AI-assisted reflection in child welfare

Maximilian Kraus, Jennifer Burghardt, Christopher Koska

Introduction

Recognizing children's rights and the state's responsibility to ensure and promote these rights plays a crucial role in ensuring children's well-being. The most important international human rights instrument for children was created with the UN Convention on the Rights of the Child (United Nations 1989). In countries in which a state responsibility for the protection of children is anchored, different national approaches have been developed to carry out this task. This assignment is often taken on by social workers. Social work is a practice-based profession and an academic discipline which promotes the empowerment and liberation of people based on the principles of social justice and human rights (IFWS 2014). With regard to risk-oriented approaches to social work, the common challenge lies primarily in the diagnosis and prognosis of child welfare hazards (Schrödter 2020). A perfect, faultless decision hardly seems possible. It would require complete information, sufficient time and unlimited cognitive abilities. We are also aware of the distorting influences on human judgement. While the best interests of the child are obviously the guiding principle for professionals, stereotyping, bias and personal experiences of decision-makers can influence their judgment (Gutwald et al. 2021). Considering this, there has been a growing interest in leveraging the potential of Artificial Intelligence (AI) to enhance the decision-making process and promote better outcomes for children and families.

The primary purpose of this article is to explore the role of AI-based reflection in child welfare cases. Reflection in this context refers to the comprehensive analysis and circumstances as well as potential outcomes associated with child welfare cases. By employing AI as a reflective tool, social workers and decision-makers can augment their judgment, enhance accuracy, and gain deeper insights into the dynamics of each case.

The article aims to shed light on the potential benefits and challenges of integrating AI into child welfare practices, with a focus on how AI-based reflection can positively impact the field. It will examine AI applications in data collection, analysis, and decision-making, while emphasizing the importance of transparency and accountability in this domain. Additionally, statements in this article will be underpinned by the findings of the KAIMo project.

The article will argue that it is essential to recognize that while AI has the potential to offer valuable support in child welfare cases, it is not a substitute for human involvement and judgment. Therefore, the scope of this article includes an exploration of how AI can complement and enhance existing child welfare practices while recognizing the unique qualities that social workers bring to their roles. The vision is a harmonious interplay between human expertise and AI-generated insights, which together can provide a more robust, comprehensive framework for safeguarding children and families.

Furthermore, this article goes into specific ethical concerns related to the deployment of AI in sensitive contexts such as child welfare. It emphasizes the multifaceted nature of bias in AI systems, highlighting the necessity to examine and address biases at various levels – from information acquisition and data collection to the interpretation of results. AI not only has the potential to enhance efficiency and accuracy but also enables the exposure of existing prejudices.

The limitations of AI-based reflection in child welfare will also be discussed, including challenges related to data availability, data security, potential resistance to technological adoption, and the need for continuous monitoring and improvements of AI algorithms.

In summary, this article aims to provide an application-oriented guideline to AI-based reflection in child welfare cases. By highlighting the potential benefits, ethical considerations, and challenges, this research seeks to foster informed discussions and responsible implementations of AI technologies in the child welfare domain. As AI continues to advance, understanding its role in supporting child welfare endeavors becomes increasingly crucial in the pursuit of a safer, more nurturing environment for children and families in need.

Understanding Child Welfare Cases

Definition of Child Welfare in Germany

In Germany, the definition of child endangerment, as specified in the case law of the Federal Court of Justice, forms the basis of state action in child protection. Accordingly, the well-being of the child is endangered »if a current danger is identified that is of such magnitude that, as things develop, there is a reasonable likelihood of significant damage to the child's mental or physical well-being« (BGH 2016).

However, determining that a child is endangered does not automatically legitimize state measures to protect children. In Germany, the care and upbringing of children is primarily the natural right of parents and their primary duty (Art 6 (2) GG). It is therefore initially the task of the parents to maintain or restore the protection of the child. The basis for state action and thus intervention in the rights of parents is only given if a child's well-being is found to be endangered and the parents are unwilling or unable to avert these dangers themselves.

Importance of Accurate Reflection

Reflective practice is a fundamental aspect of effective social work and child welfare. It involves the critical examination of one's actions, decisions, and assumptions in the context of the cases being handled. In particular, the requirement for forecasting future developments and the long-term effects of decisions represents a major challenge in the decision-making process of professionals. This is further intensified by aspects such as time pressure and pressure to act, lack of (human) resources, lack of information and the personal responsibility of professionals for the consequences of the decision. These conditions further complicate the already complex task of decision-making (Kinderschutz-Zentrum Berlin, 2009).

Therefore, in addition to a precise diagnosis and risk assessment, collegial case counselling and self-reflection are also required before a decision is made. While collegial case counselling serves to systematically analyze the situation and develop options for action in a goal-oriented manner, self-reflection focuses on analyzing one's own attitude and emotions, e.g. towards the parents and the child. This allows professionals to gain a deeper insight into the complexity of the situation in question and recognize possible personal biases or blind spots that may influence their judgement.

