



TOWARD RESPONSIBLE AI-ENABLED CAREER COUNSELLING: ETHICAL FRAMEWORKS, EDUCATIONAL POLICY CHALLENGES, AND HUMAN-CENTRED PRACTICES

Dr. Nosheen Malik¹

DOI: <https://doi.org/10.63544/ijss.v3i4.320>

Affiliations:

¹ Assistant Professor,
Department of Education,
The Islamia University of
Bahawalpur,
Punjab, Pakistan

Email: nosheen.malik@iub.edu.pk

ORCID ID: <https://orcid.org/0009-0007-5278-415X>

Corresponding Author's Email:

¹ nosheen.malik@iub.edu.pk

Copyright:

Author/s

License:



Article History:

Received: 22.11.2024

Accepted: 18.12.2024

Published: 31.12.2024

Abstract

Career counselling is one of the most significant events that can happen to students' educational or careers with an impact that affects students' future employability, socioeconomic mobility and personal identity development. Using AI technologies like chatbots, recommendation algorithms, and predictive analytics in career guidance systems provides unprecedented opportunities for personalized, accessible, and data-driven career counselling. This tech integration also poses ethical risks such as algorithmic bias, decision-making transparency issues, privacy concerns, and reliance on automated recommendations that could reduce human decision-making. This systematic conceptual review integrates research across three areas: AI ethics, educational technology, and career development theory and policy to explore the potential of AI in career counselling with an eye to responsible use and practice, accountability and fairness, transparency, privacy protection, and a human-centred approach to decision making. In this paper, the main ethical challenges outlined in 21 peer-reviewed studies and policy documents, an original Responsible AI Career Counselling Framework, which combines four core dimensions of ethical infrastructure, transparent decision system, human counsellor integration and inclusive outcomes, and policy implications for educational institutions and governments, are identified. The report's key findings indicate that responsible deployment of AI tools into career counselling must address data governance and algorithmic fairness; involve human counsellors in key functions across all AI-generated recommendations; embed transparency and explainability in the use of the tools; and adopt culturally responsive practices that address inequities in the digital education system. The paper contributes to the progress of theory seeking to bring AI principles to life in career guidance settings and pinpoints key areas of research needed to better understand the downstream consequences of AI on student outcomes and career progression, then offers practical implications that affect institutional governance, counsellor training and development, ethical procurement guidelines, and student safeguarding policies. Finally, a more human-focused and equitable approach to implementing AI into career counselling requires a shift from tech-based delivery to well-defined ethical and policy frameworks and procedures that prioritize human impact.

Keywords: Artificial Intelligence Ethics, Responsible Innovation, Career Counselling, Educational Policy, Algorithmic Fairness, Human-AI Collaboration, Digital Equity, Higher Education



1. Introduction

1.1 *The Transformation of Career Counselling Through Artificial Intelligence*

Career counselling has also completely changed during the last two decades from what it was in the counsellors' offices and interviews to digital and AI-aided advice systems that can be accessed from platforms and in various locations. A traditional humanistic approach to career counselling based on the principles of empathy, listening, active exploration of career identity, and collaboration has found its way into the realm of complex technology, to support the provision of individual career counselling to a much greater number of students. This shift is fuelled by several overlapping trends, most obviously a rise in demand for career counselling amid an evolving labour market, the number of education choices and career alternatives, the budgets of educational institutions and the scarcity of counsellors, and the lure of AI's promise to provide scalable, less error-prone, data-driven career advice by using granular industry knowledge and algorithms that override human ability to make sense of that information.

Today's AI use in career guidance involves various types of implementations. Technically intelligent career recommendation systems use an extensive knowledge base of employment information, jobs descriptions and student information to develop significant, customized recommendations that are based on personal skills, interests and the labour market. Career ChatBots with Conversational AI offer instant replies to career inquiries, cutting down on reliance on precious counsellor time and guaranteeing support around the clock regardless of time changes or distance. Predictive analytics systems can predict the "at risk" population of students as well as provide recommendations for interventions to be taken based on educational performance, demographic information, and past low, medium, or high rates of employment. The automated skill assessment tools assess student skills by problem-based testing and recommend and predict future career tracks. Labour market intelligence platforms are a collection of real-time data about jobs, salaries, and specific skills required in different market segments to serve as up-to-date labour market information for career decisions. In many ways, these techniques change the entire dynamics of career counselling from a core relational and dialogic process to a data-based, algorithmic approach driven by computational mechanisms that increasingly substitute for, mediate, and alter the counsellor-student-career-industry relationships.

This technological change is more than just about the efficiency of operations. The nature of the career counsellor's role shifts from 'principal decision-maker and confidante', to one that may be more enacted as an ability to independently evaluate algorithmic outputs, to oversee ethical issues and augment student emotional and identity-based career development factors not currently covered by Algorithms. Learner experiences are completely different when career information comes from a dialogue with an assessment model and data on behaviours, jobs and outcomes rather than from talking with a counsellor who has a complete prior understanding of the student's context, values and aspirations. The information ecology also evolves: a time when counselling professionals could depend on their own professional judgement, knowledge of the labour market and application of career development theory, it now faces an environment of more complex technological systems making recommendations with less underlying logic than ever, even for the trained health care worker.

1.2 *The Need for Ethical AI in Career Counselling*

Career counselling is one of the most morally complex fields in which AI can be used in the education sector. In career counselling, AI makes direct impact on the base-level life decisions of students, such as decisions regarding educational specialization, career field entry, geographic mobility, and decisions relating to a student's professional identity formation. So these decisions ripple through the life of the individual from their employment decisions and income path to the health and safety implications of their work, to stress and demands on life and work, to socio-economic status and social mobility. Career counselling might be the most significant institutional process that influences important life transitions from education to employment for many students beyond the "mainline" of higher education attention, including for marginalised groups, students from rural locations, those with lower income statuses, and those who are under-represented in higher education.

The ethical issues are more acute when it comes to the futures that are at stake. Career counselling is designed to help students at key developmental periods when they can be sensitised to outside influence, lack



of information and experience in the labour market, and limited experience in developing their career identity and self-concept as future employees. Students from racial and ethnic minority groups, students from economically disadvantaged backgrounds, and first-generation college students are documented as having critical experiences with high-quality career guidance that have a significant equity impact on students' completion of secondary or postsecondary education, employment, and economic outcomes. These very same populations, however, tend to be both the most susceptible targets of the biases of AI systems and the least able to confront or challenge the algorithmic recommendations. A system intelligence reflecting inequalities from the past and using historical public employment trends to advise students to avoid higher-paying jobs will inadvertently increase inequalities, for example, by directing students of colour to lower-paying careers; a system intelligence that uses history to advise avoiding technical careers is there for women; a system intelligence that uses history to advise rural students about job and career opportunities in distant locations. This being system intelligence is used to guide the future course of employment, which, when trained on and reflecting past inequalities, works to exacerbate inequality in the aggregate rather than reducing it; AI systems that deliberately steer clear of higher-paying careers for students of colour, advise women to steer clear of technical careers, and limit information about work and career opportunities in distant locations for rural students.

Moreover, career counselling is intertwined with the sense of agency, autonomy, and dignified recognition as human beings of the students. The Transparent and Explainable AI principal advocates that users should always be informed when communicating with an AI system, whether their personal data is used and why, and how decisions made by the AI system make sense during the communication. (Arrieta et al., 2019) This requires an indication of who is communicating with the user, information about whether one's personal data is being utilized and for what purpose, and an understanding of decisions the AI system makes during the communication, using explainable techniques that switch transparency to the users' profile. Career counselling touches on the trajectory of identity formation and one's future self-concept, so it requires especially strong safeguards, including to make sure that students are not left unaware that they are receiving guidance from algorithms, that they are aware of what data algorithms have obtained about them, and that they are able to understand and challenge algorithmic recommendations before they follow them.

1.3 Emerging Ethical Challenges

AI tools have quickly made their way into job paths without proper institutional management and ethical frameworks, creating a few notable and possible problems. Algorithmic bias is likely the most talked about issue. The ethical issues of the so called black-box models of AI use is their random appropriation of sensitive factors such as citizens race, age or gender; which can lead to an unfair decision. (Arrieta et al., 2019) Algorithmic bias can occur through several routes in career counselling scenarios. Occupational segregation, wage disparities, and unequal access to careers pathways in the training system can lead to algorithmic recommendations for the future. (Arrieta et al., 2019) The sources for bias can be traced to skewed data reflecting bias within data acquisition processes, tainted data involving errors in data modelling definition and wrong feature labelling, limited features using too few variables leading to false feature relationships, sample size disparities when using sensitive features inducing bias across subgroups, and proxy features that are correlated with sensitive variables and can induce bias even when sensitive features are not directly present in the dataset.

