ISSN Online 3108-026X, ISSN Print 3139-4833
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 9, September 2026 | DOI 10.17148/IMRJR.2026.030902

Social Work Ethics in the Age of Algorithmic Decision-Making: Rethinking Confidentiality and Informed Consent

Jobin J[1] & Dr. Sandhya R S[2]

Research Scholar, Department of Sociology, University of Kerala[1]

Professor, Department of Sociology, University of Kerala[2]

Abstract: Algorithms are entering child protection and welfare casework faster than the ethical frameworks that govern social work practice can accommodate them. This paper is not concerned with whether such tools are effective; it is concerned with whether confidentiality, informed consent, and professional accountability continue to function once decision-making authority is distributed between a caseworker and a computational system.

We ground the argument in Floridi's information ethics, read alongside the ethics of care and the data justice literature, to show that algorithmic mediation destabilises these commitments at a structural rather than incidental level. Drawing on critical policy analysis of the NASW Code of Ethics, the IFSW Global Statement of Ethical Principles, a systematically identified body of recent empirical and conceptual scholarship on algorithmic decision-making across child welfare, public administration, and healthcare, and current regulatory instruments — the EU AI Act as amended by the 2026 Digital Omnibus, the OECD AI Principles, and India's newly notified Digital Personal Data Protection Rules, 2025 — we identify three structural gaps. Consent frameworks were not designed to anticipate downstream, predictive reuse of client data. Confidentiality provisions presume a practitioner control over information that has, in practice, migrated into opaque proprietary systems. Accountability mechanisms remain fixed on the individual practitioner even as decisional authority has shifted toward institutional and commercial actors.

We examine the counter-argument that algorithmic tools may reduce inconsistency and bias in casework, and we show why this claim, even where empirically supported, resolves only one of the three gaps and leaves consent and accountability untouched. The paper closes with an original accountability framework organised around layered and renewable consent, mandated algorithmic transparency toward practitioners, and institutional co-responsibility for automated decisions, with attention to its implications for social work practice, professional education, and policy.

Keywords: social work ethics, algorithmic decision-making, informed consent, confidentiality, information ethics, data justice, digital welfare

1. INTRODUCTION

Public welfare agencies are adopting algorithmic decision-support tools at a pace that has outstripped the ethical frameworks built around the individual practitioner. Predictive risk scores now inform child maltreatment hotline screening. Eligibility algorithms shape which households receive a home visit. Foster care matching software ranks placement options before a caseworker reviews a case file. These are not speculative scenarios; they describe current practice in child welfare and social service systems across the United States, the United Kingdom, several EU member states, and, increasingly, India (Chouldechova et al., 2018; Gabriels et al., 2025; Saxena & Guha, 2024).

This paper does not evaluate whether such tools improve decision-making outcomes; a substantial and growing empirical literature already addresses that question, with mixed results (Gabriels et al., 2025; Yu & Rose, 2026). Instead, it asks a narrower and, we argue, more consequential question: do social work's foundational ethical commitments — confidentiality, informed consent, and professional accountability — continue to function once decision-making authority is partly transferred to a computational system that the practitioner did not design, does not fully understand, and cannot fully control?

We argue that they do not, and that the erosion is structural rather than incidental. Confidentiality was conceived around a practitioner who could observe, and therefore govern, the movement of a client's information. Informed consent was conceived around a bounded and foreseeable use of information disclosed at a single point in time. Professional accountability was conceived around an identifiable individual exercising discretionary judgment. Algorithmic mediation

unsettles all three assumptions simultaneously. It redistributes effective control to actors who operate entirely outside the licensing and disciplinary structures that govern social work: software vendors, data engineers, procurement officers.

1.1 The Research Gap

Existing scholarship on algorithmic decision-making in human services tends to isolate individual ethical concerns rather than integrate them. One body of work concentrates on algorithmic bias and disparate racial impact in child welfare screening (Chouldechova et al., 2018; Cheng et al., 2022; Gabriels et al., 2025). A second, situated primarily in human-computer interaction and public administration scholarship, concentrates on transparency and explainability as technical and organisational problems (Langer et al., 2021; Busuioc, 2021; Cobbe et al., 2021). A third, situated in health law and bioethics, concentrates on informed consent for the secondary use of patient data in predictive models (Kotsenas et al., 2021; Moulaei et al., 2025). What remains comparatively underdeveloped is a framework that treats confidentiality, informed consent, and professional accountability not as three separate technical problems to be solved independently, but as three manifestations of a single structural condition: the detachment of information control from the practitioner who remains professionally and legally responsible for it. This paper's contribution is to supply that integrative account, grounded in information ethics, care ethics, and data justice, and to translate it into a governance framework addressed specifically to the social work profession rather than to AI developers or public administrators in general.

