, long time lags between decisions to train and entry into the labor market. Task-based theories offer a more dynamic approach. While new automation technologies tend to shift the distribution of labour between labour and capital, they can also generate new labour-intensive tasks that come alongside the new machines (Acemoglu & Restrepo, 2019). This means that you do not want to label whole occupations as 'safe' or 'at risk' when giving career advice. It should assist users to see which tasks are shifting, and which competencies will be transferable, and which learning investments will provide options to multiple plausible futures. The ILO's work on occupational exposure also differentiates between technological exposure and irreversible unemployment, while pointing to the fact that transformation rather than replacement is more frequent than is often thought (ILO, 2025). Search and matching models reveal the presence of information asymmetry, search cost, and imperfect signalling as causes of misallocation. While it can help to ease these tensions by making skills and requirements easier to understand, it can also lead to new tensions when scores are hidden from view or when informal and context-specific skills are not included. The intermediary role should be assessed not only on the basis of predictive accuracy, but also on the extent to which it enhances the diversity, quality, and reversibility of choices.

Career development: agency, self-efficacy, and adaptability

SCCT describes career interests, choices and performance as a result of the interactions between self-efficacy, outcome expectations, goals, learning experiences and contextual support or barriers (Lent, Brown, & Hackett, 1994). In this context, an AI system can support career development by increasing the awareness of realistic opportunities, offering mastery feedback, identifying supports and barriers, and allowing users to establish and modify goals. It stifles development when it gives recommendations as a final judgment or when it translates inequality from the past as an individual's prediction. The theory of career construction emphasizes the psychosocial resources

that are important, resources that focus on how to cope with transitions and disruptions, known as career adaptability (Savickas & Porfeli, 2012). When labour markets are volatile, being adaptable is a more secure guidance outcome than only one predicted occupation. AI should then be used to help explore scenarios, to compare with reflection, and to develop transferable skills. Positive career orientations are also supported by evidence of the career adaptability and career decision self-efficacy, as developmental outcomes to pay attention to (Chui, Li, & Ngo, 2022). The combined logic of SCCT and career construction theory suggests a principle of augmented agency: technology should increase the user's understanding and agency, but leave the choice for questions, rejection, or reinterpretation of these capabilities. Personalisation is beneficial only if it empowers your child to make their own decisions, not direct their behaviour.

Human-centred and responsible AI

Human-centred AI does not require human or machine control, nor does it demand full automation. When systems are designed to be reliable, safe, transparent, and user-masters, there can be high levels of automation and meaningful human control (Shneiderman, 2020). Equivocal decision making is a very relevant example, as AI can process a vast amount of data and pick up on patterns while humans can bring in contextual judgement, moral reasoning, empathy and accountability (Jarrahi, 2018). Responsible-AI frameworks bring these principles into governance. UNESCO's Recommendation on the Ethics of Artificial Intelligence focuses on human rights, dignity, fairness, transparency, accountability, privacy, and human oversight (UNESCO, 2021). The NIST AI Risk Management Framework outlines practice in governance, mapping, measuring, and managing risk and trustworthiness characteristics are: valid, reliable, safe, secure, explainable, enhances privacy, and mitigates harmful bias (Tabassi, 2023). The EU Artificial Intelligence Act introduces new legally binding requirements that apply to specific education and employment use cases, such as the use of stringent controls for them (European Union, 2024). Career guidance sits between education, work and life. A system can be low-risk, if it simply accesses information on occupations that is publicly available, but it can become high-risk if it profiles a person, infers attributes about the person, ranks opportunities, or affects access to education and employment. Governance should thus be tailored to the use case and not be synonymous with the term 'AI' in general terms.

Table 1. Forms of skills mismatch and relevant guidance responses

Mismatch form	Illustration	Potential AI contribution	Principal caution
Vertical mismatch	Qualification level exceeds or falls below job requirements	Compare skills evidence with task requirements; identify progression or bridging routes	Qualifications are imperfect proxies for competence and job quality
Horizontal mismatch	Field of education differs from occupational field	Map transferable skills and adjacent occupations	Field mismatch may be voluntary or beneficial; avoid deficit assumptions
Skill gap	Worker lacks capabilities needed in a current or target role	Generate gap profiles and learning plans	Taxonomies may be incomplete; self-report and inferred skills require validation
Skill shortage	Employers cannot recruit required capabilities	Aggregate demand signals and inform programme planning	Job-posting data reflect employer behaviour, not necessarily socially optimal demand

Mismatch form	Illustration	Potential AI contribution	Principal caution
Skill underutilisation	Capabilities are not used in current work	Identify roles or task redesign using existing strengths	Mobility constraints and discrimination may block recommended moves
Skill obsolescence	Technological or organisational change reduces value of existing skills	Monitor task change and propose reskilling scenarios	Forecasts are uncertain and can become rapidly outdated

Conceptual Framework: AI-Enhanced Guidance as a Labour-Market Intermediary

The proposed framework views career guidance as an ecosystem and not a standalone chatbot or prediction model. It links four functional layers—data and competency profiling, labour-market intelligence, explainable guidance and pathway generation, and human decision and action. These layers generate proximal developmental outcomes and over time, may lead to better education-work linkages. The whole process is surrounded by contextual moderators and responsible-AI controls (Figure 1).