One of the second type of ethical concerns is the lack of transparency and explainability. (Chowdhury et al., 2023) The inherent lack of transparency in the process of decision making by AI makes organizational agents hesitant about using AI for important decisions. Recommendations by "black-box" algorithms place all of a sudden uncertainty on career counsellors and students as to why a certain career trajectory was suggested, what evidence the system valued the most, and the validity of the suggestion or whether it's an artifact of the algorithm. When a certain group of people in society is consistently discriminated against in the workforce or the organization, HR managers and educators may be called upon to respond to user and societal questions on the issue, but when the user or the HR manager does not have an understanding of the affordances or variants of the algorithms, it can cause them to fail to use them effectively and influence by prejudice. The lack of transparency has implications for career counsellors that serve students: They are unable to move forward



with research-based decision-making practices that evaluate employers' recommendations before they can recommend to students.

Thirdly, privacy issues and data ownership are challenges. Personalized guidance is taking centre stage in career counselling systems as granular data are gathered on students—sitting tests, learning styles, participation in social networks, family structure, economic status, health status, psychological traits, and more, ostensibly for the purpose of providing them.” (Arrieta et al., 2019) Privacy in the context of AI systems is particularly important due to the fact that ML models need huge amounts of data and most importantly, because XAI tools and data fusion techniques pose new challenges to preserve the privacy of protected records. The information compiled to serve the career counselling function presents dangers of surveillance, possible secondary uses of the information, potential violation of student confidentiality, and unequal power dynamics involving institutional systems analysing private information not fully understood or controlled by students.

The risks of automated decision-making arise when AI systems become decision makers and make decisions regarding access to opportunities, allocation of programs and resources, or a student's eligibility for programs. Human counsellors may still make contextual judgments, consider individual cases, have empathy, make exceptions, etc., and have ethical discretion but an all-algorithmic system may not. If the same algorithm is deployed across thousands of students, the potential for resulting from any embedded bias can be amplified, resulting in ubiquitous discrimination on a large scale.

These issues of digital exclusion and inequality are also important; especially pertinent in developing educational systems. Despite the availability of AI career guidance systems, students lack access where they do not have consistent internet connectivity, digital devices, or understanding of technology. Privileged students have easier access to AI-infused career guidance but those in rural areas, with lower incomes, and in lower income countries lack such access and may have a greater career opportunity gap than privileged students do.

The overreliance on AI recommendations can be detrimental to users' psychological development and mental well-being. Having algorithmic recommendations for career choices without allowing students to reflect and engage in career exploration, identity development and agentive decision making can close off essential developmental pathways. In traditional career counselling, the three key factors motivating people to make good decisions about their careers (adaptability, self-awareness, and agency) are encouraged as necessary elements of a meaningful career that are crucial in the current and future context of uncertain labour markets. Excessive dependence on AI recommendations could jeopardize these educational goals.

1.4 The Research Problem and Gaps

While there are numerous uses of AI in educational guidance for careers, there is limited and scattered research that examines the responsible, ethical and equitable implementation of AI in educational career guidance. Although there is much to be written about so-called ethics principles of artificial intelligence, as well as career counselling theory or implementation of educational technology, there are some weaknesses that still exist at the intersection between these two bodies of knowledge:

There are a few limited ethics frameworks which also specifically target AI career counselling. General principles of AI ethics (fairness, transparency, accountability, privacy, human agency, sustainability) are rooted in broader contexts and shaped by specific power dynamics, developmental aspects, and equity concerns of career counselling. The principles underlying career counselling have not yet been expressed in a way that is sensitive to the goals and priorities of other countries that can be found in the literature, thus it will be necessary to translate these principles into concrete practices, policies and technological specifications for career counselling systems.

Second, there is less than optimal linkage between AI ethics scholarship and educational policy analysis. The ethics literature on AI remains mostly technical/organizational, centred around design, technical regulation and corporate application of algorithms. AI ethics has not been a major focus in the field of educational policy literature. This means that AI ethics standards are not translated into institutional policies, professional standards, regulatory frameworks, or funding arrangements that provide clear guidelines for or support policymakers.



Thirdly, in general, there is little empirical evidence available on the long-term outcomes of AI-driven career counselling interventions on student outcomes, career trajectories, the quality of employment, and equity. Compared to the research on live guidance, most research on AI in education centres on learning analytics or Adaptive Learning Systems. There are no published studies that examine whether recommendations by AI can lead to better academic performance, whether there are empirically documented harms from bias in recommendations, whether students feel as if they had agency and autonomy in the process, or whether equity is enhanced or diminished in the process.

Fourth, there is limited research exploring practice in developing country contexts. Most of the AI ethics and educational technology scholarship comes from and is shaped by the views of the countries that are most developed, and have access to more institutional resources, technological infrastructure and regulatory capacity. But there are certain challenges and constraints faced by developing nations that are quite distinct from first-world countries where the resources available for education are limited, career counselling infrastructure is minimal, and AI may provide a true opportunity to scale up counselling.

Fifth, there is a lack of information for career counsellors on how to handle the use of AI tools. Career counsellors, in turn, need to be trained to be literate in AI, proficient in critically evaluating algorithmic suggestions, to have knowledge of ethical issues with AI in counselling and to come up with ways to preserve in their counselling practice the 'relational' and 'human-centred' aspects while incorporating technological tools into conversations. However, these competencies have not significantly been integrated into professional development programs or into the training for counsellors.

1.5 Research Objectives and Questions

These gaps are addressed in this paper, which is a systematic conceptual review in an integrated fashion, drawing mostly from theory, empirical research, policy research and developing new conceptual frameworks. The aims of the study are to:

1. Examine the ethical issues that arise as a consequence of the adoption, use and integration of AI tools in career counselling systems, with an analysis of four categories of algorithmic bias, lack of explainability, privacy and lack of autonomy issues in specific career counselling settings.
2. AIIE Principles and Frameworks for Responsible Implementation of AI in Education: This objective aims to dissect the principles and frameworks governing the responsible implementation of AIs in general and career counselling specifically, as well as analysing the practicality and constraints involved in the institutional implementation, from an abstract to a practical level.
3. Investigate policy ramifications at institutional and government levels for equity, transparency and accountability in how AI is used to counsel students in career exploration, such as governance structures, regulation, professional standards, and ways to protect students.
4. To create an innovative and original human-centred approach to career counselling which enables the use of AI responsibly while addressing issues of transparency, explainability, based on good ethical AI infrastructure, substantive involvement of the human counsellor, and inclusive results – covering diverse educational contexts in developed and developing systems.
5. To outline research gaps, methodological challenges, and future research pathways underpinning future evidence-based practices in empowering career counselling with AI.

1.6 Research Questions

The review is structured on five key research questions:

The key ethical issues that arise with the use of AI tools in career counselling, and how they show up in the career development systems through the design, application and deployment of AI are:

1. What are the specific ethical issues that arise when making use of AI instruments for career counselling, and how do they look when they are designed, applied and deployed within the career development systems?
2. What principles, frameworks and practices distinguish responsible, ethical and equitable implementation of AI in career counselling, and how can general ethical principles be made tangible in the terms of institutional practice and technological specifications?



3. What policy requirements, governance structures, and regulatory mechanisms are necessary to ensure that AI-enabled career counselling systems advance rather than undermine educational equity, student autonomy, and career opportunity?
4. What is the place for human career counsellors in AI-based career counselling systems and how can institutional designs ensure that the role of humans is not marginalized but central in career counselling where there is relational support and human judgment and ethics?
5. What is the balance between innovation and opportunity of the use of AI tools in the educational landscape and what does it differ in relation to ethical responsibilities and in each educational context – developed and developing countries?

2. Literature Review and Theoretical Foundation

2.1 Responsible Artificial Intelligence Frameworks and Core Principles

Responsible AI has grown to a body of scholarship and practice for designing, deploying, and governing AI systems that serve human welfare, human rights, equity and limit harms. (Arrieta et al., 2019) Methodological principles guarantee that the requisites of the various segments, such as no discrimination, sustainability, privacy or accountability, are integrated into the process, as well as other universal ones of similar importance relevant for explainability: its underlying purpose. Many international bodies, corporations and research institutions have produced guidance and rules about the values and principles which should guide the development of AI.

One of the most detailed policy-specific guidelines is provided by the ECHAMBARB'S Ethics Guidelines for Trustworthy AI which focus on a variety of assessment criteria. (Arrieta et al., 2019) The European Commission has recently published ethical guidelines for Trustworthy AI through an assessment checklist that can be completed by different profiles related to AI systems, based on a series of principles: 1) human agency and oversight; 2) technical robustness and safety; 3) privacy and data governance; 4) transparency, diversity, non-discrimination and fairness; 5) societal and environmental well-being; 6) accountability. These principles are comprehensive and cover aspects of artificial intelligence, such as explainability, fairness, human agency and as wide-ranging issues affecting systems, such as sustainability, privacy, security and safety.

As a fundamental law of fairness, the AI systems should not perpetuate or worsen discrimination and inequality. The purpose of Fair AI is to devise AI systems which explicitly advance inclusion, offer equal opportunity to all without discrimination, and advance for the common good, thus the outputs of AI systems shouldn't result in any kind of discriminatory action with respect to race, religion, gender, sexual orientation, disability, ethnic source or any other personal state (Arrieta et al., 2019). There are many definitions of fairness and fairness criteria to consider when operationalizing fairness; not all mathematical fairness definitions can agree on one another and fairness when applied to algorithms is a value judgment rather than a mathematical judgment.