The paper's contribution can be stated directly: unlike previous studies that examine algorithmic bias, transparency, or informed consent independently, this paper develops an integrated ethical framework that conceptualises confidentiality, informed consent, and professional accountability as interrelated structural consequences of algorithmic mediation in social work, rather than as three separate technical problems requiring three separate fixes.

This paper makes four specific contributions. It integrates three previously separate ethical theories — information ethics, care ethics, and data justice — into a single interpretive lens for algorithmic harm in social work. It synthesises AI governance scholarship from public administration, healthcare, and human rights with social work's own professional ethics literature, fields that have so far developed largely in isolation from one another. It proposes an original accountability framework, specific to social work's licensing and disciplinary structure rather than to AI governance in general. And it extends the analysis, through India's Digital Personal Data Protection Rules and its algorithmic welfare infrastructure, beyond the US- and EU-centred literature that currently dominates this field.

The paper proceeds as follows. Section 2 situates the argument within four adjacent bodies of scholarship — algorithmic governance in public administration, explainable AI, AI ethics in healthcare and human rights, and critical algorithm studies — to establish where social work's ethical literature currently sits relative to cognate fields. Section 3 develops the theoretical framework. Section 4 describes the methodology. Section 5 reviews the contemporary empirical landscape, including algorithmic welfare governance in India. Section 6 identifies the three structural gaps. Section 7 engages the strongest counter-arguments and considers contextual variation. Section 8 develops the proposed accountability framework in detail. Sections 9 and 10 address limitations and conclude.

2. RELATED SCHOLARSHIP

Before turning to social work's own literature, this section locates the paper's concerns within four adjacent fields that have engaged algorithmic decision-making earlier, and in some respects more systematically.

2.1 Algorithmic Governance and Public Administration

Public administration scholarship has developed a substantial literature on algorithmic accountability in government decision-making, largely independent of social work. Bovens' (2007) framework — distinguishing the actor, the forum, and the obligation to explain and justify conduct — remains a standard reference point and has been applied directly to automated decision-making (Busuioc, 2021; Cooper et al., 2022). Cobbe, Lee, and Singh (2021) proposed a “reviewable” automated decision-making framework requiring systems to preserve a record sufficient for ex post accountability, with direct relevance to the confidentiality and accountability gaps in Sections 6.2 and 6.3. Wieringa's (2020) systematic review of this literature found that most existing accountability frameworks address transparency and oversight in the abstract, without close attention to the professional or sectoral context in which a given algorithm operates — a gap this paper addresses by locating the analysis within social work's own licensing and ethical infrastructure.

2.2 Explainable AI and the Limits of Technical Transparency

A parallel literature in computer science and human-computer interaction has developed explainable AI (XAI) as a technical response to opacity. Langer and colleagues (2021) offer an influential stakeholder-oriented account distinguishing what an end-user, an auditor, and a regulator each require from an explanation. This distinction matters here: most XAI research targets developers, auditors, or affected individuals, but rarely the frontline practitioner who

ISSN Online 3108-026X, ISSN Print 3139-4833
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 9, September 2026 | DOI 10.17148/IMRJR.2026.030902

must act on a score in real time, with limited technical training, inside a high-caseload environment. Recent public administration scholarship confirms that this gap persists in deployed systems, since explanation techniques are typically evaluated for technical fidelity rather than for whether they support the comprehension of the person actually using the tool. Section 8.2 argues that mandated transparency toward practitioners — rather than toward the public or the regulator alone — is a distinct and currently neglected requirement.

2.3 AI Ethics in Adjacent Fields: Healthcare and Human Rights

Healthcare offers the most mature comparator literature on AI ethics, informed consent, and confidentiality, and is instructive because clinical and casework settings share a comparable fiduciary structure. Wang and colleagues' (2025) PRISMA scoping review of 309 sources on AI ethics in healthcare found privacy, confidentiality, transparency, and accountability among the most frequently addressed concerns, but that disclosure of AI-generated outputs to the patient — the direct analogue of practitioner transparency toward the client — was comparatively neglected. Kotsenas and colleagues (2021) argue explicitly for rethinking patient consent frameworks in radiology once AI outputs enter the clinical record, on grounds close to those developed in Section 6.1. Moulaei, Akhlaghpour, and Fatehi's (2025) scoping review of thirty-eight studies on patient consent for the secondary use of health data in AI models identified failure to obtain informed consent and violations of confidentiality as the most frequently cited barriers to ethical deployment — findings that mirror, in a different institutional setting, the consent gap identified in Section 6.1.