Procedures in Risk Assessments

Various methods and instruments are available to support the decision-making process of the professionals. The risk assessment tools help with the systematic collection and assessment of relevant information. They differ primarily in how the information is evaluated (Schrödter 2020). The different instruments can be classified into two basic categories. On the one hand there are statistical forecasting methods that are based on empirical knowledge of certain characteristics and behaviors that apply specifically to the target group of children and young people at risk. To ensure easy manageability, these are mostly used in the sense of checklists with fixed weightings of individual factors. On the other hand, there are interpretative, hermeneutic forecasting methods in which the domain knowledge of the professionals plays a major role. The evaluation of information usually takes place in the discourse of several actors (Lehmann 2024).

Regardless of the debate about AI, the statistical forecasting methods in particular are critically discussed in the technical discourse. Critics of these instruments argue, among other things, that (group) statistical predictions cannot be applied to individual cases and that categorizing facts does not serve the actual »understanding of the case«. »In this sense, professional action is not characterized by technology orientation and dogmatic adherence to rules, but by an understanding of the case for which scientific knowledge is only one necessary element. This must be supplemented by empirical knowledge and hermeneutic sensitivity to the case« (Ferchhoff and Kurz 1998, p. 23).

The Federal Association of Child Protection Centers e.V. (Bundesarbeitsgemeinschaft der Kinderschutz-Zentren e.V.) also criticizes the use of risk assessment forms. »The use of the questionnaires represents a quality risk in the risk assessment and encourages professional errors«, for example if the design of the questionnaires »a) suggests objectivity and certainty, b) leads to a binding result on its own (points system, traffic light system) and thus leads to the questionnaire and no longer the person making the decision (in the extreme in the sense of a PC software-based evaluation)« (Die Kinderschutz-Zentren 2011, p. 3).

Advocates of statistical prediction methods, on the other hand, emphasize the transparent and verifiable weighting and the influence of certain characteristics on the result of the risk predictions and refer to the numerous individual and meta-studies that prove a significantly higher prediction quality of statistical methods compared to, for example, hermeneutic or clinical methods (Johnson et al. 2015, van der Put et al. 2017, Baird/Wagner 2000, Søbjerg et al. 2021). Particularly with regard to the assessment of child endangerment, it is argued that »a risk assessment, as envisaged by the legislator in child protection, cannot be carried out as a purely interpretative judgement; it is ultimately a classification process« (p. 9). The use of statistical instruments is therefore »helpful, because they can be used to make accurate predictions and more appropriate group categorizations – both occurances that professionals can only do inadequately on the basis of individual case-related interpretative procedures. Classification is therefore not only not wrong, but first and foremost necessary and common. It is most accurately possible through statistical procedures.« (p. 19ff).Given the ongoing development of AI, it is foreseeable that statistical methods will become established (Gutwald et al. 2021, Burghardt and Lehmann 2023). »It will probably be undisputed in the future that computer-aided forecasts are more accurate than forecasts generated by specialists without the support of computers. In the future, it will no longer be a question of the effectiveness of statistical forecasting methods, nor will it be a question of whether social work using statistical methods should allow the practice to take hold at all.« (Schrödter 2020, p. 255ff). In the following chapters, we will critically examine the integration of AI in child welfare risk assessments, highlighting its potential to enhance decision-making while acknowledging the vital role of human expertise.

Artificial Intelligence in Social Services

Overview

Artificial Intelligence refers to the simulation of specific aspects of human intelligence in machines that are programmed to process and analyze data and make decisions based on that analysis. Though these machines ›learn‹ from patterns and experiences in data, it is important to distinguish that they do so without the conscious or subjective feelings humans tend to have when learning from experience. With advancements in machine learning, natural language processing, and data analytics, AI has gained significant traction in various domains, including healthcare, finance, and customer service, where it has demonstrated the ability to streamline processes, improve decision-making, and optimize resource allocation. In recent years, the potential of AI to enhance social services has also been explored more thoroughly, with a growing interest in leveraging AI technologies to address the complexities of child welfare cases as well as predictive policing in law enforcement (Mugari & Obioha 2021).

In the context of social services, AI can be used to support decision-making of social workers and other professionals involved in child welfare. AI technologies, such as machine learning algorithms and natural language processing, enable systems to analyze vast amounts of data quickly, detect patterns, and generate insights that can inform decision-making of social workers.

Applications

AI can potentially be applied at various stages of the child welfare process to support social workers and enhance the overall efficacy of interventions. Some of the key applications of AI in child welfare include:

Data Collection and Analysis

AI can streamline the data collection process by automating the extraction of relevant information from various sources, including case files, reports, and historical data. Through advanced data analytics, AI can identify patterns and trends that may not be apparent to human analysts, thus facilitating a more comprehensive understanding of each case (Waldfogel 2000).

Risk Assessment

AI-powered risk assessment tools can assist social workers in evaluating the level of risk to a child's safety and well-being within their family environment. By analyzing a broad range of risk factors and protective factors, these tools can help anticipate future developments, prioritize cases based on urgency and allocate resources more efficiently (Gillingham 2006).

Decision-making Support

AI can serve as a decision-making support system, offering factual insights and recommendations to social workers when considering different intervention strategies. This collaborative approach can augment decision-making practices based on standardized templates and enables social workers to make more informed and data-driven decisions while maintaining their professional expertise and judgment (Heggdalsvik et al. 2018).