AI-enhanced career guidance as a governed labour-market intermediary

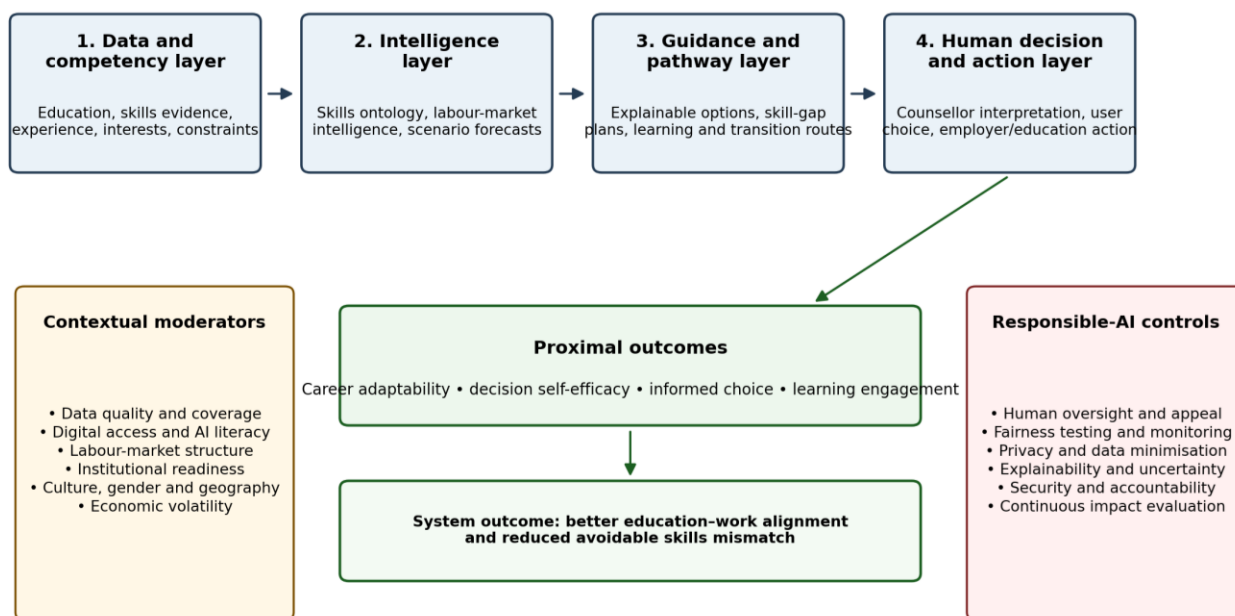


Figure 1. Conceptual framework for governed AI-enhanced career guidance. The model emphasises iterative feedback, contextual moderation, human judgement, and accountable governance rather than deterministic occupational prediction.

Layer 1: data integration and competency representation

The first layer is an "evidence-based but contestable" representation of the user. This can encompass qualifications, Micro-Credentials, verified Micro-Credentials, work experience, volunteering experience, personal portfolios, assessment, interests, accessibility requirements, geographical limitations and career aspirations. Skills can be gleaned from unstructured records and aligned with occupational taxonomies using NLP. Inferred attributes should never be assumed to be the ground truth, however. The opportunities for users to examine, correct, add and remove information about themselves. It is crucial to use data minimisation. Inferences based on personality, demographic, biometric, emotional and health data may lead to disproportionate

privacy and discrimination. A career system ought to gather just the variables that are essential for a defined function, and the collection of variables for user-controlled individualisation is not the same as the collection of variables for monitoring fairness, which is protected.

Layer 2: skills and labour-market intelligence

The second layer connects individual evidence to information on occupations, tasks, skills, wages, job openings, educational pathways, and opportunities in the region. They can be official occupational classifications, graduate outcomes, employer surveys, job postings, professional standards and skills forecasts. Cedefop's skills-intelligence infrastructure highlights the possibility for combining forecasts, vacancy analysis online and mismatch indicators whilst keeping methodological documentation (Cedefop, 2024). Real-time data don't necessarily have to be better. Online vacancies overrepresent formal jobs, digitally advertised vacancies, and can underrepresent small firms, informal work, rural jobs, and occupations recruited via networks. Forecasts should thus be given in terms of ranges and scenarios, and the date of the forecast, the geographic scope of the forecast, and uncertainty should be explicitly stated.

Layer 3: matching, explanation, and pathway generation

The third layer produces multiple likely solutions (not a single 'best' answer). Current skill overlap, learning distance, user values, accessibility, expected demand, job quality, geographical feasibility, and transition cost are some of the factors a multi-criteria model might include. Every response should include an explanation of the evidence used, assumptions made, what is not known and what alternative route was considered. Pathway generation is a process of turning an occupational option into action: modules, projects, mentoring, work-based learning, credentials, financial support, and milestones. In higher education, evidence of adaptive systems, with respect to learning sequence and pacing, suggests potential for enhanced performance, engagement, and some limited evidence for the long-term impacts on career outcomes (du Plooy, Casteleijn, & Franzsen, 2024).

Layer 4: human interpretation, choice, and institutional action

Recommendations can only be turned into guidance through interpretation and action. A pathway is what the user chooses to embrace based on their identity, responsibilities, risk tolerance and aspirations. Counsellors can help to bring up assumptions, deal with anxiety, challenge stereotyped suggestions, and relate information to reality. Institutions and employers are then required to make genuine opportunities – courses, placements, job change, reasonable adjustments. If these responses are not present, an AI tool cannot reduce mismatch; it only describes it. Feedback should be both model-performance feedback and human feedback. Some helpful factors are profile corrections, option diversity, understanding of users, decision self-efficacy, learning actions completed, transition quality, job quality, and differences between groups. It is possible to incentivise any persuasive but damaging suggestions by optimising CTR or acceptance rates.