The human rights literature contributes a complementary vocabulary. Bakiner (2023) surveys the promises and limits of framing algorithmic harms as human rights violations, noting that human rights instruments are typically enforced against states rather than against the private vendors who build these systems — a limitation directly relevant to the accountability vacuum in Section 6.3. Jobin, Ienca, and Vayena's (2019) widely cited mapping of AI ethics guidelines found broad global convergence around principles such as transparency, fairness, and accountability, but far less consensus on how those principles should be operationalised within any specific professional context — the operationalisation this paper attempts for social work.

2.4 Critical Algorithm Studies and Accountability Scholarship

A distinct critical tradition, associated with Cooper, Moss, Laufer, and Nissenbaum (2022) and Widder and Nafus (2023), treats accountability as relational and distributed across an "AI supply chain" rather than residing in any single actor. Widder and Nafus's account of "dislocated accountabilities" among AI developers describes a phenomenon structurally identical to the one this paper identifies in child welfare practice: each actor — developer, vendor, procuring agency, frontline user — can plausibly point to another as responsible. This paper extends that diagnosis into social work's professional ethics, where the dislocation is compounded by a licensing regime that continues to fix liability on the frontline practitioner alone, regardless of how dislocated actual decisional authority has become.

Together, these four literatures establish that the structural problem addressed here — confidentiality, consent, and accountability detaching from practitioner control — is neither unique to child welfare nor novel in the abstract. What is missing is an account that integrates these insights within social work's own professional and ethical infrastructure, the task the remainder of this paper undertakes.

3. THEORETICAL FRAMEWORK

3.1 Floridi and the Informational Self

Floridi's (2013) information ethics treats informational entities — and, by extension, the informational traces a person leaves within institutional systems — as bearing moral weight independent of any subsequent physical or material harm. On this account, a client's case record is not merely a description of the client; it is partly constitutive of the client as an informational organism embedded within an infosphere. Damage to informational integrity therefore constitutes a genuine ethical harm rather than a proxy for some other, ostensibly more tangible injury.

This distinction matters for social work because the profession's ethical vocabulary developed within a largely pre-digital infosphere. Confidentiality rules were designed to guard against a practitioner's indiscretion, not against a vendor's training pipeline. Floridi's framework allows a stronger claim, without conceptual strain: a client is harmed at the moment their case data is repurposed to train a predictive model. This is true independent of whether that repurposing ever produces a visible downstream consequence for the client. The harm is informational, and it is immediate.

3.2 Ethics of Care

Where Floridi supplies the ontological basis for informational harm, the ethics of care supplies its relational stakes. Care ethics, developed through Gilligan, Noddings, and subsequently Tronto (1993), holds that ethical practice is constituted through relationship — through attentiveness, responsibility, competence, and responsiveness — rather than applied to a

relationship from outside it. A caseworker's ethical obligation to a client is not a rule imposed upon the relationship; it is the relationship.

Algorithmic mediation introduces a third party into that relationship without the client's meaningful awareness. A risk score generated by a proprietary model does not merely inform the caseworker's judgment; in high-caseload environments, it frequently displaces it. Saxena and Guha's (2024) two-year ethnographic study of a United States child welfare agency documented caseworkers relying heavily on algorithmic tools for decisions concerning mental health needs, foster placement, and trafficking risk, despite lacking the statistical training required to interrogate what a given score actually represented. Care ethics supplies the vocabulary for what is lost in this arrangement: not merely accuracy, but responsiveness. A model cannot be attentive to a client in Tronto's sense. It can only be calibrated.

3.3 Data Justice

Data justice, as developed by Dencik and colleagues (2022), reframes the analysis from individual harm to distributional harm: who bears the burden of datafication, and who benefits from it. Predictive tools in child welfare and social services are disproportionately deployed on populations with the least capacity to contest their use — low-income families, racial and ethnic minorities, and, in the Indian context, populations navigating welfare bureaucracies with limited digital literacy or legal recourse. Gabriels, Kuiper, and Harder's (2025) scoping review of algorithmic tools in child protection found that ethical concerns, particularly racial disparities, surfaced in nearly every study reviewed, including studies of tools judged technically promising. Data justice asks a sharper question than whether a tool is accurate. It asks accurate for whom, and at whose expense.