Data collection and Analysis

In child protection, it is crucial for professionals to get a direct impression of the child and its personal environment. All information relevant to the risk assessment should be collected as comprehensively as possible. Based on knowledge of specific risk and protective factors for the child's well-being, systematic observations are carried out and statements are obtained from various actors (e.g. family members, educators/teachers) (Bathke et al. 2019). It must be considered that both personal observations and third-party statements are not only influenced by the subjectivity of the observers but are also shaped through interpretative skills, enabling the integration of collected experiences into a meaningful context, thus go beyond mere subjectivity.

Understanding that data in child protection is an abstraction and can never fully capture the complexity of reality, it is essential that professionals are equipped not just with access to relevant information but also with the methodological and technical means for its appropriate analysis. This section highlights some technical and methodological possibilities for utilizing data within the confines of legal stipulations, while also incorporating preliminary ethical considerations in data handling.

Following the exploration of the technical and methodological possibilities (›What can I do with data?‹ as per the Data Literacy Charter, Schüller et al. 2021), AI technologies offer opportunities to streamline data collection processes and enhance the quality and comprehensiveness of information gathered. A key method of potential AI-based data collection in child welfare is Natural Language Processing (NLP):

NLP algorithms enable systems to analyze unstructured data, such as interview transcripts and case narratives, extracting essential details and identifying patterns in language that may indicate potential risks or concerns (Chowdhary 2020). The technology is capable of processing large data sets, providing relevant information to social workers in a time-efficient manner.

Ethical Considerations in Data Handling

AI-based data collection leads to a complex net of ethical considerations, which are closely linked to legal issues and are intertwined in current regulatory efforts. As a result, many people find it difficult to distinguish genuine ethical considerations from legal or technical issues (Filipović and Koska 2019). In practice, this often leads to ethical considerations being reduced solely to data protection aspects (such as informed consent, purpose limitation, data minimization, transparency etc.) or to specific measures for organizational and technical assurance of information security (especially institutional responsibilities and competencies).

The EU AI Act stands, after the EU GDPR, as another testament to the EU's commitment to highlighting this linkage. As important and valuable as these efforts are, this approach risks that ethical deliberation only takes place in selected expert circles.

Considering the risk classification of the AI Act, the topic of AI assistance to support reflection in child welfare cases, as discussed by Colloseus in this volume, is to be classified as a high-risk application with regulatory requirements. Such a classification not only demands human oversight of the IT system by an expert who can press a stop button at any time, but also the involvement of all affected parties (Koska and Filipović). However, the ethical dimension of the requirement to involve those affected in the data collection and usage process goes far beyond the legal cornerstone of informed consent. From a philosophical perspective, the central challenge is to implement the aspect of ›meaningful human control‹ not only for a High-Level Expert Group (HLEG) – as proposed in the AI Act – but also for social work professionals as well as the affected families and especially the affected children. After all, no one has such privileged access to their own wishes, goals and interests as the people concerned (Koska 2023).

Analyzing and Extracting Insights from Data

AI-powered data analysis has the potential to transform the field of child welfare by enabling social workers to derive valuable insights from vast amounts of information quickly and efficiently. Several key techniques are employed for data analysis in this domain.

One crucial technique is Pattern Recognition, where AI algorithms can identify patterns and correlations within the data that might not be immediately evident to human analysts. More advanced LLMs (Large Language Models) can analyze file cases, and automatically find hints for or against risk status of a child. Most crucially, they have shown emergent abilities in reasoning about whether certain hints contradict or support each other (Wei et al. 2022). This aids in identifying risk factors and protective factors associated with child welfare cases, allowing social workers to make more informed decisions.

Although prone to bias, Sentiment Analysis is another valuable technique, using Natural Language Processing (NLP) to assess the sentiments and emotions expressed in written or verbal communications (Mao et al. 2023). This helps social workers better understand the emotions and perspectives of the children and families involved, enhancing their ability to offer appropriate support.

Enhancing decision making

AI systems driven by data are being utilized more frequently to enhance human decision-making in intricate social situations, such as those found in social work or the legal profession (Kawakami et al. 2022b). AI can support social workers to make informed decisions and tailor interventions to suit the unique circumstances of each case, ultimately leading to better outcomes for vulnerable children (Lysaght et al. 2019).

The dimensions of AI-assisted decision-making in child welfare encompass various critical aspects:

AI-driven risk assessment tools have the potential to analyze a wide array of risk and protective factors associated with a child's situation, providing social workers with a comprehensive evaluation of the level of risk to their safety and well-being (Kawakami et al. 2022a). For that, guidelines like the framework for algorithmic decision-making adapted for the public sector (ADMAPS) can help in the design of high-stakes algorithmic decision-making tools in the child-welfare system (Saxena et al. 2021).

Additionally, studies in the field of healthcare have shown that AI's predictive modeling can anticipate potential outcomes that may arise from a given set of risk factors (Axelrod and Vogel 2003). Mapping these approaches to risk factors in a child's life can help enabling social workers to implement timely and preventive interventions. Early identification of concerns can significantly reduce the likelihood of more severe issues developing later on, offering a proactive approach to child welfare.