Second core principles are Transparency and Explainability. Individuals should always know if they are dealing with a person or AI system, if their personal details are being used by the system and for what purpose, and be able to have some understanding of the decisions it is making (for users, by using XAI techniques tailored to their profile). For career counselling, transparency must be provided so that students know they are getting information through an algorithm, what personal information the algorithm has collected and analysed, why certain career courses were being recommended, and the limitations and potentials for bias in the algorithm.

Privacy and Data Governance are crucial protections. Throughout its life cycle AIs should pay attention to privacy and security standards, thereby establishing the principle of Privacy and Security by Design, which should be embedded within the processes that are already in place within a company and should be inspired by larger dimensions of Responsible Research and Innovation which involve sustainability, public engagement, ethics, gender equality, open access, and governance (Arrieta et al. 2019). In career counselling settings, this includes seriously considering the amount of data collected, only gathering what is needed to provide the student with appropriate career counselling; secure data storage and processing; informing students of what data is being collected and the ways it will be used; limiting the amount of time data will be



kept on her or his record; and providing safeguards that ensure the sensitive information of the students is protected.

Accountability and Auditability demand organizations to be responsible for the outcomes of the AI system and that systems can be open to scrutiny outside the organization. If AI systems impact fundamental rights – including safety-critical applications (Arrieta et al., 2019) – the systems shall always be audited by an external third party, and negative impacts shall be minimized as well as reported, including reporting actions and decisions resulting in negative impacts on fundamental rights and assessment of the negative impacts and what may be done to address them. Accountability in career counselling requires that institutions be held responsible for the effects of the algorithmic recommendations on students, that they conduct regular audits including assessments of whether systems produce biased outcomes, make such results transparent, and create processes for students to report harms and/or challenge algorithmic recommendations.

Regardless of how powerful the AI systems become, Human Agency and Oversight are still very important. The potential that AI has should be augmented by human empathy and understanding in the context of the global, organizational and personal sphere, which makes humans the gatekeepers in using the findings of the AI in an appropriate manner, thus complementing the decision-making process instead of replacing it. (Chowdhury et al., 2023) For career counselling, a combination of three things is required: 1) human (in this case, the career counsellor) maintain control over high-stakes recommendations that will impact students' futures, 2) algorithmic recommendations add value to, but are not substitutes for, human decisions and judgement, and 3) that students remain in control and autonomous regarding their own decisions as to what career courses and degrees they pursue, as well as what careers or jobs they take.

2.2 Ethics of AI in Education: Lessons for Career Counselling

Ethics of AI applications in wider educational contexts make for important lessons and precedents that can be applied to the field of career counselling. AI Education Ethics has evolved to Community-Wide Frameworks, drawing in education, student, policy and other perspectives. Like AI career counselling, these AI applications in education have some core features: they gather student information about their learning behaviours and identify opportunities for instruction and assessment; they inform the teacher or instruct student about educational opportunities; they relate to educational inequities; and they are present when students are in developmental contexts.

There are clear trends to be seen in the research on the themes of educational AI ethics. Initially, technical excellence and ethical responsibility are not one in the same. The history of AI in education between 2018 and 2024 illustrates the breadth of applications that technological advances are opening up as well as the ways in which they are not yet being applied ethically. AI systems often appear to be very sophisticated in algorithmic abilities but are ethically questionable when deployed because the Subject Matter experts fail to consider the issues of bias, transparency, privacy and equity when designing and developing them.

Secondly, ethical implementation of AI requires the involvement of stakeholders and a wide range of views. To ensure the viability of Educational Decision Support Systems (EDSS), especially those with Artificial Intelligence, to support equity, they must be designed and assessed as socio-technical and cultural systems that can have both direct and indirect influences on every student, particularly those within historically marginalized groups. APs developed favourably by an educational technologist, but lacking meaningful input from the educator, the student, the parent, and the community, tend to reflect the assumption and bias of their developers and fail to consider the context that affects educational equity and impact.

Third, often side effects and systemic impacts are far bigger than merely optimizing a technique for its specific purpose. Systems designed for the sole purpose of predictive accuracy or efficiency can cause systematic harms on a wide scale if they are adopted on a large scale and education. For instance, one of the many benefits of optimising learning analytics systems to detect "at-risk" students from behavioural signals is that automatic identification of students may sometimes result in lower expectations of and reduced support for those who are already marginalised students, rather than those academically at-risk.

2.3 Career Counselling Theoretical Foundations

Theoretical underpinnings of career counselling that focus on human agency, relationality, identity development, and context approximately permeate this work and are central to understanding the process of



making career decisions. One of the frameworks which are influential among these frameworks is Career Construction Theory (CCT). It argues that career development is not only a skill-job matching or an optimization of the objective variables but is essentially a process that is narrativized and shaped through the processes of identity. People continuously "make meaning" and develop career narratives that shape what they value, what they like or dislike, what capacities they possess, and what contextual opportunities are available; they begin to develop career stories that weave their values, interests, capacities, and contextual opportunities together into coherent calls for future possible selves and work lives.

Career Construction Theory is a career development process that focuses on life and career construction in which people are active participants. This includes a personal interpretation, taking a reflective look at life events, and being able to make sense of them, explore one's own identity and the prospect for the future. Narrative identity is a meta capacity which allows a person to direct his or her activity and plan his or her career through a dialogue, continuously developing internal and social discussion of which a person is a part. Contextualizing this narrative process reinforces three very important elements of career development that are not recruited when assessing by the most objective means: subjectivity, intentionality, and reflexivity (Rossier et al., 2020). In this theoretical perspective, the aspect of meaning-making is pivotal. It emphasizes how people bring together various parts of themselves and how they deal with the uncertainties in life by shaping narratives that make sense and give a purpose. This process is, of all things, complex: emotional, cognitive and existential features – all humans.

Career counselling supportive of CCT smoothes this process of narrative construction by helping clients to explain, analyse and edit their career story, drawing attention to the importance of meaning, identity, flexibility and experiential learning. Career Narrative generated guided imagery (CGI) has been proved to improve the process. An idiographic script problem-solving technique has been developed that employs client-centred stories to elicit deeper exploration of identity and meaning-making in counselling sessions, encouraging clients to actively envision alternative future events and futures, and to pursue alternative career options (Stoltz et al., 2018). Thus, career counsellors are co-constructors and partners in the narrative. They are able to help clients navigate their interpersonal dynamics, recognize the minefield of constraints or opportunities in the context, and develop agency when making career decisions which requires nuanced interpersonal understanding, empathy, and flexible responses to individual situations.

Social Cognitive Career Theory focuses on the interplay between personal agency, self-efficacy beliefs and contextual factors on how careers develop and are made. Under this model, people's decisions on career take place when there is a dynamic interaction between their sense of ability in specific areas, their outlooks for various careers, their interests and values, and the environment that supports and can prevent them from pursuing their career. Importantly, the distinction is made that while career development has mainly been defined as a matching process (matching the person to the job that is ideal for them), it also is viewed as a process of acquired efficacy beliefs, managing environmental constraints, and keeping agency despite contextual barriers. From here there is a need to boost student agency and efficacy rather than allow algorithms to determine their career path for them. Counsellors might leave the counselling process with increased belief in their efficacy if they spend time exploring, reflecting and actively choosing their way to the algorithmic course recommendations, but they could leave with decreased efficacy beliefs if they simply accept the algorithm's suggestions (Lent & Brown, 1996; Buhăescu-Ciucă, 2023)

Humanistic Counselling Theory stresses that the counsellor-client relationship is therapeutic in nature and will be a key component to positive outcomes. In this tradition, the counsellor's delivery of unconditional positive regard, empathic understanding, and authentic presence allows clients the freedom to explore their self-concepts, retell stories that are muddled, and/or make choices that are congruent with their authentic values. The relational aspects of counselling (counsellor-student relationship, the experience of being affirmed and cared for, and the prospect of engaging in an authentic conversation) are not an addition-on to career counselling, but integral aspects of it and vital to its effectiveness. An AI system can pass you the information about the careers, but it can't give you that feeling of being seen and supported by another person. This human relational element of career counselling is often the most important component for today's students, especially the individuals who are marginalized, doubting or are experiencing identity confusion (Amari, 2022; Renger



& Macaskill, 2021; Spors et al., 2023; Velasquez & Montiel, 2018).

Together, these conceptualizations suggest that AI can be a useful and valuable tool in career guidance and counselling – sharing's information on the labour market, identifying pertinent career opportunities, complementing counsellors' ability – but it should not replace subjective judgement, relational support, meaning making, ethical guidance and oversight in activities of career counsellors. These dimensions owe their place in the career counselling practice to their essential human value rather than technology's ease of use and should be sustained and enhanced.