Read together, these three frameworks perform distinct analytical work. Floridi establishes that informational harm is harm in its own right. Care ethics explains why algorithmic mediation degrades the relational core of social work practice specifically, as distinct from other professions. Data justice explains why that degradation is unevenly distributed across client populations. The three structural gaps identified in Section 6 sit at the intersection of all three frameworks.

4. METHODOLOGY

This is a conceptual and critical policy paper rather than an empirical study; fieldwork for the authors' broader doctoral research programme on datafied welfare governance has not yet commenced, and this boundary is stated explicitly rather than implied. The paper draws on three categories of secondary source, identified through a structured though non-systematic review process appropriate to conceptual scholarship.

First, professional ethical codes: the NASW Code of Ethics (2021 revision) and the IFSW Global Statement of Ethical Principles (2018), read in full against the analytical categories of consent, confidentiality, and accountability.

Second, empirical and conceptual scholarship on algorithmic decision-making. Searches were conducted in Google Scholar, Scopus, PubMed, and the ACM Digital Library between January and July 2026, using combinations of the terms “algorithmic decision-making,” “child welfare,” “social work ethics,” “algorithmic accountability,” “explainable AI,” “informed consent AI,” “data justice,” and “algorithmic governance,” restricted where possible to 2019–2026 to prioritise contemporary regulatory and empirical developments, with earlier foundational theoretical sources (Floridi, 2013; Tronto, 1993; Bovens, 2007) included regardless of date. Sources were included where they addressed algorithmic decision-support in a public service, human service, healthcare, or public administration context, and specifically engaged at least one of consent, confidentiality, transparency, or accountability. Purely technical machine-learning papers without an ethical or governance dimension were excluded, as were opinion pieces without an identifiable evidentiary or theoretical basis. This process identified approximately fifty sources meeting inclusion criteria, of which the thirty-two most directly relevant are cited in this paper; systematic reviews and scoping reviews were prioritised as synthesis sources over individual primary studies where both were available, consistent with an evidence-hierarchy approach appropriate to a conceptual paper of this scope.

Third, primary regulatory texts: the EU AI Act and its 2026 Digital Omnibus amendments, the OECD AI Principles (2019), and India's Digital Personal Data Protection Rules, 2025, notified by the Ministry of Electronics and Information Technology in November 2025. These were read against the empirical and theoretical material using a critical policy analysis approach, asking not whether the codes and regulations achieve their stated aims, but what assumptions about practice, consent, and accountability they encode, and whether contemporary algorithmic practice continues to satisfy those assumptions.

ISSN Online 3108-026X, ISSN Print 3139-4833
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 9, September 2026 | DOI 10.17148/IMRJR.2026.030902

5. THE CONTEMPORARY LANDSCAPE IN CHILD WELFARE AND SOCIAL SERVICES

The empirical record on algorithmic tools in child welfare is more ambivalent than a straightforward account of harm would suggest, and that ambivalence is itself analytically significant.

Yu and Rose's (2026) systematic review of algorithmic-assisted decision-making tools in child welfare practice, based on nine studies meeting PRISMA inclusion criteria, found that responsible implementation remains the exception rather than the rule, with substantial methodological and empirical gaps in how fairness and equity are evaluated once a tool reaches deployment. Only two of the included qualitative studies applied an explicit theoretical framework to interpret their findings — a gap that is analytically significant rather than a minor citation omission, since it indicates that most implementations proceed without a systematic account of what kind of harm might follow from redistributing decisional authority.

A parallel scoping review by Gabriels, Kuiper, and Harder (2025), covering twenty-four studies published between 2000 and 2022, reached a comparable conclusion by a different route. The Allegheny Family Screening Tool and the Child and Adolescent Needs and Strengths (CANS) instrument emerged as the most frequently referenced systems, with genuine evidence that both could improve accuracy in identifying children most in need of intervention. Ethical concerns, particularly regarding racial disparities, nonetheless surfaced in almost every study examined, leading the authors to conclude that algorithmic decision support in child protection remains, in their assessment, still in its infancy. Independent investigative reporting on the Allegheny tool found that it flagged a disproportionate share of Black children for mandatory neglect investigations, a finding sufficiently serious that neighbouring Oregon discontinued its own comparable tool rather than continue to defend its use. Evidence of technical promise and evidence of structural harm are therefore not distributed across different studies of different tools; they recur within the same studies of the same tools.
Saxena and Guha's (2024) ethnographic account adds a dimension outcome-focused reviews cannot capture. Within their two-year study of a United States child welfare agency, caseworkers legally required to use CANS to score a child's mental health needs — a score that in turn determined foster parent compensation — began adjusting scores upward in response to a chronic shortage of quality foster placements, in order to make placements financially viable. This is not a failure of predictive accuracy; it demonstrates that once a score becomes load-bearing for resource allocation, practitioners route around it in ways its designers did not anticipate, invisible to both the vendor and, eventually, to the agency itself.