Mechanisms of mismatch reduction

- Information-asymmetry reduction: increasing the visibility of skills, requirements and routes for individuals and institutions.
- Anticipatory development: Identifying the set of changing task requirements at an early stage so that learning can take place in anticipation of them.
- Transferability recognition: discovery of matching occupations and skills across the bridges that are not covered by traditional qualification matching.
- Personalisation for development: adapting exploration and action, enhancing adaptability and self-efficacy.

- Institutional coordination: consolidating anonymous signals to enhance curriculum, employer engagement and workforce planning.
- Access expansion: Providing initial information and navigation beyond the "office hours" and location, which is actively dealing with digital exclusion.

Methodology

Review design

A critical integrative review and conceptual synthesis have been used in the paper. Given that the evidence comes from diverse empirical research, ideas, technical prototypes, policy documents, and governance standards that are not suitable for meta-analysis, this design is appropriate. The goal is to build a theory and to clarify mechanisms, not to calculate an overall intervention effect.

Search scope and sources

This review was based on literature from 2015 to 2025, but also on older literature related to career theory, labour economics and human centric computing. Authoritative sources from the ILO, OECD, Cedefop, UNESCO, NIST, and the European Union were included in search domains, along with Scopus, Web of Science, ERIC, IEEE Xplore, ACM Digital Library and ScienceDirect. Search strings included the terms for career guidance and recommender systems with AI, machine learning, natural language processing, labour-market intelligence, skills mismatch, employability, fairness, privacy and human–AI collaboration. A representative Boolean structure was: ("career guidance" OR "career counselling" OR "career recommender" OR "vocational guidance") AND "artificial intelligence" OR "machine learning" OR "large language model" OR "natural language processing") AND "skills mismatch" OR employability OR "occupational matching" OR "labour market"). There was a search for foundation and related studies using backward and forward citation searching.

Eligibility and critical appraisal

Included sources: Provided career exploration/recommendation support via AI; Extracting skills from content, matching with occupations, or providing labour-market intelligence; Career-development outcomes; Responsible-AI risks in career, education or workforce services; Implementation in career, education or workforce services. Any purely technical papers without an appropriate guidance application were not included, nor were promotional vendor materials or unsupported opinion pieces. Official documents were kept when they provided definitions, comparative indicators, policy requirements or documented policy frameworks. Relevance, methodological transparency, sample adequacy, ecological validity, clarity of outcome measurement and applicability across contexts were appraised. The technical accuracy measures were not considered to be equal to guidance effectiveness. The research that reported the user outcomes, implementation conditions, and fairness evaluation, or real-world deployment, received particular weight.

Synthesis procedure

The analysis was a combination of thematic coding and mechanism-oriented synthesis. They were organised into the following studies on competency assessment, labour-market forecasting, recommendation and adaptive learning, human–AI collaboration, equity and bias, privacy and governance, institutional readiness and cross-national transferability. The claims were contrasted with one another between disciplines, contradictions were not eliminated with the aid of averaging, and the claims evolved through the four-layer framework.

Findings and Critical Analysis

Competency profiling and skill-gap identification

AI can enhance the level of granularity of skills representation by pulling evidence from CVs, portfolios, course records and work history and attaching this evidence to structured occupational taxonomies. This may uncover skills acquired through non-linear career paths, projects, volunteering and informal work, in comparison with qualification-only matching. Research on career recommender also shows that there is a growing trend of personalisation of options using content-based, collaborative, knowledge-based, and hybrid approaches (Mavuso, Jere, & Van Greunen, 2023). Construct validity is a major drawback. A skill listed in a document does not constitute demonstration of that skill at the appropriate level and absence from a digital record does not constitute absence of that person. Models can confuse 'polished language' and competence, enable 'users' to access credentialing and/or portfolio support, or under-recognise culturally different descriptions of experience. Automated extraction and user verified, performance supported and counsellor reviewed, therefore results in quality improvements. The creation of taxonomic design is also of critical importance. Occupational standards are not fine enough for the varying tasks that may occur in any job; job descriptions are noisy, are inflated, and/or may be discriminatory when created by the employer. Versioned taxonomies, multiple standard mapping, provenance, and additions from the local level should be included within the system. They should also be able to separate competence, potential for competence, interest, and constraint and not combine them into one 'fit' score.

Labour-market intelligence and occupational forecasting

With the help of machine learning and text analytics, a massive amount of vacancy and administrative data can be fed into the system to detect shift in skill language, occupational groups, and/or the demand in specific regions. These functions can reduce the time lag between labour-market change and content on guidance. Skills forecasting and vacancy analytics for skills via the internet already assist policy and vocational education planning in Europe; international organisations have been starting to create a task-level analysis of AI exposure alongside the occupational classification (Cedefop, 2023; ILO, 2025). Forecasts remain susceptible to self-fulfilling and distribution shift errors, however. Policy change, recession, conflict, migration, climate events and technological discontinuity may kill historic demand. Vacancy data reflect the preferences of employers and not necessarily a requirement for or desired performance. When institutions replicate curricula in a mechanical, prescriptive way in terms of the jobs available, they can perpetuate the short-term goals of employers and ignore the learning that is civic in nature, fundamental, and transferable. Using it as scenario intelligence is a more responsible use. There should be several horizons, confidence ranges, data dates and alternative assumptions in systems. When there is uncertainty, guidance should focus on transferable skills and option value. It is important to avoid making point forecasts without specifying the geography, source, limitations of the model and the difference between the demand for the occupation as a whole and the probability of an individual being in that occupation.