In conclusion, AI-generated insights can complement social workers' expertise by providing data-driven information that informs their decisions. This collaborative approach empowers social workers to make well-informed choices while preserving their professional judgment and deep understanding of individual circumstances.

Addressing Bias and Fairness Concerns

Automated decision-making systems (ADM) are often associated with greater objectivity and are therefore – at least in principle – considered inherently fairer than human decision-makers (Hauer 2023). In this context, fairness is typically reduced to the concept of non-discrimination. The primary goal of this approach is to identify biased data sets and avoid distortions in the algorithms used. Various strategies are available for this purpose, such as the use of diverse and representative data, methods for bias detection and correction, Explainable AI, and human oversight.

In our research project on the development of an AI-based assistance system for the assessment of child welfare risks, we found that bias can occur at several points in the assessment, planning and controlling process.

Bias can already occur in individuals who report a suspected child welfare risk to the Youth Welfare Office. These individuals may, for example, keep certain families or groups under disproportionate scrutiny due to prejudice. Another point at which bias can occur is the documentation of case files by social work professionals. These professionals could inadvertently incorporate their own prejudices into the case files, for instance, by highlighting certain information or omitting other. Bias can also play a role in the process of colleague case consultations. Social work professionals may, for example, prefer or reject specific strategies based on prejudice.

Building on the processual view of judgment formation, specific forms of bias can be addressed at particular points in time. In the context of the KAIMo project, nearly 200 potential cognitive biases were screened and assigned to various phases of the decision-making process.[1] With the aim of creating an additional level of reflection for social work professionals, a context-sensitive selection of biases was subsequently displayed to the users for the evaluation of the prototype.

1 The interactive graphic from the Institute for Media and Communication Management at the University of St. Gallen, available online at https://bias-map-vl.web.app, provides a good overview and is useful for an introduction tot he topic.

KAIMo: Overview and Methodology

Project goals and assumptions

The research project KAIMo examines whether and to what extent specific elements of the ethical assessment and decision-making process that guides the actions of youth welfare offices can be encapsulated and translated into algorithms, and whether digital tools may strengthen institutional actions. Particularly due to the complexity of the situation, the legal mandate and the scope of the decisions, the question arises as to whether digital systems can help to make institutional decisions more transparent and well-founded. The research project thus opens a fundamental reflection on the field of tension between social work, ethics and computer science.

The basic assumption of the project is that digital systems can relieve and support the daily work of social workers. Through specific forms of analysis and visual presentation of case data, a faster, more transparent, and more efficient case processing becomes possible.

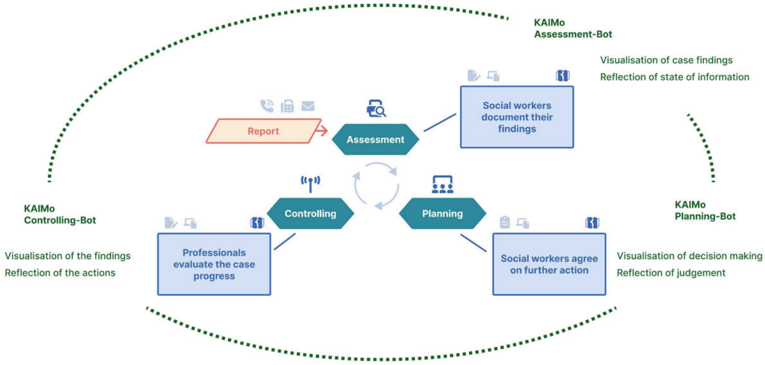
Initially, it was hypothesized that algorithmic decision support could introduce an objective, data-based dimension to the decision-making processes concerning a child's well-being. However, in the course of the project it has become evident that achieving true objectivity is unfeasible due to the inherent subjectivity of the data involved. Considering this, the project's objective has evolved to utilize algorithmic support primarily as a means to reveal blind spots and resolve inconsistencies among the various contributors within the available data, thereby enriching the decision-making process with more comprehensive and subtle insights.

The Three-Agent Approach

Following this guideline, new technologies should not simply be imposed on a social practice, as this can lead to various risks of rejection or improper use.

The KAIMo project worked out an approach to AI integration involving three components (agents) that were designed to ensure a smooth integration in existing processes and established guidelines present in social services. The three agents work together on different levels to provide their respective expertise to the human caseworkers for their case evaluation. They aim to specifically support social work professionals in information gathering (KAIMo Assessment-Bot), decision-making (KAIMo Planning-Bot), and monitoring the implemented measures (KAIMo Controlling-Bot). *Assessment*, *Planning* and *Controlling* are key sub-processes in child welfare services (Figure 1).

Figure 1: Child welfare process



In the first step, existing case data (call logs, notes from professionals, records of home visits) are aggregated and semantically analyzed. The result of this phase is a classified list of case features that provide professional indications for or against the presence of child endangerment.

Building on this classification, the input data is then visualized in a knowledge graph. The structured visual representation of the case characteristics is intended to support professionals in the case analysis and provide case-specific expertise. For example, blind spots can be identified in this phase through a completeness check (comparison with the feature set in similar scenarios), or indications of possible subjective biases of the actors involved can be revealed.