Kawakami and colleagues' (2022) participatory design research offers a constructive counterpoint: frontline workers, given a genuine role in the design process, can meaningfully shape how algorithmic tools function within casework rather than simply absorbing a vendor's finished product. Participatory design nonetheless remains the exception within current procurement practice; most tools reach agencies as completed, proprietary products, with practitioners lacking any formal channel through which to interrogate, let alone modify, their operation.

Considered as a whole, this literature supports neither an unqualified optimistic nor an unqualified pessimistic reading. It supports the more specific structural claim advanced here: these tools operate within institutional environments — chronic understaffing, resource scarcity, high caseworker turnover — for which the ethical codes governing individual practitioners were never designed. The difficulty is not that these are poorly engineered tools; it is that they enter a governance vacuum that current professional ethics has not yet been reconstructed to address.

5.1 Algorithmic Welfare Governance in India

The literature reviewed above is concentrated in the United States and Europe, and the structural gaps this paper identifies recur, in a different institutional form, within India's own digital welfare infrastructure. Aadhaar, India's biometric identity system, has become a de facto precondition for access to major welfare programmes, including the Public Distribution System and the National Rural Employment Guarantee Act. Khera's (2017) review of Aadhaar's impact on welfare delivery found limited evidence of the corruption reduction the system was designed to achieve, alongside substantial evidence of exclusion: authentication failures — biometric mismatches, connectivity gaps, worn fingerprints among manual labourers and elderly beneficiaries — routinely operate as a de facto eligibility filter, denying entitlements to individuals who remain legally qualified for them. Chaudhuri's (2022) ethnographic study of the Aadhaar-enabled Public Distribution System describes this dynamic as “programmed welfare”: algorithmic sorting that is enacted, in practice, through an unstable assemblage of biometric devices, databases, and street-level intermediaries, producing outcomes considerably less deterministic — and less accountable — than the technology's design would suggest.

This Indian case exhibits the same three structural gaps identified in Section 6, in a welfare-benefits register rather than a child-protection register. Consent to biometric enrolment was, for most beneficiaries, effectively compulsory rather than freely given, since enrolment became a precondition for accessing existing legal entitlements. Confidentiality and

ISSN Online 3108-026X, ISSN Print 3139-4833
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 9, September 2026 | DOI 10.17148/IMRJR.2026.030902

data control rest with the Unique Identification Authority of India and connected databases rather than with the frontline ration dealer or block-level welfare officer who faces the beneficiary directly. And accountability for a wrongful denial of entitlement remains difficult to locate: the authentication failure may originate in a biometric sensor, a connectivity outage, a database mismatch, or a street-level functionary's discretion, with no single actor clearly positioned to answer for the outcome. India's DPDP Rules, 2025, discussed in Section 6.1, represent a first step toward addressing the consent dimension of this problem, but do not yet speak directly to the accountability diffusion Chaudhuri's fieldwork documents. This suggests that the accountability framework proposed in Section 8, developed primarily from social work's child welfare literature, has direct application to India's algorithmic welfare infrastructure as well, and that social work practitioners engaged in welfare administration in the Indian context face a version of this paper's central problem already, independent of when child-welfare-specific algorithmic tools of the kind documented in Section 5's US literature arrive in the Indian system.