2.4 Human-Centred AI and Complementary Intelligence in Career Counselling

Recent scholarship has focused on human-centeredness in the use of AI in supporting human wellbeing highlights the need for a closer collaboration between humans and AI, suggesting that the best outcomes in human-serving contexts are achieved through complementarity, don't rely on the replacement of human abilities with AI. The application of Artificial Intelligence in different industries has led to unique enhancement where AI technological abilities excel in handling huge amount of data and automation of repetitive task automated process, and the human creativity, intuition and emotional intelligence complements the AI system for achieving enhanced decisions with hyper-informed data supplement and through optimized Human-AI collaboration. In a sense, the principle of complementarity is especially applicable to career counselling.

AI systems can offer unique skills to career counselling by analysing and condensing large amounts of labour market data, recognizing patterns among thousands of historical student-to-outcome pairs, facilitating rapid access to career information, combining information from a student's profile and making personalized career recommendations, and scaling to serve students who don't have counsellor time. Counsellors' competencies are, however, fundamentally values, contexts, aspirations of individual students, ethical judgment regarding appropriate recommendations, relational support and encouragement, flexibility and (contextual) sensitivity when modifying recommendations as new information is received, accountability to student as a whole person rather than as a piece of data, and professional skills in career counselling as developed through training, supervision, and ethical practice standards (Sharma et al., 2021 ; Kolhe & Bhat, 2024).

The best career counselling systems will be designed to channel AI's capabilities to improve counsellor effectiveness – not take the place of counsellor judgment. For instance, one algorithm might generate a collection of job listings that align with a student's desired fields, and raise flags regarding the fields (e.g., unstable job outlook, low salary). The algorithm could then provide those flags as part of the labour market analysis prior to the counsellor meeting with a student, freeing up some of the counsellor's time to do what the front-line caregiver should focus on: relational support, helping the student reflect on their own values and aspirations, and engaging in informed decisions. On the other hand, counsellors would have authority to contest and situate the algorithmic recommendations due to his or her knowledge of the student and professional judgment.

2.5 Algorithmic Bias and Equity Issues in Career Guidance

While algorithms in career advice systems are intended to be based on the objective and data-driven decision-making process, they can and do explicitly maintain and amplify current employment inequalities and discrimination. The mechanisms by which bias can get into algorithms used to guide into careers is crucial for designing mitigation practices.

Career Recommender Algorithms can be biased in a variety of ways. Training data based on past employment patterns are always subject to past discrimination. (Arrieta et al., 2019) Common causes of bias include skewed data that already contains bias in the data collection process, errors in captured data that would constitute tainted data (errors in data modelling, feature labelling, etc.), a small number of features (the use of too few features means that the data creates false relationships), small differential feature sizes (the use of small sample sizes causes differences in features introduced in the distribution (possibly induces bias when using sensitive features by the lack of a sufficient number of samples in various subgroups)), and proxy features (features that are correlated with sensitive features can cause bias even if sensitive features are not present in the dataset). When applied specifically to the field of career counselling, algorithms based on



historic career statistics that exhibit segregation (e.g., low proportions of women working in technology, overrepresented among men working in early childhood education or nursing, and low proportions of racial minorities in higher-paying professions) will continue these patterns of segregation in their outbound recommendations. If the algorithm is trained with data indicating a historical trend for women with specific academic characteristics to take up roles in less well-paying industries, it could recommend a similar path to women with the same academic profile today, thereby reproducing that historical discrimination.

Impacts on Vulnerable Populations are not distributed evenly. Women might get different referrals from men who have similar skills, and from women in more lucrative or leadership positions. Women may be directed into lower-paying or lower leadership status career paths than men with similar qualifications. Students of racial and ethnic minority backgrounds can also be exposed to algorithms that do not explicitly take race into account but end up making recommendations through other proxies for race for example (geographic region, type of school attended, socioeconomic status). Limited information to rural students of urban opportunities/career paths. Parents with disabilities might experience algorithms that either overlook their abilities or define disability as a job restriction instead of an ability that should be accommodated. Because first generation college students often don't know much about a range of careers, they may be encouraged into career areas they wouldn't have imagined or algorithms look outward based on self-reported interests and do not broaden perspective.

Bias Mitigation Mechanisms, that are to be implemented, are available. For bias detection, XAI proposal statements can be utilized utilizing frameworks that qualitatively sift through the data to visually summarize the bias of a model by individual and group fairness, making the process of understanding and quantifying bias easier, with possible bias reports, quantitative biasing justifications, a measure of how fair a model is, and explanations as to why a user or group would be treated unfairly in the provided data. These technical mitigation strategies are not sufficient, however, without associated practices: improving the diversity of people in algorithm development teams; ensuring voices from vulnerable groups to bias are also heard in the development and testing of algorithms; continuous monitoring of algorithm outputs for disparate impact; and a commitment to mitigating algorithms when bias is identified despite the potential loss of overall system efficiency.

2.6 Global Perspectives on Ethical AI Adoption: Developed and Developing Contexts

Any discussion of the implementation of responsible AI in the field of career counselling can't be divorced from the very different contexts, resources, capacities and challenges of education systems around the world. There is a possibility for implementation that is appropriate to a wealthy technologically developed country and is not possible in resource challenged environments like the developing world and vice versa, implementation priorities may differ significantly based on location of developed and developing country.

The unique situations in these contexts require multi-level forms of governance and more flexible governance frameworks. In recent years, studies have advanced the urgency for responsible AI governance frameworks to be holistic, extending beyond the issue of algorithmic fairness to consider the entire AI system lifecycle, taking into account local sociocultural contexts and capabilities (Lu et al., 2024). Furthermore, private-sector AI audits have identified the recurring challenges related to weak levels of stakeholders' engagement, financial constraints, and uncertainty over the interpretation of different levels of AI governance regulations, highlighting the necessity for capacity building in ethical governance of AI in every context (Schiff et al. 2024).

Policy proposals for developing countries highlight the importance of building policy instruments to encourage the inclusive use of AI technology and ethical governance, which includes investment in digital infrastructure, AI literacy and the development of policy, as well as the call for public-private partnerships (Folorunso et al., 2024). For relevant stakeholders in African Africa, strategic priorities encompass improvements to digital infrastructure, inclusion of AI education into school curricula, and the formation of collaborations with international players to gain access to AI skills and resources to ease implementation of AI (Laban et al., 2024). Such measures are crucial to establishing inclusive AI systems that support a wide range of populations, while avoiding incurring additional risks of deepened inequalities.

Additionally, economic dynamics are also quite different – AI implementation impacts employment



differently in developed and developing nations. In developed countries, the productivity growth and labour market changes are somewhat predictable, and policy interventions can be directed to support these changes (Masood, 2024). This is because in developing economies, infrastructure and the uptake of AI technologies is restricted, adding further difficulties to skills training and education or retraining and policy-related gaps.

To conclude, a well-informed approach to career counselling using AI requires a global perspective that is distinctive. Advanced infrastructure and regulatory frameworks are present in developed contexts, but they need to come to terms with intra-national equity and need to not impose unnecessary market-driven deployments. Contexts for development are facing critical issues at the moment such as capacities, regulations and infrastructure which may pose obstacles to equitable AI benefits. At the macro level, policies should integrate AI capacity building, flexible and accessible infrastructure development, and ethical governance practices specific to local contexts to aid in the prevention of education and digital disparities exacerbated or caused by AI. Failure to consider such factors could compromise AI innovations for their potential to transform how they guide international careers.

The synthesis is based on research findings that put multi-level governance, capacity building and ethical good practices in place within the different contexts of global AI use (Folorunso et al., 2024; Laban et al., 2024; Lu et al., 2024; Masood, 2024; Schiff et al., 2024).

3. Conceptual Framework Development: Responsible AI Career Counselling

The section introduces an original Responsible AI Career Counselling Framework (RACC Framework) synthesized from AI ethics principles and career counselling theory, educational equity scholarship, and research on implementation in organizations. Its framework comprises four interdependent elements: ethical infrastructure, decision systems that are transparent and explainable, meaningful involvement of human counsellors in the process, and inclusive outcomes, balancing the first three requires addressing them all at the same time when implementing responsible AI in career counselling.

3.1 Ethical AI Infrastructure

The first layer of the framework is the technical and organizational infrastructure which sets the groundwork for ethical AI systems. Data Governance is the key starting point. Applying the basic tenets of AI to an organization involves articulating the values and boundaries that need to be established, then applying the AI principles, designing the infrastructure and tools, communications and training on technical and non-technical matters, carrying out questionnaires to get people thinking about particular impacts of the AI system that can be expected in aspects where technical and non-technical issues are not commented upon, and using the tools to answer questions and mitigate the problems that would be found. Ethical data governance for career counselling needs:

- Data minimization: Ensure that there is only a minimal collection of personal information that is used to deliver accurate career guidance that is not excessive or tangential.
- Purpose limitation: Only relying on the collected information for the defined purpose of the career guidance (e.g., not predicting academic success for other institutional reasons, or giving it to external employers).
- Informed consent: Students will be asked for the specific, informed consent before gathering of career information, and information about what data will be collected, its purposes, its storage, access, and also its retention.
- Transparency of data: make it easy for students to access their own data and explanations on how their data has been analysed and used to provide recommendations/hint.
- Security and confidentiality: Providing technical controls to ensure that student data is not accessed, disclosed or used by others in an unauthorized or inappropriate manner.