Additionally, context-sensitive reflection questions are asked, derived from the current status and the specific work step of the professionals. These questions are inspired by the idea of the Socratic method, a form of dialogic practice often referred to as ›maieutics‹ or the ›art of midwifery‹, which aims to stimulate critical thinking and illuminate ideas through conversational questioning. For example, the system could ask a question such as »Are there certain assumptions about the family or environment that might be influencing your assessment?« Another question could be »What information is still missing, and how might these missing pieces affect the overall picture?« to point out possible blind spots in the information situation. In this way, the system supports decision-making by guiding professionals to look critically at their own thought processes rather than providing direct decision recommendations.

Ethical decision-making processes cannot be reduced to merely rational deliberations. Central to a moral consciousness are also abilities such as empathy, intuition, and a form of judgment that presupposes embodiment and the ability to choose for or against the good (Koch 2019; Fuchs 2020; Koska 2023). Since algorithms can at best mechanically replicate such characteristics, it's often misleading to consider them as actors in the context of moral decisions. This is because, given the current state of technology, machines cannot have their own moral values or intentions, as they are incapable of establishing a normative-desiring relationship to a moral decision. Therefore, algorithms should aim to improve human decision-making rather than to replace it and to ensure transparency and comprehensibility of algorithmic decisions. This includes implementing mechanisms to review and adjust algorithms to ensure they adhere to ethical standards and are continuously improved.

Human – AI collaboration

The collaboration of the three agents aims to create a balanced approach to decision-making, taking into account both the efficiency of the algorithms and the ethical integrity of the decisions. However, the human component in this system remains crucial. One central pillar and the main guideline of the KAIMo project is an approach to AI integration that is defined by a supportive role. Predictive analytics models are exclusively used to question and improve the current information status (assessment phase), the process of judgment (planning phase), and the effectiveness of measures (controlling phase). The focus is particularly on making the normative criteria guiding the actions of youth welfare offices visible. In this context, algorithms can serve as assistance to support human decision-makers in ethically complex situations. For instance, they can identify patterns in large data sets that are difficult for humans to grasp, providing valuable information for decision-making. It is important to understand the role of algorithms as supporting and complementing human expertise, rather than replacing it.

Central to the approach followed in the KAIMo project is the belief that both professionals and algorithms bring distinct strengths. When synergized, they can offer an unparalleled approach to decision-making, tailored to the unique challenges of child welfare. The KAIMo system, with its Knowledge Graph visualization and Socratic model-inspired reflection questions, acts as a supporting arm to social workers, allowing them to make decisions with a broader, data-informed perspective.

Challenges and mitigation strategies

While applications of AI technologies in social services are manifold, as outlined in the previous chapters, it is important to assess benefits and risks of AI adoption in order to evaluate if and how an introduction of this technology can influence everyday work of social workers.

Previous research in other fields such as agriculture and virology has shown its capacity to provide enhanced data insights (Rawson et al. 2019; Gil et al. 2021). By leveraging AI, child welfare agencies could be enabled to conduct more thorough analyses of data, unveiling valuable insights and trends that might be difficult to identify through traditional approaches (Alufaisan et al. 2021).

On the other hand, it's crucial to recognize that AI models can inadvertently inherit biases from the data they are trained on. Addressing bias in computer systems involves three categories, as identified by Friedman and Nissenbaum (1996, pp. 333–336): pre-existing bias rooted in societal structures, technical bias arising from system design, and emergent bias that evolves during real-world use. AI-assisted reflection on bias, as implemented within the KAIMo prototype, can help to reveal pre-existing biases. However, it's important to recognize that the introduction of AI introduces a heightened complexity, particularly on the technical and emergent levels, which may risk displacing or reinforcing bias, making bias mitigation a multifaceted challenge.

Additionally, the use of AI involves the collection and analysis of sensitive information, raising valid privacy concerns. It is imperative to implement data protection measures to safeguard the confidentiality and security of this information. Ensuring that individuals' privacy rights are respected and upheld is challenging and a paramount ethical consideration (Oseni et al., 2021).

Furthermore, the integration of AI in child welfare should not be seen as a replacement for human expertise. Instead, it should be viewed as a tool to augment and enhance the capabilities of human professionals. Collaboration between AI systems and human experts is essential for ensuring the best outcomes for children and families involved in the system (Wang et al., 2020).

In this chapter, we explore the key challenges associated with AI adoption in child welfare as they were recognized as part of the KAIMo project.

Establishing Accountability Mechanisms

Accountability is crucial in child welfare cases as decisions made can have long-lasting effects on the well-being of children and families. Additionally, depending on local law, social workers are oftentimes held accountable as private persons for their decisions and the results thereof (Döring et al. 2023). To establish accountability in AI-driven child welfare systems, it is essential to define clear roles and responsibilities for all stakeholders involved. The KAIMo project implements a multi-agent solution, which involves human social workers as well as AI algorithms.

Additionally, mechanisms for feedback and evaluation should be put in place. Social workers should be encouraged to provide feedback on the AI's performance, and there should be avenues for addressing concerns and correcting errors that may arise during the decision-making process. These accountability mechanisms are vital in ensuring that AI is viewed as a tool to support social workers rather than replace them, maintaining a human-centric approach to child welfare.