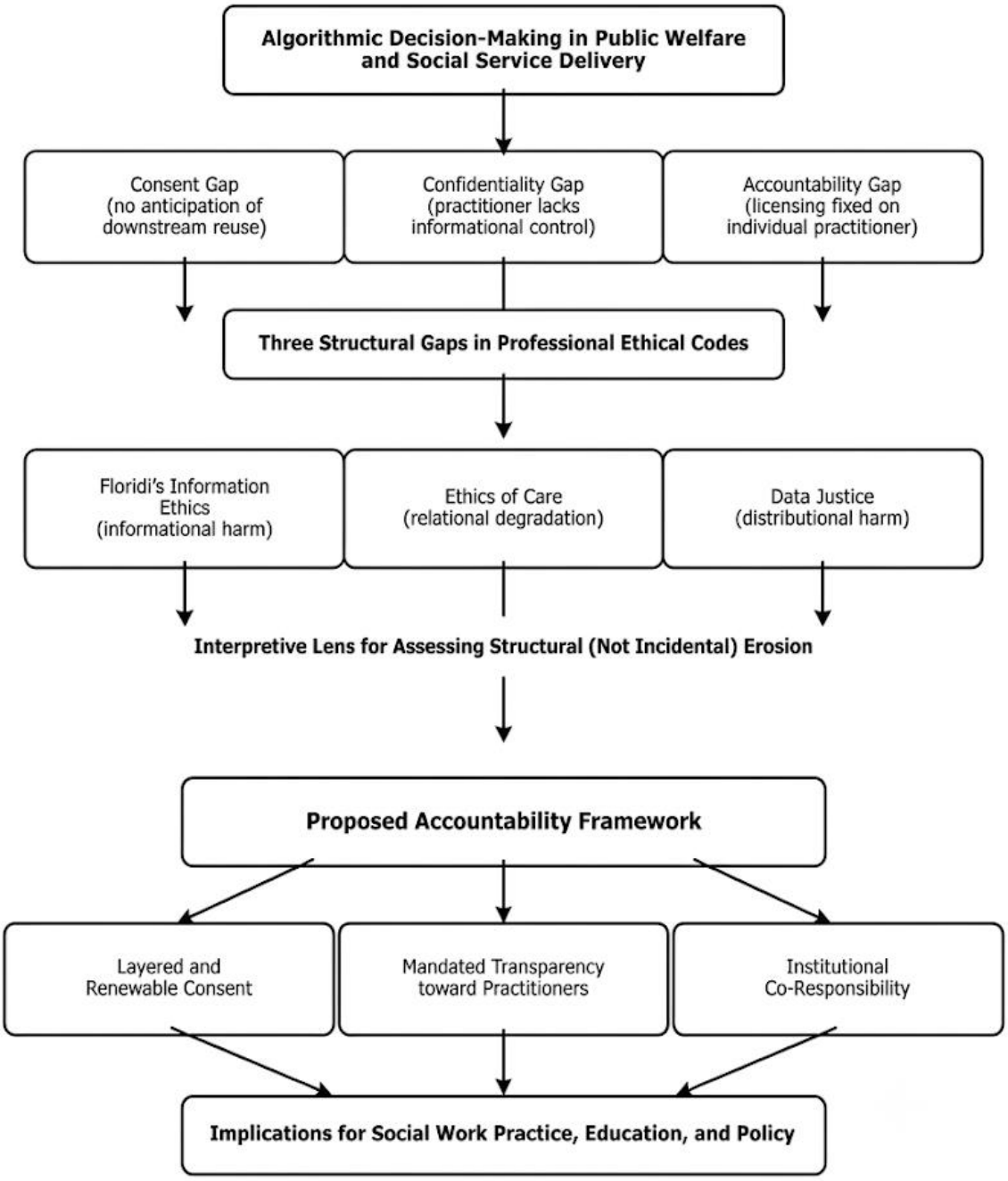

Figure 1. Conceptual framework linking algorithmic mediation, the three structural ethical gaps, their theoretical grounding, and the proposed accountability model.

ISSN Online 3108-026X, ISSN Print 3139-4833
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 9, September 2026 | DOI 10.17148/IMRJR.2026.030902

6. THREE STRUCTURAL GAPS

6.1 Consent That Does Not Anticipate Reuse

The NASW Code of Ethics and its consent provisions were drafted around a transactional model: a client agrees to a specific disclosure, for a specific purpose, at a specific point in time. This model assumes the practitioner can describe, in advance, the full range of uses to which the data might be put. Predictive analytics inverts that assumption. A risk score generated today may train tomorrow's model on a population that never consented to inclusion. Downstream reuse means repurposing casework data collected for direct service delivery into training data for a predictive system. This falls entirely outside what most clients understood themselves to be agreeing to, and frequently outside what the intake worker herself understood the agency's data infrastructure to be capable of. This pattern is not unique to child welfare. Moulaei, Akhlaghpour, and Fatehi's (2025) scoping review of patient consent for the secondary use of health data in AI models identified precisely this failure, processing personal data beyond the scope of original consent, as among the most frequently cited ethical barriers in healthcare AI deployment. Kotsenas and colleagues (2021) argue on comparable grounds for restructuring clinical consent once AI outputs enter the medical record.