The second is ethical infrastructure: Algorithmic Fairness and Bias Mitigation. This means there needs to be intentional efforts devoted to recognizing and distilling away possible sources of bias. To implement, there needs to be awareness and training on potential issues (technical and non-technical), a form or framework that makes the individuals think about impacts of the AI system, with guidance on how to acts on these if they are undesirable, tools to answer questions and mitigate problems identified (including the XAI tools and fairness tools), and, finally, a governance model that sets responsibilities and accountabilities. Practically, this



includes:

- Diverse and representative training data: Training the algorithms to draw on a varied and representative demography of students, employment contexts, and career paths (Content focus); avoiding overrepresentation of privileged populations, and underrepresentation of marginalized populations.
- Regular testing to uncover disparate impacts of the algorithm by different demographic groups (fairness testing), regular audit on the algorithm by a third party (fairness auditing).
- Removing bias – having clear protocols for identifying and remediating bias rather than taking it for granted.
- Involving affected voices: Ensuring students (especially those on the margins of discrimination) are part of the design and testing of algorithms.

3.2 Transparent and Explainable AI Decision Systems

The second dimension of the framework focuses on making decisions made using algorithms transparent and intelligible to students and counsellors. The second aspect of the framework considers making decisions using algorithms transparent and intelligible to students and counsellors. Transparency and Explainable AI requires humans to be aware at all times when speaking to an AI system, if their personal information is being used and why, as well as being able to get a sense of how the decision is being made by the AI system and, importantly, how it can be understood via XAI methods depending on their profile. This involves the following requirements for career counselling:

Clarifying if AI has been used: If students are being given career preferences, they should be told that those preferences are based on an algorithm, a counsellor's judgment or both, never confused about whether they are talking to an AI algorithm.

Intelligible explanations: When algorithms suggest certain careers students ought to pursue, students should get explanations they can understand (not technical details about the procedures of machine learning, but clear explanations of what features of the career that machine learning algorithm used; it is important for students to see what the algorithm assumed, and what the limits of the recommendation are).

Access to alternative perspectives: The algorithm's output and output from multiple counsellor interpretations and critiques of the algorithm; information the algorithm had not recommended but might have been appropriate for the student; information the algorithm had not had about the student that might have changed its output.

Limitations awareness: Students should be aware that recommendations from an algorithm are predictions, not guarantees, that they are based on historical information that does not necessarily reflect future opportunities, that they are based on other values and optimization criteria important to the algorithm, and finally, that while the algorithm is useful, students' own judgement must ultimately be given in decisions about their career.

3.3 Human Counsellor Integration

The third dimension can be understood as the real job role of human career counsellors for artificial intelligent career counselling system. The third dimension can be interpreted as the workplace role of human career counsellors in an artificial intelligence-based career counselling system. Human counsellor involvement in the process should not simply be inserted into places where AI "recommendations" are the "default" and counsellors now present and administer algorithmic recommendations but strongly be involved and have an authority in the process. AI effectiveness and implementation will depend on how efficiently AI can adapt with people's capacity to demonstrate empathy and comprehend findings with regard to global, organizational and individual context, where humans will act as gatekeepers maximizing and leveraging AI capabilities to utilize findings appropriately, thus augmenting decision making. This requires:

- In making recommendations and decisions that impact students' futures, counsellors have authority based on their expertise and judgment. Algorithms offer information and indications which direct the counsellor's judgment but cannot replace it.



- Counsellor as ethical gatekeeper – Counsellors critically assess the alignment of the algorithmic recommendations with ethical standards of practice, the consideration, sensitivity and relevance of the individual student's values and context as well as the benefits for the individual student.
- The capacity to perform a counsellor's job: To be able to make exceptions to algorithmic recommendations when context requires; to integrate information that is beyond the scope of the algorithm into consideration for the student; to address the emotional, contextual and identity issues of career decision-making which algorithms cannot.
- Technical training and skills: Career counsellors need training in AI literacy, understanding and being able to critique AI recommendations, the ability to work well with AI systems and uphold ethical boundaries, and continuous personal development to stay knowledgeable as technology advances.

3.4 Inclusive Career Guidance Outcomes

The 4th framework dimension defines the outcomes and impacts that responsible AI- career counselling should yield. In addition to technical measures (algorithmic accuracy, system efficiency), outcomes for all students of equity, empowerment, and development should be considered. (Viberg et al., 2024) Advancing equity and inclusion in educational practices with AI-EDSS requires adopting a socio-technical and cultural perspective crucial for designing, deploying, and evaluating AI-EDSS that truly advance educational equity and inclusion. This demands:

- Equitable access and outcomes: A career counselling powered by AI should be equally accessible to everyone, irrespective of geographical location, socioeconomic status, educational background, disability, or any other attribute. Outcomes should show that use of the system facilitates access for marginalised students to access and seek high quality career opportunities, rather than simply perpetuates inequities.
- More student autonomy/agency and decision-making capability: Learners should feel more knowledgeable about career counselling, better able to make their career decisions, and more empowered rather than reliant on decisions made for them by algorithms and systems and uncertain about their ability to assess career options.
- Reflective career exploration should be supported not just determined by career matches, and there should be emphasis on values, interests, abilities, aspirations, etc. when discussing career counselling. Pupils need to be able to acquire the skills necessary to be 'career ready' and have careers that involve them being 'agentic' or not only reacting but taking initiative across their entire life.
- Recognizing limitations transparently; acknowledging that career counselling – either human or based on artificial intelligence – offers a partial view of career options; acknowledging and retaining the uncertainty one feels in career decision making and acknowledging that a career path is subject to change over the course of career development.

3.5 Moderating Factors and Contextual Considerations

The framework acknowledges that implementation won't happen in a uniform manner. The textbooks – or reality – which are relevant for a specific context influence the nature of AI responsibility:

The feasibility of implementation is affected by AI Literacy of counsellor, students and the institutional leaders. Contexts with higher AI literacy can have more complex systems and governance in the context of AI, while contexts with lower AI literacy will need more simplified and transparent systems and more intensive training on the subject of AI.

Institutional Readiness includes current technological capabilities, knowledge and skills of the staff, funding capability, and institutional support for the practice of ethics. Readier institutions have the capability to put in place more thorough governance structures while resource-limited institutions may need to adopt more streamlined governance approaches.

Policy Environment encompasses national policies on data protection, education and labour market policies. Data protection and education governance are strong scaffolds for the effective implementation of AI; in contexts without data protection or the educational governance, more institutional autonomy and initiative are needed.



Cultural Context influences definitions of equity as well as values and aspirations of students, culturally appropriate/desirable career pathways, and culturally normal/respected relationships with institutions and with students. Implementing practices is a contextualized process, not employing universal definitions of responsible practice.

The access to and benefits from using AI for Career Counselling differ between and within countries. There is Digital Inequality in who can access and who can benefit from an AI-powered Career Counselling. Implementing must actively tackle digital inequalities compared to parents and children being assumed to have equal technological access.

4. Methodology

This paper draws on both theoretical and empirical research, the policy review and the development of a framework. This is not a formal systematic review, nor is it employing set of strict protocols and meta-analysis, but is instead a conceptual synthesis method tailored to the interdisciplinary research question that combines fields of AI ethics, career counselling, education policy and responsible innovations.

4.1 Research Design and Data Sources

The review included information from peer reviewed research papers from a variety of academic disciplines and practitioner groups, international reports and professional practice literature from them. The databases that were searched to find academic references are: Scopus, Web of Science, ERIC (Education Resources Information Centre), IEEE Xplore, and ProQuest Dissertations & Theses. Database search was complemented by two other methods: Google Scholar and hand-searching of key journals. Policy and practice sources encompassed records of international sources (such as OECD, UNESCO, European Commission), governmental sources, professional association policies (American School Counsellor Association, career development organizations), and academic policy centres.

4.2 Search Strategy and Inclusion Criteria

The search terms used were various combinations of the core concepts of the topic suggested by Artificial Intelligence and career counselling: Artificial intelligence AND career counselling;

ETHICALY ARTIFICIAL INTELLIGENCE IN CAREER COUNSELLING; AI POLICY AND GOVERNANCE IN EDUCATIONAL; ALGORITHMIC ETHICS IN EDUCATIONAL SYSTEMS; PRIVACY AND DATA PROTECTION IN STUDENT INFORMATION SYSTEMS; HUMAN-CENTREED DESIGN IN EDUCATIONAL TECHNOLOGY; DIGITAL EQUITY AND ACCESS IN HIGHER EDUCATION.