Legal and Ethical Implications

The integration of AI in child welfare raises crucial legal and ethical considerations that demand careful attention to safeguard the rights and well-being of children and families involved.

A child-centric approach is imperative, with AI-driven decision-making prioritizing the safety, well-being, and development of children in all recommendations. Nonetheless, human oversight must be maintained, empowering social workers to retain control over AI-assisted decisions and prevent undue reliance on AI recommendations without critical human judgment, as outline by KAIMo's supportive AI approach.

Introducing AI within the realm of child welfare reveals pivotal legal and ethical dimensions that warrant meticulous scrutiny to champion the rights and welfare of both children and their families. At the epicenter of these considerations is the tension between social work and ethics.

Establishing clear lines of accountability and liability is therefore essential, enabling determination of responsibility in case of errors or adverse outcomes resulting from AI-driven decisions. Professionals therefore need the necessary knowledge of the capabilities and limitations of AI models and the ability to ensure ethical and responsible use of this technology.

One of the foremost concerns in AI-driven child welfare is the security and privacy of sensitive data. To support privacy of sensitive data, AI systems applied in child welfare scenarios should work in a closed system, native to the child welfare institution that applies the AI solution. However, for NLP and machine learning mechanism to work, it is mandatory to grant these processes access to actual case data. After all, the AI should support social workers by giving insightful information and coherences about individual cases.

It was outside the scope of KAIMo to work out detailed guidelines for data handling in child welfare AI applications. These questions deserve dedicated research efforts and should involve legislative and political drivers.

Integration with Existing Systems

Integrating AI systems into the existing child welfare infrastructure can be a complex task. Many agencies already rely on legacy systems, and introducing AI solutions may require substantial changes to the workflow and processes. To address this challenge, the KAIMo project proposes a phased approach to integration.

Initially, AI systems should be introduced in a supplementary role, providing insights and recommendations to social workers without directly replacing any existing processes. This allows for gradual familiarization and adjustment to the new technology. As social workers become more comfortable with the AI's assistance, further integration steps could be taken, streamlining the collaboration between human and AI agents in adherence to the supportive AI approach.

Overcoming Resistance to AI Adoption

Resistance to AI adoption is not uncommon, especially in fields where human decision-making has been the norm for a long time, such as child welfare. End users of AI systems may fear that AI cannot adequately capture and account for the complexity of family systems and life-world influences, disregard professionals' tacit experiential knowledge, or undermine their autonomy in making important decisions. Furthermore, involving social workers in the design and development of the AI system fosters a sense of ownership and trust in the technology. Their feedback is continuously collected and integrated into the AI's improvement processes, ensuring that the system is tailored to meet their specific needs and concerns.

It is also essential to communicate the ethical principles guiding the AI system's development and use to the public. Transparency about how AI is employed in child welfare cases, its limitations, and the steps taken to address potential biases fosters trust and understanding from all stakeholders.

By addressing these challenges and adopting appropriate mitigation strategies, the integration of AI in child welfare cases can be done responsibly and effectively. The KAIMo project serves as a model for how AI can be harnessed to support social workers and improve outcomes for vulnerable children and families while respecting privacy, adhering to ethical guidelines, and maintaining a human-centric approach. As AI technology continues to advance, it is crucial to remain proactive in addressing challenges and ensuring that AI remains a powerful tool in the hands of caring and knowledgeable professionals.

In summary, successfully integrating AI into the child welfare system requires direct actions to ensure its ethical application. This includes employing fairness metrics to evaluate and adjust algorithms, implementing robust bias detection and mitigation strategies, and enforcing stringent data protection measures. Additionally, fostering a collaborative interplay that combines human oversight with AI's analytical capabilities is the most promising approach to enhance decision-making processes. By focusing on these specific and actionable steps, a framework can be created that results in a responsible and beneficial use of AI in this sensitive domain.

References

Alufaisan, Yasmeen, Laura R. Marusich, Jonathan Z. Bakdash, Yan Zhou, and Murat Kantarcioglu. 2021. Does Explainable Artificial Intelligence Improve Human Decision-Making? Proceedings of the AAAI Conference on Artificial Intelligence 35 (8): 6618–6626. doi: 10.1609/aaai.v35i8.16819.

Axelrod, Randy C., and David Vogel. 2003. Predictive Modeling in Health Plans. Disease Management & Health Outcomes 11 (12): 779–787. doi: 10.2165/00115677-200311120-00003.

Baird, C./Wagner, D. (2000): The relative validity of actuarial-and consensus-based risk assessment systems. In: Children and Youth Services Review 22, 839–871.

Bastian, P. (2014). Statistisch Urteilen – professionell Handeln. Überlegungen zu einem (scheinbaren) Widerspruch. Online verfügbar unter: https://www.researchgate.net/publication/265446562_Statistisch_Urteilen_-_professionell_Handeln_Uberlegungen_zu_einem_scheinbaren_Widerspruch

Bundesgerichtshof 2016. Beschluss XII ZB 149/16. Online available: https://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bgh&Art=pm&Datum=2016&nr=76862&linked=bes&Blank=1&file=dokument.pdf.