India's newly notified Digital Personal Data Protection Rules, 2025 — notified by the Ministry of Electronics and Information Technology on 14 November 2025, operationalising the 2023 Act — address this problem partially, through purpose limitation and itemised consent notices. Data fiduciaries must issue standalone, plain-language notices specifying what data is collected, the exact purpose, and a defined retention period, and Significant Data Fiduciaries face an additional obligation of algorithmic due diligence to verify that their systems do not endanger a data principal's rights. This genuinely improves on having no comparable rule at all. However, purpose limitation assumes that purposes are static and singular, whereas predictive modelling is iterative by design: a model retrained on an expanded dataset continues to serve the same stated purpose — risk assessment — while performing a function the originally consenting client did not anticipate. The rule closes the front door to unconsented use while leaving a side door open through iterative retraining.

6.2 Confidentiality Without Control

NASW's confidentiality provisions presume that the practitioner functions as custodian of client information, determining who may access it and able to trace its movement. Once case data enters a proprietary vendor's system, that presumption no longer holds. The practitioner frequently cannot inspect the model's inputs, cannot audit what data has been retained, and cannot confirm what becomes of the record once the vendor's contract with the agency terminates. Confidentiality, as a professional obligation resting on the individual practitioner, has become detached from the practitioner's actual capacity to exercise control over information. The obligation has not moved; the capacity to discharge it has.

6.3 An Accountability Vacuum

This is the sharpest of the three gaps. Current regulation addresses it least directly. Professional licensing exposes the individual social worker to liability for decisions she makes. But when a risk score materially shapes, or in practice substitutes for, that decision, the actor with the greatest influence over the outcome sits entirely outside the licensing structure. The vendor who built the model faces no professional accountability. The agency administrator who procured the tool faces, at most, an administrative consequence rather than a professional one. The individual caseworker remains the only party whose licence is exposed.

This pattern has a name in the philosophy of technology literature, and engaging with that literature more directly clarifies what kind of problem it is. Nissenbaum (1996) identified four barriers that computerisation creates for moral accountability, the first and most relevant here being the problem of many hands: when a task is distributed across enough actors, no single one of them satisfies the conditions ordinarily required to hold a person responsible, even though the outcome is clearly the product of their combined actions. Matthias (2004) sharpened this into what he termed the responsibility gap, arguing that learning systems can behave in ways their own designers could not have predicted, which breaks the traditional link between an operator's foreseeability and her liability for what the system does. Santoni de Sio and Mecacci (2021) later showed that this is not one gap but at least four distinct ones, spanning culpability, moral accountability, public accountability, and active responsibility, each with different causes and each requiring a different institutional remedy.

Read against this literature, the accountability vacuum in child welfare casework is a sector-specific instance of a much more general phenomenon in sociotechnical systems. Many hands are already distributed across the vendor's engineering team, the agency's procurement office, and the frontline practitioner. Foreseeability has already broken down in the sense Matthias describes, since the vendor cannot fully predict how a retrained model will score a given family, and the caseworker cannot audit why it did. What social work's professional codes have not yet done is what Santoni de Sio and Mecacci's typology suggests is necessary: disaggregate the single, undifferentiated notion of “accountability” in the

NASW and IFSW codes into distinct culpability, oversight, and remedial obligations, and assign each to the actor actually positioned to discharge it, rather than defaulting all of them onto the practitioner by omission.

The EU AI Act, as originally scheduled, classifies systems used by public authorities to determine eligibility for social security benefits and other essential public services as high-risk under Annex III. This triggers conformity assessment, mandatory human oversight, incident logging, and registration in an EU-wide database, representing meaningful institutional accountability, at least on paper. The European Commission's Digital Omnibus on AI was published on 19 November 2025. Parliament, Council, and Commission reached political agreement on it on 7 May 2026, and that agreement defers the compliance deadline for Annex III high-risk systems from 2 August 2026 to 2 December 2027. That is an eighteen-month gap in which exactly the child welfare and social benefit systems this paper is concerned with keep operating under the thinner, pre-existing transparency regime. Institutional accountability, in the EU framework, is scheduled but not yet operative.

7. DISCUSSION: ALTERNATIVE VIEWPOINTS AND CONTEXTUAL VARIATION

7.1 The Bias-Reduction Counter-Argument

The strongest counter-argument available holds that algorithms reduce inconsistency and bias in casework, and this claim is not without empirical support. Cheng and colleagues' (2022) research on how child welfare workers actively work to reduce racial disparities in algorithmic decisions shows that some frontline practitioners take equity considerations seriously, treating an algorithm's output as one input to weigh rather than a final determination.