Studies that were not published in English, before 2018 (to guarantee the validity of recent publications) were excluded as were articles that did not explore any interaction between the core concepts and those not gray literature (institutional reports) that did not include original, empirical findings or important theoretical content. The studies extended from the theoretical, empirical, policy, and conceptual categories, to accommodate the broad spectrum of the scholarly landscape.

4.3 Data Analysis and Framework Synthesis

Given that the literature comes from various disciplines, and that certain units addressed the same subject, the review did not utilize meta-analysis, but it did utilize thematic synthesis as the process adopted. All the articles selected were systematically coded to identify key elements such as the study design, ethical principles identified, policy recommendations made, empirical evidence of bias and fairness found, stakeholder perspectives, and implementation recommendations. Coding was done in an iterative fashion, with descriptive coding categories being developed during several re-readings, with the specific patterns emerging. Thematic clusters were derived through a process of inter-reflection and discussion between researchers, focusing on themes and their direct connection with situations of career counselling and responsible AI implementation.

Next, a conceptual framework was elaborated by synthesizing results across themes, while adhering to the concepts of responsible innovation found in the literature and the principles of human centred design. The framework aligns with the OECD AI Principles, UNESCO Recommendations on AI Ethics, and existing career services governance frameworks, but it is tailored to the context of career services.



5. Findings and Discussion: Key Themes in Responsible AI Career Counselling

5.1 Ethical Opportunities and Innovations in AI-Supported Career Counselling

AI technologies have the potential to revolutionize career counselling in meaningful, inclusive, and positive ways. Machine learning can be used to create personalized career coaching systems capable of analysing a student's preferred strengths and personality, interests, current labour market information, and education streams to make personalized recommendations far beyond the capacity of actual career counsellors (Chowdhury et al., 2023). AI can be a tool for making access to advanced career information more inclusive by offering timely career information about new career opportunities, oftentimes for students in resource-strapped environments, that include the training and soft skills needed to meet the requirements of the labour market. Early detection of student's risk of educational disengagement through learning analytics and predictive models allows for early intervention before the future of careers at risk.

But the realisation of these opportunities demands imbuing ethical considerations in the design, implementation and governance of the system. AI career counselling systems should complement rather than supplant all the elements of a human career counselling systems, such as the relational and emotional elements of career counselling support by providing data-driven insights for counsellors, while also retaining the ability to support others (Chowdhury et al., 2023). AI can improve the efficiency of career counsellors, enabling them to allocate more time to addressing emotional and motivational aspects of career advising and engaging in more complex discussions, while AI completes the initial assessment and delivery of routine information. With the responsible design, AI can help overcome geographical and socio-economic challenges to quality career guidance, to enable students in rural areas, economically disadvantaged communities and developing countries that are often lacking in trained career counsellors.

5.2 Algorithmic Fairness and Bias Mitigation in Career Recommendations

One of the most important ethical issues in AI career counselling is the fear that algorithms can reinforce the long-standing employment disparities and segmenting of different work roles. A career recommender system can be used on historical trends of the labour market that may lead to system learning and reinforcement of biased patterns, such as overall underrepresentation of women in technical positions and minorities in leadership roles, or the specific underrepresentation of students in disadvantaged backgrounds in lower status occupations (Chowdhury et al., 2023). An assurance of fairness is necessary for algorithmic decisions because they are directly related to important life events themselves.

There are several types of algorithmic bias in career counselling (Arrieta et al., 2019). Recommendation algorithms learn discriminatory patterns because they are trained on biased data that may be influenced by the bias in matters of hiring and promotion throughout history. Insufficient variables when creating a career prediction tool can result in overlooking essential factors and/or misrepresenting protected attributes (e.g., zip code as a proxy for race). The disparity in number of samples for training sets, in which underrepresented groups have fewer examples to go on, can be an exaggeration of their biases in the recommendations made. These issues are further complicated by the possible inherent selection bias in the way that historical data has been gathered, so that some groups could have been more likely to obtain careers and thus the same groups may be more likely to be represented in "successful" outcomes.

Proactive bias detection and mitigation strategies are needed for responsible development (Arrieta et al., 2019). While explainable AI (XAI) is generally encouraged to explain and interpret algorithm decisions, fair-by-design approaches should include an analysis of algorithms to uncover hidden connections in the algorithm between protected characteristics and career recommendations and test the results to see if the algorithm produces alternative recommendations when the demographic features are modified (counterfactual analysis) (Arrieta et al., 2019). It is important to build diverse and representative datasets for training that ensure equal representation across all the groups of demographics. Moving from a onetime fairness assessment to continuous algorithmic auditing, as use of the algorithm evolves and more information is generated, is a necessary step. In consultation with career development professionals and students, fairness metrics suitable to the counselling context should be chosen: There are metrics of fairness that are relevant to the context of career development counselling, including metrics of equal opportunity fairness (equal quality recommendations across groups) and individual fairness (equal recommendations for similar students



regardless of their group).

It's crucial for organizations adopting AI career counselling to communicate clearly about its boundaries. Pupils and counsellors need to be aware that career recommendations are based on observed general patterns, not predictions of a student's success with a given field, and that there is a significant level of uncertainty involved in any algorithmic recommendation due to the importance of individual agency, career motivation, discriminatory hiring practices, and changing labour markets (Chowdhury et al., 2023).

5.3 Privacy, Data Protection, and Student Autonomy

Aptitude tests, interest inventories, academic achievement, background information, family history, socioeconomic status, health status and counsellor notes for issues such as mental health, family conflict, and personal aspirations are some of the types of personal information collected naturally in career counselling systems. Together with other institutional data systems and connected to external data sources (such as social media profiles, online presence, credit checks), this information forms full digital profiles of students that can be used to deduce sensitive characteristics and draw meaningful inferences about a student's future.

Privacy protection is an ethical requirement, and it is one of the pillars (Arrieta et al., 2019). Data governance mechanisms should clearly define who can access career data, for how long, when and how it can be shared, and what the data can be used for beyond supporting career advising. Student and family informed consent processes need to clearly communicate how the data will be used, especially if AI systems will be using that data to provide recommendations. Collection of data must be in keeping with data minimization, which means that it should be limited to the types of data that are essential to the career counselling process, and that it does not follow the practice of "collect everything, analyse later" that leaves a child unnecessarily open to potential misuse.

Various specific privacy concerns have emerged with the use of AI in counselling (Kasneci et al., 2023). AI systems can be very data intensive, pressuring to bring all of the career data together for a number of students and years, reflecting in the exposure and risk of the data if compromised. Algorithmic processing can determine patterns or make inferences which students may have been expecting to remain undisclosed, such as to inform an inference about sexual orientation, disability status or socioeconomic stress from combinations of behavioural indicators. Inspired by shared use of systems across institutional and commercial platforms, there is the possibility of data concentration risk as a single breach potentially affects thousands of students at once. Student autonomy is also undermined when information in the form of algorithmic recommendations is given as black box information, perhaps positioning learners to pursue jobs that are not their preferences, but are those of the institution or the sucker punch company.

The technical, organizational and policy level safeguards are necessary for responsible data practices (Arrieta et al., 2019). Safe infrastructure, access controls and data encryption are used to guard information stored against unauthorized access. The benefits of analysing data can still be achieved while maintaining relevant privacy-preserving techniques, like federated learning (training using dataset on local devices without transferring to central); differential privacy (noise that is added to protect individual records); and data anonymization. Advised "when new" privacy issues arise from systems on vulnerable populations, they should be identified and responded to through regular data audits and impact assessments conducted independently of the system in question. Students should keep meaningful control of their data, that means they have the right to access, correct and delete information, and there should be systems designed to allow these rights.

There is a need to balance privacy concerns with the authentic counselling benefit that will accrue from the fullness of career information. This needs context-specific privacy-by-design solutions, where privacy protection is built into the system's design (rather than added later) and where privacy protection influences the types of data collected and how the algorithms work.

5.4 Transparency and Explainability in AI-Supported Career Guidance

In AI career counselling, transparency and explainability are essential to "trust" and accountability, and to provide human oversight. However, students and counsellors will need to know why certain career pathways are suggested by the algorithms (including what factors were taken into account, how these factors were prioritised, what other pathways were also examined and why they were deemed inferior) (Chowdhury et al., 2023). If there is no transparency, students can't assess a recommendation, they can't see any errors or



biases, and they can't make an autonomous decision.

Achieving algorithmic transparency is itself a significant challenge, especially for sophisticated machine learning systems such as deep neural networks (DNNs) (Chowdhury et al., 2023). Even the creators of the algorithms can't fully explain why their models give the model outputs, as a number of millions of parameters and interactions between the parameters within the algorithms is too complex to comprehend by humans. Additionally, proprietary interests are a key factor in reducing transparency, as such elements often work as a kind of "secrets" for companies. However, students and counsellors have a right to be part of a system of which matters affecting their work are of concern.