Personalised pathways and adaptive support

AI can align a target pathway to courses, micro-credentials, work experience, projects, and support services. With generative models, technical information can be converted to accessible language, interviews can be simulated, reflective prompts can be provided, and iterative feedback can be given. Adaptive learning research indicates that personalisation has a positive impact on engagement, if it is based on diagnostic evidence and pedagogical design (du Plooy et al., 2024). However, the potential for epistemic mistrust occurs when there is generative fluency. Occupational, legal, salary, or qualification details generated by a LLM may be plausible, but are

not accurate; outputs may also vary between prompts and model versions (Kasneci et al., 2023). For factual claims, therefore, retrieving information from curated, dated sources, and validating them according to a rule is preferable. A career assistant is to cite sources, express uncertainty, deny any unbacked conclusions, and refer complex cases to a professional.

Personalisation should not be too individualistic in relation to structural problems. The re-skilling recommendation may suggest that the person is to blame for the malfunction of the institution if it is repeated, particularly regarding discrimination, job creation, transport safety, and care options. When reskilling is repeated as a recommendation for individuals who are discriminated against, have weak job creation, lack transport safety, or lack care options, it may suggest that the person is responsible for the failure of the institution. Effective guidance will include identification of barriers, linkages to financial and social support and feedback of aggregated evidence to institutions and policy makers.

Algorithmic bias and distributive effects

Historical data, labels, sampling, proxies, feature engineering, optimisation objectives, interfaces and deployment practices can introduce bias. However, algorithmic hiring research reveals that fairness cannot be simply translated into the absence of protected attributes: variables that are not the focus of the algorithm can also be detrimental to fairness, and various fairness requirements may contradict each other (Fabris et al., 2025; Rigotti & Fosch-Villaronga, 2024). Career systems are at a similar risk, as their suggestions can impact aspirations first before a formal selection is made. If a model is trained with the history of segregated labour, it may draw the same conclusions about inequality as preference or aptitude and steer women, minorities, disabled users, rural residents or low-income learners toward narrower or lower-paid opportunities. The debiasing approach alone is not enough as user studies suggest that users might prefer stereotyped recommendations if they are in line with social patterns they are used to (Wang et al., 2021). So there is a need for technical mitigation, critical career education, inclusive opportunity design, and down stream outcome monitoring for fairness. Developers need to do at least the following tests: representation, calibration, error rates, option diversity, recommendation rank, and adverse impacts on relevant groups and intersections. An alternate protected attribute or proxy can be counterfactual tested to determine if changing the attribute or proxy alters a recommendation for no good reason. It is necessary to carry out a qualitative evaluation, with participation of the communities affected, when assessing fairness, as statistical parity may be masking an important and dangerous assumption or, conversely, an unobtainable pathway.

Privacy, security, and informational self-determination

Educational records, employment history, psychological testing, disability information, family barriers, place of residence, and inferred interests are all examples of the types of information that can be included in a career guidance system. The linking of these data could result in the formation of a profile of the person's life that could be used in ways detrimental to education, working life, insurance, or reputation. In addition to data breaches, there are other privacy risks, such as function creep, secondary use, indefinite retention, inappropriate inference, and loss of a person's ability to control how they are represented. Purpose limitation, data minimisation, separation of identifiers from analytical data, encryption, role-based access, retention schedules and deletion mechanisms are all features of a defensible architecture. Consent is necessary and must be voluntary in case of unequal power or where access to service is dependent on consent. Institutions must have a lawful basis, privacy impact assessment, vendor due diligence, incident-response plan and an understanding of who is the controller and who is the processor. Users must have access to the primary data used, be able to fix them, ask for human input and object to any consequential recommendations. There may be some sensitive attributes required for fairness auditing purposes,

but they should be rendered sensitive and kept apart from operational recommendation features, unless there is a justified and lawful reason. The system cannot assume emotional state, personality, health status or protected identity based on the fact that a model can make an assumption.

Human–AI collaboration: allocation of functions

The evidence is strong to augment, not replace. AI can support data-rich, repetitive, and repetitive functions, such as: Retrieving occupational information, comparing skills, creating alternative scenarios, monitoring milestones, and identifying inconsistencies. Human professionals are essential when the work is not explicit, not devoid of sentimentality, not clearly related to an identity, not morally or ethically laden, or when information not observable by the system is vital to the task. There's still the possibility of automation bias. An apparent 'human in the loop' does not ensure meaningful oversight if counsellors are under time pressure, lack training, authority or access to the reasoning in the model. Oversight should be built around active verification: individuals should be able to access evidence, edit and/or refuse outputs, justify when and why, and raise issues of systemic error. They need to track if AI alters counselor conduct, rather than just if a human pressed APPROVE. The most efficient allocation of labor is different for different stages, and different for different users. The digitally confident user is likely to require minimal professional engagement when accessing public information. An intense human support may be needed for a student in a family conflict, a worker who has lost his or her identity, or an individual who is being discriminated against. A single workflow is not the most efficient approach for systems, as it fails to take into account individual risk levels, nor is it the most effective way to provide graduated support.