Burghardt, Jennifer, Lehmann, Robert. 2023. Künstliche Intelligenz und Kinderschutz. Dexheimer, Andreas; Rothballer, Marc (Hg.): Künstliche Intelligenz in der Kinder- und Jugendhilfe. Zeitschrift Jugendhilfe, 05/2023, S. 410–414.

Chowdhary, K. R. 2020. Natural Language Processing. In Fundamentals of Artificial Intelligence, ed. K. R. Chowdhary, 603–649. New Delhi: Springer India.

Die Kinderschutz-Zentren (2011). Empfehlung der Kinderschutz-Zentren zur Nutzung von Gefährdungseinschätzungsbögen in den Kinderschutz-Zentren. Verabschiedet vom Fachausschuss der Kinderschutz-Zentren. Köln. Online verfügbar unter: https://jugendhilfeportal.de/fileadmin/user_upload/fkp_quelle/pdf/Empfehlungen%20zur%20Nutzung%20von%20Gefaehrdungseinschaetzungsboegen.pdf

Döring, Linn K., Mörsberger Thomas, Wapler Friederike (eds.). 2023. Garanten für den Kinderschutz? Nomos Verlagsgesellschaft mbH & Co. KG.

Ferchhoff W, Kurtz T (1998) Professionalisierungstendenzen der Sozialen Arbeit in der Moderne, Neue Prax 28(1)12-26.

Filipović, Alexander, Koska, Christopher. 2019. Corporate Digital Responsibility muss mehr als geltendes Recht abbilden: Zu den »Zehn Regeln für den verantwortungsvollen Einsatz von KI in Human Resources« des Ethikbeirats HR-Tech. https://www.future-of-hr.com/2019/10/corporate-digital-responsibility-muss-mehr-als-geltendes-recht-abbilden/.

Friedman, Batya und Helen Nissenbaum. 1996. Bias in computer systems. In: ACM Trans. Inf. Syst. 14 (3), 330–347. DOI: https://doi.org/10.1145/230538.230561.

Fuchs, Thomas. 2020. Verteidigung des Menschen: Grundfragen einer verkörperten Anthropologie.

Gil, Yolanda, Daniel Garijo, Deborah Khider, Craig A. Knoblock, Varun Ratnakar, Maximiliano Osorio, Hernán Vargas, Minh Pham, Jay Pujara, Basel Shbita, Binh Vu, Yao-Yi Chiang, Dan Feldman, Yijun Lin, Hayley Song, Vipin Kumar, Ankush Khandelwal, Michael Steinbach, Kshitij Tayal, Shaoming Xu, Suzanne A. Pierce, Lissa Pearson, Daniel Hardesty-Lewis, Ewa Deelman, Rafael Ferreira Da Silva, Rajiv Mayani, Armen R. Kemanian, Yuning Shi, Lorne Leonard, Scott Peckham, Maria Stoica, Kelly Cobourn, Zeya Zhang, Christopher Duffy, and Lele Shu. 2021. Artificial Intelligence for Modeling Complex Systems: Taming the Complexity of Expert Models to Improve Decision Making. ACM Transactions on Interactive Intelligent Systems 11 (2): 1–49. doi: 10.1145/3453172.

Gillingham, Philip. 2006. Risk Assessment in Child Protection: Problem Rather than Solution? Australian Social Work 59 (1): 86–98. doi: 10.1080/03124070500449804.

Gutwald, Rebecca, Jennifer Burghardt, Maximilian Kraus, Michael Reder, Robert Lehmann, and Nicholas Müller. 2021. Soziale Konflikte und Digitalisierung Chancen und Risiken digitaler Technologien bei der Einschätzung von Kindeswohlgefährdungen. Ethik Journal 7 (2): 1–20.

Heggdalsvik, Inger Kristin, Per Arne Rød, and Kåre Heggen. 2018. Decision-making in child welfare services: Professional discretion versus standardized templates. Child & Family Social Work 23 (3): 522–529. doi: 10.1111/cfs.12444.

IFWS – International Federation of Social Workers (2014). Global definition of social work. Online available: https://www.ifsw.org/what-is-social-work/global-definition-of-social-work/.

Johnson, W.; Clancy, T. & Bastian, P. (2015): Child Abuse/Neglect Risk Assessment under Field Practice Conditions: Tests of External and Temporal Validity and Comparison with Heart Disease Prediction. Children and Youth Services Review, 56, S. 76–85.

Kawakami, Anna, Venkatesh Sivaraman, Hao-Fei Cheng, Logan Stapleton, Yanghuidi Cheng, Diana Qing, Adam Perer, Zhiwei Steven Wu, Haiyi Zhu, and Kenneth Holstein. 2022a. Improving Human-AI Partnerships in Child Welfare: Understanding Worker Practices, Challenges, and Desires for Algorithmic Decision Support. In CHI Conference on Human Factors in Computing Systems, 1–18. CHI '22: CHI Conference on Human Factors in Computing Systems, New Orleans LA USA. 29 04 2022 05 05 2022. New York, NY, USA: ACM. doi: 10.1145/3491102.3517439.

Kawakami, Anna, Venkatesh Sivaraman, Logan Stapleton, Hao-Fei Cheng, Adam Perer, Zhiwei Steven Wu, Haiyi Zhu, and Kenneth Holstein. 2022b. »Why Do I Care What's Similar?« Probing Challenges in AI-Assisted Child Welfare Decision-Making through Worker-AI Interface Design Concepts.