Explainable AI (XAI) techniques present several avenues for both more transparency and not greying out the algorithms too much, as described in Arrieta et al. (2019). The Feature importance methods determine which aspect/s of a particular recommendation were most influential—such as determining that STEM aptitude measures held more significance for a software engineering recommendation, while gender had minimal significance. Known as counterfactual explanations, these scenarios cover different contexts where various factors would differ from those in the future, and the results would be different than in the actual case, such as, "If you had better skills at communication, the algorithm would rank management jobs higher. Local interpretable model-agnostic explanations (LIME) generate simplified local models which can approximate an algorithm in given instances, without needing to know the model of the algorithm and thus can explain black-box systems.

But transparency has to be created for an intent audience. Career counsellors require explanations different from students, and different from parents and/or policy makers. Technical explainability (understanding mathematically what processes are happening) differs from intelligibility (understanding real applications of mathematics). Responsible transparency design involves engaging with students, counsellors, and families in constituting what they need to explain to make decisions. Responsible transparency design means engaging students, counsellors, and families in considering what explanations are needed in order to make decisions—explanations that are technical rather than there.

The implementation of transparency in career counselling needs human centred design. Recommendations should be given in terms of a clear range of uncertainty and interval of confidence, remembering that a prediction is not a certainty. Elements that support recommendations and elements of a recommendation that would not generally be included in the algorithm should be emphasized. Interactive explanation systems are presented to the user to which he/she can turn "what if" questions to, which enables an active approach to information rather than the passive one. A professional capacity of the counsellor should be cultivated to understand and convey algorithmic reasoning to students of various skill levels.

Organisational accountability mechanisms go hand-in-hand with transparency. It is important to recognize that independent audits of explanation quality can check that explanations are accurate and meaningful that is, performed by a neutral, third party, other than the system developer. Grievance procedures that allow students to appeal adverse recommendations and for decisions to be appealed to human experts offers practical remedy in event of harmful recommendations based on algorithms. Reporting algorithmic errors, biases and harms are essential to learning and improving in systems.

5.5 Human-Centred AI and the Evolving Counsellor Role

Career guidance by AI needs to be duly centred on the Human - AI collaboration. The AI needs to complement and not supplant the professional judgment of the individual. (Chowdhury et al., 2023) Human career counsellors have unique capabilities: they can listen with empathy to student concerns, appreciate situations where family or personal issues are ambiguous or complex, evaluate conflicting values, and be able to engage in authentic and meaningful relationship building that inspires career development efforts. These are aspects of counselling that cannot be computerised and are integral to the core development of counselling for Career.

AI can bring something unique to the table: it can process massive amounts of data from the labour market to identify trends among numerous students and career outcomes, monitor for new opportunities and shifting job markets, and be available around the clock to provide routine information (Chowdhury et al., 2023). The best Human-AI synergy is that which allocates effort based on the comparative advantage—



algorithms excel in data intensive tasks of pattern recognition, information delivery, while humans excel in interpreting situations, ethical judgment and decision making, building relationships, and problem solving in complex scenarios.

But, achieving the complementary effect between human and AI calls for intentional organizational design. While incorporating AI into traditional counselling approaches can be effective, it can equally leave them with subordinate outcomes compared to either the human counsellor or AI, this is known as "automation bias" when humans passively accept decisions from the algorithm in an uncritical manner, and "selective adherence" when the human rejects the recommendation from AI based on poor or biased reasons. The risks can be prevented only by taking a holistic approach. Counsellors need to be trained on the capabilities, and the limitations of the algorithms they are using, as well as the frequency of failure modes and biases, and indications of how they can be used or misapplied to generate recommendations. Job redesign needs to explicitly uses AI different from human – human should only be used for emotional support, complex decision making, ethical guidance, relationship building with clear demarcation from no mistake in judgment call region to AI.

Cultural aspects are more aligned with the notion of collective intelligence: When humans and AI think together, they can produce better results together (Chowdhury et al., 2023). This means that when building the system, and in its application, multidisciplinary teams of practitioners ensures a continued dialogue between AI developers, career counsellors, psychologists, and ethicists. It demands technology design processes that seek continuous counsellor feedback on the nature or forms of algorithmic support that would make a difference, instead of creating technology that presumes a specified process. It also demands that professional identity and autonomy of career counsellors be maintained, that is, AI should augment, not supplant, the capacity of career counsellors.

6. Policy Implications and Governance Requirements

6.1 National and Institutional AI Governance Frameworks

The governance for the adoption of Responsible AI needs to be aligned across national policy, institutional governance and system level requirements. Governments should formulate national level AI education policies that explicitly mention the ethical, fair, and transparent concerns around AI-based in education (Ng et al, 2023). These strategies must not be just aspirations; they must create enforceable ethical standards that include an obligation on educational institutes, as well as on the AI vendors to adhere to a minimum standard.

Education policy should specify explicit demands for policy-relevant AI systems with respect to students' educational and career trajectory. They may involve required bias auditing checks by third parties prior to deployment in the education context, transparency assessment whether student and counsellor can trace reasoning underlying these algorithms, privacy impact assessment to determine if there are risks to student data, and grievance processes that allow students to appeal algorithmic decisions (Ng, et al., 2023). This calls for implementation to be conducted in line with emerging international standards (ex. EU AI Act, proposed UNESCO AI ethics guidelines for education) while taking into account local differences in national contexts and implementation resources.

The governance frameworks for institutions should include accountability for AI systems. AI career counsellors should be part of an ethics review board within the university or school that is composed of faculty, students, ethicists, and community members tasked with reviewing AI systems before their implementation and tracking the impact of AI systems after they go into effect. These boards should have a decision-making role and should only be advisory in role when there are ethical issues for implementation decisions where systems do not meet ethical criteria. Transparency and accountability of responsibility within institutions for algorithmic decisions, data protection, transparency, and remedy measures in case of harm is crucial for meaningful governance.

6.2 Stakeholder Responsibilities in Responsible Implementation

Clear and responsible governance entails harmonization of different but mutually dependent roles of various players.

Policy makers and governments have a responsibility to set binding moral and ethical ground rules by



legislation and/or regulation and create minimum protection in educational institutions for all, irrespective of inclination towards ethics. Governments need to safeguard educational institutions serving lower-resource groups that may not be able to adequately govern and prevent a 'two tiers' system, whereby the well-resourced can have more responsible governance around AI and the lower-resourced have less accountable systems. Additionally, governments should invest in research and development of responsible AI in the education sector, since the “responsibility premium” associated with creating and deploying ethical AI systems could not be realised in many educational institutes, unless publicly funded.

Educational institutions need to adopt clear policies and responsibility at all governance levels. Before using AI systems, universities and schools should conduct technology impact assessments, studying the impact on various student cohorts, issues with privacy and fairness and ability to establish ethical governance. Strong counsellor capacity needs to be developed in order to work effectively with the AI systems, and the capacity to counter any algorithmic recommendations that counsellor judgment deems that it is warranted. Institutions need resources to continue monitoring systems, testing for biases, and revisiting them on an ongoing basis—the “deploy and forget” framework is bound to be harmful as labour markets and student demographics change over time, and systems will be closer to the past data inputs.

Career counsellors and educators are responsible professionals who must remain primarily faithful to the educational and career development of the learners, not to efficiency nor to algorithmic convenience. Experts need to gain enough knowledge about AI systems to be able to represent students' perspectives in cases where recommendations are biased or potentially harmful and also have the appropriate position of ‘expert’ to ensure they have the freedom not to be simply a useful tool for carrying out the recommendations. Professional associations ought to create and apply methods on ethical use of AI in counselling practice, drawing these clear boundaries from the use of opaque algorithms, lack of any human oversight always, or the use of technology without analysing the human dimension.

It is essential that AI developers and tech firms take an active role in creating ethical systems, not adding an after-thought. This means planning for explainability, incorporating audit features to test for bias after the system has been launched, structuring systems for the least possible gathering of data or putting it out there explicitly, and avoiding commercial schemes that profit from the students' data or distort recommendations for commercial or institutional advantage. Developing AI systems used in education should not be allowed to be marketed in low resource environments unless there is also support for their responsible use, to avoid the geography of ethics, and an education system's algorithms get used in the global south while being limited in the global north.

Students and families should play an informed role in determining, in part, the adoption of the tools; communities impacted by AI systems should have meaningful agency regarding implementation of such systems. This will involve a significant increase in the accessibility of information on the capabilities, limitations, and data practices of the systems incorporating Artificial Intelligence, communicated in accessible language. Student data rights (right to access, correct and delete data) needs to be meaningful and practical rather than theoretical.

6.3 Addressing Global Inequities and Ensuring Digital Equity

Maintaining the integrity of the implementation of responsible AI also demands that it does not further widen educational and economic disparities. One danger is that, if there is a successful role of what I'm calling responsible AI career counselling, that it can only be put in place in universities that are more well-funded, and that, less fortunate populations may get access to these systems, but they are not systems that are as 'responsible' as I would envision them being. This discriminatory effect needs to be proactively avoided in policy frameworks.