Table 2. Recommended allocation of guidance functions

Function	AI-leading role	Human-leading role	Required safeguard
Information retrieval	Search, summarise, update and compare documented sources	Check relevance and explain implications	Source citation, date stamping, hallucination controls
Competency profiling	Extract and map possible skills	Validate evidence and contextualise proficiency	User correction, provenance, no hidden sensitive inference
Option generation	Produce diverse scenarios and adjacent occupations	Explore meaning, feasibility, identity and constraints	Diversity requirements; no deterministic ranking
Learning plan	Sequence resources and milestones	Adapt to motivation, accessibility, finance and wellbeing	Verified providers; cost and time transparency
High-stakes decision	Provide evidence and uncertainty	Retain judgement, consent and accountability	Mandatory human review and appeal
Emotional/ethical support	Recognise need for escalation only	Provide empathy, safeguarding and advocacy	No simulated therapy; referral protocols
Institutional intelligence	Aggregate trends and gaps	Decide policy and resource response	De-identification, representativeness and governance

Organisational readiness and implementation capacity

Adoption involves more than software; it requires a variety of other factors. Institutions should have a well-defined challenge, solid data stewardship, professional ownership, technical integration, procurement capability, staff training and tools to evaluate and maintain. Common pitfalls in pilot projects include starting with a model first, lack of responsibility coalescing, or a service redesign that begins first. We should first start implementing some functions and tasks that are relatively simple and high-gain, for example, curated information retrieval, appointment triage, or counsellor facing summaries. Evidence and governance are needed to support more consequential uses like profiling, ranking, prediction or automated learning prescriptions. Data portability, audit access, model updates, access, language support and exit rights should be part of any build versus buy decision. Professional capability is part of the essential conditions. Career practitioners need a level of AI literacy that enables them to understand how data relates to uncertainty, bias, and proper challenge. To prevent the technical teams from simplifying complex decisions into similarity scores, the need for career development knowledge is crucial. When co-designed with students, workers, counsellors, employers, and marginalised groups, it demonstrates additional harms that will not be picked up using performance testing alone, and improves fit.

Cross-national transferability and digital inequality

AI career systems are not neutral in respect to culture or institution. The typologies of jobs, qualification frameworks, pathways to enter occupations, expectations of family members, gender norms, social protection provision and informal employment vary from country to country. Especially in less-formal work settings, in low-tech environments, and in less developed countries, a system which has been trained using formal job postings in high-income regions may not provide optimal guidance. Digital inequality relates to devices, connectivity, language, disability access, data literacy, and confidence with challenging automated advices. These every factor may turn a seemingly scalable service into an additional exclusion mechanism (Assefa et al., 2025). Equity means that there must be ways of communicating that require less bandwidth or are mobile-friendly; the design must be accessible; communication materials must be available in more than one language; access must be assisted; there must be alternatives to digital communication, and human services must be in addition to rather than a complete replacement for technology. The steps of cross-national adaptation should incorporate local labour market data, local experts, user testing and assessment of outcomes for subgroups. Transferability is not an assumption, but is an empirical question. They can have the same architecture and governance standard with different taxonomies, datasets, explanations, referral, policy constraints etc.

Overall strength of evidence

There has been more research to support the technical feasibility than the social effectiveness. Numerous studies measure decisions made by users or satisfaction with the prototype or engagement over a short time frame only, and relatively few investigate if users make better decisions, learn relevant skills, create better products, decrease mismatch over time, or input better work. Findings are also focused in higher education and formal labour markets, and missing for some adult career changes, informal labour markets, disability, rural areas, and low resource areas. But, the argument that AI fixes skills mismatch is still unconvincing. Pitfalls that create mismatch are also job quality and wage setting, discrimination, regional development, social protection and provision of appropriate work, for which AI can help to optimize. One intermediary in a larger system of transformation of skills depends on the success of intermediary success skills services, and intermediary success skills services depends on the response of institutions to the information generated.

Discussion

Answering the research questions

RQ1: How can skills mismatch can be minimised with the guidance improved with AI? These mechanisms introduced in the framework are: reduction of information asymmetry, forecasting new gaps, identification of transferable competences, individualisation of the developmental supports, amplification of access, and coordination of institutions. These are possible and in part validated in terms of the quality of the information, its actionability and the presence of genuine education and employment opportunities as communicative tools.

RQ2 was related to conditions of effectiveness. Outcomes are intensively moderated by data quality, representativeness, digital access, AI literacy, labour-market stability, organisational readiness, professional capability and governance. In some cases, the same model can be effective and in others, detrimental. The context, implementation, and equity should, therefore, be addressed by a system evaluation rather than just being viewed as inherent by examining accuracy.

Human–AI collaboration was explored in RQ3. The scale is graduated augmentation: AI is for scalable analysis tasks, while humans are for relational, contextual, deontic and judgement tasks of high value. Human control should not be delegated, it have to be resourced and empowered. Quantitative measures, such as the percentage of counselling procedures automated, are not to be used to determine the success of the combined service, but rather how the combined service has increased user agency, understanding and quality of the transition.

RQ4 addressed governance. To be implemented responsibly, the purposes must be limited, data minimized, explainable, uncertainty communicated, fair assessed, secure, subject to human review, subject to appeal, accessible, and subject to ongoing effects monitoring. Between the stages of a problem being defined and a product being taken out of service, governance shouldn't just be carried out once a product or service is bought in accordance with standards, but there should be a full governance cycle.