In Designing Interactive Systems Conference, 454–470. DIS '22: Designing Interactive Systems Conference, Virtual Event Australia. 13 06 2022 17 06 2022. New York, NY, USA: ACM. doi: 10.1145/3532106.3533556.

Kinderschutz-Zentrum Berlin e.V (2009). Kindeswohlgefährdung. Erkennen und Helfen. 11. überarbeitete Auflage. Online available: https://www.bmfsfj.de/resource/blob/94156/178873b3c5a6eeb604568df609e16683/kindeswohlgefaehrdung-erkennen-und-helfen-data.pdf.

Koch, Christof. 2019. The feeling of life itself: Why consciousness is widespread but can't be computed. Cambridge, Massachusetts, London, England: The MIT Press.

Koska, Christopher. 2023. Ethik der Algorithmen. Auf der Suche nach Zahlen und Werten. (1. Auflage 2023.) Berlin: Springer Berlin; J.B. Metzler (Techno:Phil – Aktuelle Herausforderungen der Technikphilosophie, 6). doi: doi.org/10.1007/978-3-662-66795-8.

Koska, Christopher, and Alexander Filipović. 2019. Blackbox AI – State Regulation or Corporate Digital Responsibility? In Digitale Welt 3 (4), 28–31. doi: 10.1007/s42354-019-0208-5.

Lehmann, R. (2024, im Druck). Herausforderungen der künstlichen Intelligenz in der Sozialwirtschaft. In L. Kolhoff (Hg.), *Aktuelle Diskurse in der Sozialwirtschaft V* (S. XX–XX). Springer.

Lysaght, Tamra, Hannah Yeefen Lim, Vicki Xafis, and Kee Yuan Ngiam. 2019. AI-Assisted Decision-making in Healthcare: The Application of an Ethics Framework for Big Data in Health and Research. Asian bioethics review 11 (3): 299–314. doi: 10.1007/s41649-019-00096-0.

Mao, Rui, Qian Liu, Kai He, Wei Li, and Erik Cambria. 2023. The Biases of Pre-Trained Language Models: An Empirical Study on Prompt-Based Sentiment Analysis and Emotion Detection. IEEE Transactions on Affective Computing 14 (3): 1743–1753. doi: 10.1109/TAFFC.2022.3204972.

Mugari, Ishmael, and Emeka Obioha. 2021. Predictive Policing and Crime Control in The United States of America and Europe: Trends in a Decade of Research and the Future of Predictive Policing.

Oseni, Ayodeji, Nour Moustafa, Helge Janicke, Peng Liu, Zahir Tari, and Athanasios Vasilakos. 2021. *Security and Privacy for Artificial Intelligence: Opportunities and Challenges.*

Rawson, Timothy M., Raheelah Ahmad, Christofer Toumazou, Pantelis Georgiou, and Alison H. Holmes. 2019. Artificial intelligence can improve decision-making in infection management. Nature human behaviour 3 (6): 543–545. doi: 10.1038/s41562-019-0583-9.

Saxena, Devansh, Karla Badillo-Urquiola, Pamela J. Wisniewski, and Shion Guha. 2021. A Framework of High-Stakes Algorithmic Decision-Making for the Public Sector Developed through a Case Study of Child-Welfare. Proceedings of the ACM on Human-Computer Interaction 5 (CSCW2): 1–41. doi: 10.1145/3476089.

Søbjerg, L. M., Taylor, B. J., Przeperski, J., Horvat, S., Nouman, H., Harvey, D. (2020). Using risk factor statistics in decision-making: prospects and challenges. European Journal of Social Work, 0(0), 1–14. Routledge.

Schrödter, M., Bastian, P., Taylor B. (2020). Risikodiagnostik und Big Data Analytics in der Sozialen Arbeit. In: Kutscher et al. (Hg.). Handbuch Soziale Arbeit und Digitalisierung. 255–263.

Schüller, Katharina, Henning Koch, and Florian Ramplet. 2021. Data Literacy Charta. Stifterverband (2021): Data Literacy Charta, Online available: https://www.stifterverband.org/data-literacy-charter.

United Nations (20.11.1989). Convention on the Rights of the Child. Online available: https://www.coe.int/en/web/compass/convention-on-the-rights-of-the-child.

van der Put, C. E., Assink, M., Boekhout van Solinge, N. F. (2017). Predicting child maltreatment: A meta-analysis of the predictive validity of risk assessment instruments. Child Abuse & Neglect, 73, 71–88.

Waldfogel, Jane. 2000. »Child welfare research: How adequate are the data?« *Children and Youth Services Review* 705–741.

Wang, Dakuo, Elizabeth Churchill, Pattie Maes, Xiangmin Fan, Ben Shneiderman, Yuanchun Shi, and Qianying Wang. 2020. »From Human-Human Collaboration to Human-AI Collaboration: Designing AI Systems That Can Work Together with People.« *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* 1–6.

Wei, Jason, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, et al. 2022. »Emergent Abilities of Large Language Models.« *Transactions on Machine Learning (08/2022)* 1–30.