A critical lever is open-source development of educational AI systems based on 'public interest' and not profit maximization. Helping to prevent monopolization of the use of responsible AI by well-sponsored organizations, public funds might be invested in developing open-source, auditable AI career counselling tools that would be freely available to schools of any resources. Ethical structures might become standard and not a top-up, so there will be no need for more investment to obtain them, but they will be standard. Ethical systems might become standard and not premium service – so no need if customers want them, they will have



to cover their investment rates, as they are not present by default.

Ensuring students have access to AI career coaching demands meeting students with technology and technical requirements (reliable internet, devices, digital literacy training) to bring all of these things to their hands, instead of the assumption that the technology is available. Digital inclusion initiatives for students who do not have access should be in place alongside the implementation of AI systems, since students without access to the system cannot benefit from the use of AI tools. This is especially relevant in developing countries and in rural areas where the magnitude of the digital divide is still large.

Sustaining equity requires AI to be multilingual and culturally adapted. In order to maintain equity, AI must be multilingual and culturally adapted. Large language models and algorithmic systems are primarily derived from the English language of abundant resources, and have been trained on predominantly English-speaking, high-resource nations, which may exhibit cultural assumptions that are not appropriate in other settings (Kasneci et al., 2023). To be implemented responsibly this has to happen on a global scale, which means localization, cultural adaptation, and testing in various groups and populations, and a commitment to creating systems in ways that do not harm non-dominant populations.

7. Limitations and Future Research Directions

While our major focus in this review has been on papers published in English language sources, and papers coming from developed countries, the record of AI's applications in school education has left us with limited understanding on how the principles of responsible AI are applied in education in developing countries with institutional capacity, financial resources, and cultural context different from the one where these papers have been developed and written. There are few studies focused specifically on lower income educational systems and responsible implementation of AI. Future research needs to involve voice of teachers and students from global south contexts, as well as scholarship to be developed from them.

Although this review was a synthesis of literature, there is limited empirical research to examine the outcomes resulting from the implementation of AI career counselling. The question of whether responsibly using AI for career counsel can truly enhance students' career decisions, increase access to quality career counselling, and decrease occupational segregation by demographic has received scant attention in the literature. Additional data on students' careers, as a result of their AI-generated decisions, and their post-career status would provide more evidence to support responsible use and implementation. Additionally, the pace of rapid technological evolution presents new challenges by delivering novel capabilities, such as generative AI, large language models, faster, and faster, that are difficult to study and keep up with.

8. Conclusion

AI technologies hold considerable promise for career counselling, providing more advanced counselling and guidance, tailoring recommendations to individuals, and assisting educators with data-informed strategies. Yet, these opportunities must be leveraged without perpetuating current inequalities, or safeguarding fundamental rights, and be embedded within the very design, implementation and governance of AI. This paper has been seeking to explain that a set of multiple ethical principles (fairness, transparency, privacy, human oversight, accountability) needs to be synthesized and integrated into responsible AI career counselling, as well as the need for there to be governance mechanisms that clearly divide the responsibility between institutions, professions, policy actors, and actors building the technology.

The conceptual model presented here offers a vision aligning concepts on responsible innovation with career guidance practice, guiding institutions on the opportunity to benefit from AI while upholding professional integrity and ensuring students' interests are well protected. Making this work requires investment of time, training for counsellors, further development of technology with ethical priorities, independent auditing and governance mechanisms, as well as equitable access by all, irrespective of their socioeconomic status. And these investments are not luxury expenses but indispensable for the ethically sound use of AI in a causal field that entails the support of the students' life opportunities: career guidance.

There is a need for dialogue between different worldviews, between the two professional worlds (business and research) and the scholarly communities. Importantly, scholars, AI developers, educators and students whose lives are impacted by such systems must have genuine voice in developing and deploying these systems: career development scholars must be engaged with the literature of artificial intelligence ethics,



while AI developers must be acquainted with career development theory, and policy makers must understand both fields in their integration, before learners are most of all. Responsible AI in career counselling is not just a technical issue, it's a very human one – one that's about how educational institutions are, and should be, going about their business while they are undergoing transformative change.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Funding Statement

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

Informed Consent

Informed consent was obtained from all individual participants included in the study.

Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Data Availability

The datasets generated during and analysed during the current study are available from the corresponding author on reasonable request.

References

- Akinnagbe, O. B. (2024). Human–AI collaboration: Enhancing productivity and decision-making. *International Journal of Education, Management, and Technology*, 2(3), 387–417. <https://doi.org/10.58578/ijemt.v2i3.4209>
- Amari, N. (2022). Trust, acceptance, and power: A person-centered client case study. *Person-Centered & Experiential Psychotherapies*, 21(1), 16–30. <https://doi.org/10.1080/14779757.2022.2028662>
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- Buhăescu-Ciucă, Ș. (2023). Career exploratory intentions and decidedness among students: A career self-management perspective. *Journal of Pedagogy (Revista de Pedagogie)*, 71(1), 29–53. <https://doi.org/10.26755/revped/2023.1/29>
- Chowdhury, S., Dey, P., Joel-Edgar, S., Bhattacharya, S., Rodriguez-Espindola, O., Abadie, A., & Truong, L. (2023). Unlocking the value of artificial intelligence in human resource management through AI capability framework. *Human Resource Management Review*, 33(1), Article 100899. <https://doi.org/10.1016/j.hrmr.2022.100899>
- Folorunso, A., Olanipekun, K., Adewumi, T., & Samuel, B. (2024). A policy framework on AI usage in developing countries and its impact. *Global Journal of Engineering and Technology Advances*, 21(1), 154–166. <https://doi.org/10.30574/gjeta.2024.21.1.0192>
- 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>
- Kolhe, D., & Bhat, A. (2024). Enhancing career guidance with AI in vocational education. In *Transforming education with artificial intelligence* (pp. 539–550). IGI Global. <https://doi.org/10.4018/979-8-3693-7570-9.ch031>
- Laban, O., Owin, O., & Mugenzi, M. M. (2024). Artificial intelligence, the world's giant: Transformative benefits and impact on modern business in Uganda. *International Research Journal of Modernization in Engineering Technology and Science*. <https://doi.org/10.56726/irjmets59710>
- Lent, R. W., & Brown, S. D. (1996). Social cognitive approach to career development: An overview. *The*



- Career Development Quarterly*, 44(4), 310–321. <https://doi.org/10.1002/j.2161-0045.1996.tb00448.x>
- Lu, Q., Zhu, L., Xu, X., Whittle, J., Zowghi, D., & Jacquet, A. (2024). Responsible AI pattern catalogue: A collection of best practices for AI governance and engineering. *ACM Computing Surveys*, 56(7), 1–35. <https://doi.org/10.1145/3626234>
- Masood, F. (2024). The role of AI in shaping the future of labor markets: A comparative analysis of developed vs. emerging economies. *International Journal of Emerging Multidisciplinaries: Social Science*, 3(1). <https://doi.org/10.54938/ijemdss.2024.03.1.346>
- Ng, D. T. K., Leung, J. K. L., Su, J., Ng, R. C. W., & Chu, S. K. W. (2023). Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. *Educational Technology Research and Development*, 71(1), 137–161. <https://doi.org/10.1007/s11423-023-10203-6>
- Renger, S., & Macaskill, A. (2021). Developing the foundations for a learning-based humanistic therapy. *Journal of Humanistic Psychology*, 65(5), 1039–1060. <https://doi.org/10.1177/00221678211007668>
- Rossier, J., Cardoso, P. M., & Duarte, M. E. (2020). The narrative turn in career development theories: An integrative perspective. In S. D. Brown & R. W. Lent (Eds.), *The Oxford handbook of career development* (pp. 169–180). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190069704.013.13>
- Schiff, D. S., Kelley, S., & Camacho Ibáñez, J. (2024). The emergence of artificial intelligence ethics auditing. *Big Data & Society*, 11(4). <https://doi.org/10.1177/20539517241299732>
- Sharma, S., Kumawat, S., & Garg, K. (2021). Predicting student potential using machine learning techniques. In *Proceedings of the International Conference on Artificial Intelligence and Smart Systems* (pp. 485–495). Springer. https://doi.org/10.1007/978-981-16-2594-7_40
- Spors, V., Flintham, M., Brundell, P., & Murphy, D. (2023). Care-full data, care-less systems: Making sense of self-care technologies for mental health with humanistic practitioners in the United Kingdom. *Frontiers in Computer Science*, 5, Article 1230284. <https://doi.org/10.3389/fcomp.2023.1230284>
- Stoltz, K. B., Apodaca, M., & Mazahreh, L. G. (2018). Extending the narrative process: Guided imagery in career construction counseling. *The Career Development Quarterly*, 66(3), 259–268. <https://doi.org/10.1002/cdq.12147>
- Velasquez, P. A. E., & Montiel, C. J. (2018). Reapproaching Rogers: A discursive examination of client-centered therapy. *Person-Centered & Experiential Psychotherapies*, 17(3), 253–269. <https://doi.org/10.1080/14779757.2018.1527243>
- 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 (AI-EDSS). *British Journal of Educational Technology*. Advance online publication. <https://doi.org/10.1111/bjet.13507>