Theoretical contribution

The paper is an extension of the labour-market matching theory by designating a digital intermediary that conveys more than just vacancy information. It's a skills booster, it creates scenarios, organizes learning and returns aggregated intelligence to educational organisations. Meanwhile, arguments in career-development theory rule out turning the intermediary into an optimisation machine. Self-efficacy, contextual barriers, adaptability and meaning are situated at the centre of the theory of effectiveness, which is found in SCCT and in career construction theory. The key contribution of the resulting concept is the theory of governed augmented agency. Where appropriate, AI can be useful in extending the person's evidence and options, and in their capacity to act, and things that can be managed with governance mechanisms are constraining inappropriate inferences and institutional power. This is not the same as it is in techno-optimistic automation and general rejecting. It also highlights that personalisation and fairness can be a trade-off; for example, while the most accurate recommended action (considering past experience of users) may be more personalised, it could be less fair than a more broadly prescribed one that is guided towards opportunity extension.

Practical contribution: from recommendation engine to ecosystem

The most common prototypes are those that are based on a limited user/algorithm interface. The ecosystem model introduces counsellors, education providers, employers, public agencies, regulators and community organisations. It's important because it's about resources and opportunities that are required for recommendations. A system that can locate a nursing pathway but are unable to link to prerequisite education, finances, placement need, childcare and/or licensing or support is an information providing system without capability. The Learning

Functionality for the ecosystem is also achieved. Data can be aggregated and protected to expose predictable gaps, limited access, local or regional shortages, or groups that are systematically over or underserved in curriculum offerings. This intelligence must be used for redesign of services – but it is not surveillance and is not a performance management tool for the individual.

A Responsible Implementation Model

Seven implementation principles

- Begin with a clear service issue and a theory of change rather than a technology preferred.
- Use the least intrusive and the least autonomous handling system that will accomplish the desired outcome.
- From the start, think in terms of a design that will be different, that people can correct, explain, lack confidence in, and appeal to.
- Ensure humans have time, authority and training to intervene in high-stakes interpretation assignments and hold them accountable.
- Take a more holistic perspective on developmental and distributive results—now looking beyond accuracy, efficiency, or user satisfaction.
- Co-design with affected users and representation of marginalised groups in impact review, governance and testing.
- View deployment as an ongoing process of a socio technical experiment, including a monitoring regime, an incident reporting framework and an exit plan.

Lifecycle governance

The NIST RMF model has four iterative functions; govern, map, measure, and manage (Tabassi, 2023). Governance sets ownership, policy and risk appetite, document and accountability. Mapping identifies users, context, benefits, potential or anticipated harms, affected rights and dependences. Measurement assesses the technical validity, quality of explanations, privacy, security, accessibility, fairness, and human outcomes. Management emphasizes controls, identifies and rectifies incidents, notifies of limitations, and makes decisions on system scaling, redesign, stop, or disregard. Independent review is of particular importance if a system impacts education access, employability assessment, or opportunity ranking. Documentation, audit evidence, model change notifications, incident cooperation, data export, deletion and service continuity are important aspects that should be covered in a procurement contract.

Minimum evidence before scaling

Prior to the release to broad scale, institutions should show: factual accuracy in light of current local sources that an institution has verified as being acceptable, usability and understanding across the different groups in its developmental target set, particularly regarding non-deterioration or documented mitigation plan, reliable escalation and appeal, a secure trial and understanding of privacy testing, positive impact on decisions and developmental outcomes, and no deterioration in human support access. There should be stop rules included for serious harm, systematic misinformation or unmanageable inequity with a predefined progression criteria for a small pilot.

Table 3. Minimum responsible-AI controls for career guidance

Risk domain	Minimum control	Illustrative evaluation question
Validity and reliability	Local validation, source provenance, change control, uncertainty thresholds	Are recommendations stable, current, and correct enough for the intended decision?
Fairness	Subgroup and intersectional audits; option-diversity review; community input	Do similarly situated users receive comparable quality and breadth of opportunity?
Privacy	Data minimisation, lawful basis, retention and deletion, privacy impact assessment	Is every data element necessary, and can users inspect and control its use?
Transparency	Plain-language explanation, source dates, model limitations, uncertainty	Can users understand why an option appeared and what could make it wrong?
Human oversight	Defined escalation, trained reviewers, authority to override, appeal route	Can a meaningful human review occur before harm or exclusion?
Security	Threat modelling, access control, encryption, incident response	Could an attacker expose, alter, or manipulate career profiles or recommendations?
Accessibility	Inclusive design, multilingual and low-bandwidth access, non-digital alternative	Can disabled and digitally excluded users obtain an equivalent service?
Accountability	Named owner, audit log, impact monitoring, redress and retirement plan	Who is responsible when the system causes or contributes to harm?

Implications for Stakeholders

Higher education and training providers

AI should be a supplement, not a substitute, for human resources and resources for a career-development service. Not everything that has to do with the educational purpose should reduce to a need for a vacancy. Aggregate skills intelligence can be used to inform curriculum review, but it shouldn't be an educational purpose of a need for a vacancy. Programmes should be designed to include technical, transversal and career management skills and learning routes must be made explicit to students at the early stages of study. The governance models required for universities also need to link between career services, academic units, the data protection team, the IT security team, the equality & accessibility team, students and external employers. It is not possible to give the entire responsibility to a vendor or the implementation of one innovation project.

Career practitioners

AI is a tool for interrogation and not repetition. Provenance checking, explaining uncertainty, stereotyped or structurally unrealistic advices and documenting the professional overrides are some of the core competencies. While the repetitive search and administration tasks on the time sheet can be automated by AI, complex, relational, under-served cases must be prepared for additional investment of time in the form of service models.

Policymakers and workforce agencies

Public agencies can enhance the data infrastructure underlying trustworthy guidance, including interoperable Skill Taxonomies, Open OIS, and regional forecasts, quality assured course data and

privacy-preserving outcome statistics. There should be consequential education and employment uses and impact assessment and meaningful appeal on their use in the regulation and procurement specifications. However, precarious jobs, inadequate anti-discrimination enforcement, lack of social protection capabilities, inability to mobilize between places of working, and job-deprivation are all part of the scope of the policy challenge on the demand side, which cannot be solved by guidance.

Developers and vendors

It should separate retrieval of facts from generative dialogue, have dated and localized source repositories, and be able to reveal confidence and provenance. Option diversity, comprehension, corrections, downstream consequences and subgroup effects are included in product evaluation. Any claims made in marketing should articulate what the demonstrated benefits are rather than what capabilities exist. Vendors should issue model cards, data documentation, accessibility conformance, security evidence and the clear protocols of their updates.

Research Agenda

Future studies should take the prototype phase and cross-sectional surveys of acceptance one step further. Pragmatic trials for standalone AI recommendations and counsellor services versus those delivered alongside ongoing involvement from a counsellor are included in priority designs, as are studies of wellbeing, employment outcomes and additional skill development in the process of education before and after being deployed, particularly focused on changes in professional practice. Outcomes should not just be the accuracy of recommendations, but also decision self-efficacy, adaptability, breadth of options, informed consent, quality of transition, quality of job, and equity. Research conducted on an inter-national basis should check which parts go and which do not go across the border. Research is particularly required for disabled people, displaced people, workers in the informal labour markets, rural areas and mid career and older workers in low and middle income countries. Participatory research requires study of user interpretations of explanations, challenging of advice and the acceptance or the rejection of debiased advice. Technical research should include uncertainty-aware/recommendation methods when the data is under uncertainty, fair recommendation methods, data linkage under privacy concern, and tools for detecting sufficiency of evidence and escalation. Research into procurement, professional identity and accountability, cost effectiveness, and even a shift in workload because of AI need to be explored.

Limitations

This article is not a meta-analysis as it is an integrative review of the literature, though one is imperative, that is. There is no consensus on the effect size in the heterogeneous literature, and the fast changing nature of development leads to differences in technical capacities and obligatory requirements. Evidence is skewed by successful prototypes and higher education and formal labour markets. In this sense, the framework is explanatory and normative, causal pathways need to be empirically tested. This review does not cover all publications but only those in the English language and may not cover locally-publicized evidence and career-development traditions in a way that gives them in-depth coverage..

Conclusion

Career guidance is a possible labour-market intermediary, but without a clear definition of how it is used. It does the best job of structuring complex information, recognizing transferable skills, creating a variety of scenarios, and linking users with actionable learning and support. It cannot create jobs, eliminate discriminate, predict the numbers of jobs or substitute for the relational and

ethical work of career professionals. The suggested design places AI in a controlled landscape and measures the effectiveness by means of enhanced agency, developmental results, institutional action, and fair access. Responsible systems should be plausible, verifiable, privacy oriented, just, available, and efficient in monitoring their operation throughout their entire life-cycle. So career guidance may not be fully automated in future or technology free. It is a conscious alliance in which computation scales adds to and enhances, not takes away from human judgement, adaptability, and opportunity.

References

- Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2), 3–30. <https://doi.org/10.1257/jep.33.2.3>
- Adalet McGowan, M., & Andrews, D. (2015). *Labour market mismatch and labour productivity: Evidence from PIAAC data* (OECD Economics Department Working Papers No. 1209). OECD Publishing. <https://doi.org/10.1787/5js1pzx1r2kb-en>
- Akyazi, T., del Val, P., Goti, A., & Oyarbide, A. (2022). Identifying future skill requirements of job profiles for a sustainable European manufacturing industry 4.0. *Recycling*, 7(3), Article 32. <https://doi.org/10.3390/recycling7030032>
- Albaroudi, E., Mansouri, T., & Alameer, A. (2024). A comprehensive review of AI techniques for addressing algorithmic bias in job hiring. *AI*, 5(1), 383–404. <https://doi.org/10.3390/ai5010019>
- Assefa, Y., Gebremeskel, M. M., Moges, B. T., Tilwani, S. A., & Azmera, Y. A. (2025). Rethinking the digital divide and associated educational in(equity) in higher education in the context of developing countries: The social justice perspective. *International Journal of Information and Learning Technology*, 42(1), 15–32. <https://doi.org/10.1108/IJILT-03-2024-0058>
- Bankins, S., Jooss, S., Restubog, S. L. D., Marrone, M., Ocampo, A. C. G., & Shoss, M. K. (2024). Navigating career stages in the age of artificial intelligence: A systematic interdisciplinary review and agenda for future research. *Journal of Vocational Behavior*, 153, Article 104011. <https://doi.org/10.1016/j.jvb.2024.104011>
- Bobba, M., Flabbi, L., Levy, S., & Tejada, M. M. (2021). Labor market search, informality, and on-the-job human capital accumulation. *Journal of Econometrics*, 223(2), 433–453. <https://doi.org/10.1016/j.jeconom.2019.05.026>
- Buchanan, R. (2025). Exploring the role of AI in career access through a social justice lens. *Nordic Journal of Transitions, Careers and Guidance*, 6(1), 1–13. <https://doi.org/10.16993/njtgcg.77>
- Cedefop. (2023). *Skills in transition: The way to 2035*. Publications Office of the European Union. <https://doi.org/10.2801/438491>
- Cedefop. (2024). *Next generation skills intelligence for more learning and better matching: Skills anticipation trends, opportunities and challenges in EU Member States*. Publications Office of the European Union. <https://doi.org/10.2801/180485>
- Chui, H., Li, H., & Ngo, H.-Y. (2022). Linking protean career orientation with career optimism: Career adaptability and career decision self-efficacy as mediators. *Journal of Career Development*, 49(1), 161–173. <https://doi.org/10.1177/0894845320912526>
- du Plooy, E., Casteleijn, D., & Franzsen, D. (2024). Personalized adaptive learning in higher education: A scoping review of key characteristics and impact on academic performance and engagement. *Heliyon*, 10(21), Article e39630. <https://doi.org/10.1016/j.heliyon.2024.e39630>

- Fabris, A., Baranowska, N., Dennis, M. J., Graus, D., Hacker, P., Saldivar, J., Zuiderveen Borgesius, F., & Biega, A. J. (2025). Fairness and bias in algorithmic hiring: A multidisciplinary survey. *ACM Transactions on Intelligent Systems and Technology*, 16(1), Article 16, 1–54. <https://doi.org/10.1145/3696457>
- Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., Moro, E., Wang, D., Youn, H., & Rahwan, I. (2019). Toward understanding the impact of artificial intelligence on labor. *Proceedings of the National Academy of Sciences*, 116(14), 6531–6539. <https://doi.org/10.1073/pnas.1900949116>
- Gmyrek, P., Berg, J., Kamiński, K., Konopczyński, F., Ładna, A., Nafradi, B., Rosłaniec, K., & Troszyński, M. (2025). *Generative AI and jobs: A refined global index of occupational exposure* (ILO Working Paper No. 140). International Labour Organization. <https://doi.org/10.54394/HETP0387>
- International Labour Organization, & Organisation for Economic Co-operation and Development. (2023). *Global skills gaps measurement and monitoring: Towards a collaborative framework*. International Labour Organization. <https://www.ilo.org/publications/global-skills-gaps-measurement-and-monitoring-towards-collaborative-0>
- Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human–AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577–586. <https://doi.org/10.1016/j.bushor.2018.03.007>
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior*, 45(1), 79–122. <https://doi.org/10.1006/jvbe.1994.1027>
- Mavuso, N., Jere, N. R., & Van Greunen, D. (2023). A systematic review of customised career recommender systems for South Africa. In *2023 2nd Zimbabwe Conference of Information and Communication Technologies (ZCICT)* (pp. 1–7). IEEE. <https://doi.org/10.1109/ZCICT59466.2023.10528488>
- McGuinness, S., Pouliakas, K., & Redmond, P. (2018). Skills mismatch: Concepts, measurement and policy approaches. *Journal of Economic Surveys*, 32(4), 985–1015. <https://doi.org/10.1111/joes.12254>
- Organisation for Economic Co-operation and Development. (2023a). *OECD employment outlook 2023: Artificial intelligence and the labour market*. OECD Publishing. <https://doi.org/10.1787/08785bba-en>
- Organisation for Economic Co-operation and Development. (2023b). *OECD skills outlook 2023: Skills for a resilient green and digital transition*. OECD Publishing. <https://doi.org/10.1787/27452f29-en>
- Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No. 300/2008, (EU) No. 167/2013, (EU) No. 168/2013, (EU) 2018/858, (EU) 2018/1139, and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797, and (EU) 2020/1828 (Artificial Intelligence Act), OJ L, 2024/1689 (2024). <http://data.europa.eu/eli/reg/2024/1689/oj>

- Rigotti, C., & Fosch-Villaronga, E. (2024). Fairness, AI & recruitment. *Computer Law & Security Review*, 53, Article 105966. <https://doi.org/10.1016/j.clsr.2024.105966>
- Savickas, M. L., & Porfeli, E. J. (2012). Career Adapt-Abilities Scale: Construction, reliability, and measurement equivalence across 13 countries. *Journal of Vocational Behavior*, 80(3), 661–673. <https://doi.org/10.1016/j.jvb.2012.01.011>
- Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, 36(6), 495–504. <https://doi.org/10.1080/10447318.2020.1741118>
- Tabassi, E. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.100-1>
- United Nations Educational, Scientific and Cultural Organization. (2021). *Recommendation on the ethics of artificial intelligence*. <https://unesdoc.unesco.org/ark:/48223/pf0000380455>
- Viberg, O., Kizilcec, R. F., Wise, A. F., Jivet, I., & Nixon, N. (2024). Advancing equity and inclusion in educational practices with AI-powered educational decision support systems [Editorial]. *British Journal of Educational Technology*, 55(5), 1974–1981. <https://doi.org/10.1111/bjet.13507>
- Wang, C., Wang, K., Bian, A., Islam, R., Keya, K. N., Foulds, J. R., & Pan, S. (2021). *User acceptance of gender stereotypes in automated career recommendations* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2106.07112>
- Westman, S., Kauttonen, J., Klemetti, A., Korhonen, N., Manninen, M., Mononen, A., Niittymäki, S., & Paananen, H. (2021). Artificial intelligence for career guidance—Current requirements and prospects for the future. *IAFOR Journal of Education*, 9(4), 43–62. <https://doi.org/10.22492/ije.9.4.03>