Consistency, however, is not equivalent to fairness. A biased algorithm applied consistently produces consistent bias at scale, without the friction of individual hesitation that sometimes interrupts a biased human decision before it is made. Gabriels and colleagues (2025) found racial disparity concerns in nearly every study of these tools, including those judged most technically promising, which suggests that consistency can function as a mechanism through which bias is industrialised rather than corrected. Nor does the bias-reduction claim address consent or accountability: a perfectly unbiased predictive tool would still repurpose client data without renewed consent and would still diffuse decisional responsibility across actors who bear no professional accountability for the outcome. Bias reduction, if achieved, resolves one of the three structural gaps identified in this paper; it leaves the other two untouched.

7.2 What AI Developers and Policymakers Might Argue

A fair account of this debate should consider the perspective of those who design and regulate these systems. AI developers might reasonably argue that predictive tools formalise decision criteria that were previously implicit and unaudited within human casework, and that formalisation creates the documentary trail necessary for meaningful accountability in the first place. This argument has some force, but the difficulty, addressed in Section 6.2, is that this documentary trail is frequently accessible to the vendor and the agency's data infrastructure, but not to the practitioner or the client — converting a potential accountability gain into a further consolidation of informational control away from the people the record concerns.

Policymakers, for their part, often frame algorithmic tools as a necessary response to caseworker shortages and unsustainable caseloads, a framing with clear justification given the workforce pressures documented in Saxena and Guha's (2024) fieldwork. This argument, however, risks treating algorithmic decision-support as a substitute for adequate staffing rather than a complement to it, normalising resource scarcity as the baseline condition under which algorithmic mediation becomes acceptable, rather than treating adequate staffing as a precondition algorithmic tools should support rather than replace.

7.3 Contexts Where Predictive Tools May Improve Ethical Practice, and Variation Across Welfare Domains

It would be inaccurate to suggest no context favours algorithmic support. Kawakami and colleagues (2022) show that participatory design processes, where practitioners help shape a tool before deployment, can produce systems that practitioners trust and use appropriately rather than over-relying upon or covertly circumventing. Screening-stage triage — flagging cases for closer human review rather than determining outcomes directly — represents a narrower and comparatively defensible use case than the load-bearing resource-allocation role CANS came to occupy in Saxena and Guha's fieldwork. The distinction that matters, across this literature, is not whether a tool is used, but whether its output remains contestable and subordinate to practitioner judgment, or becomes, in practice, the decision itself.

The ethical stakes identified in this paper are also unlikely to be uniform across welfare domains. Child protection decisions carry a distinct combination of urgency, irreversibility, and state coercive power that elder welfare eligibility determinations or disability benefit assessments do not share in the same configuration, even though the consent and confidentiality issues identified in Section 6 recur across all of them in some form, as India's Aadhaar-linked welfare

ISSN Online 3108-026X, ISSN Print 3139-4833
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 9, September 2026 | DOI 10.17148/IMRJR.2026.030902

infrastructure illustrates. The accountability framework proposed in Section 8 is intended as a general structure to be calibrated to context rather than a uniform prescription; a jurisdiction's child protection algorithms plausibly warrant more stringent practitioner-transparency requirements than a non-coercive community support waitlist tool, even though both fall within the scope of the underlying argument.

8. TOWARD AN ACCOUNTABILITY FRAMEWORK

This section develops the paper's central original contribution: an accountability framework organised around three principles, each specified here with attention to implementation, institutional and practitioner responsibilities, and policy implications.

8.1 Layered and Renewable Consent

Principle. Consent obtained at intake should not function as a permanent authorisation for all subsequent uses of a client's data, including uses not contemplated at the time of collection.

Implementation. Agencies should distinguish, in their consent architecture, between consent to direct-service data collection and consent to the use of that data in training, retraining, or validating a predictive model. The latter should be renewed at defined intervals — we suggest annually, or at any point a model is materially retrained — and clients should be able to withdraw the latter without losing access to direct services. India's DPDP framework already establishes relevant infrastructure through its Consent Manager mechanism; nothing prevents child welfare and social service agencies elsewhere from adopting an equivalent renewal cycle voluntarily, ahead of any legal mandate that may eventually require it.

Institutional responsibility. The procuring agency, not the individual caseworker, should bear responsibility for building this renewal mechanism into its data infrastructure and vendor contracts.

Practitioner responsibility. Practitioners should be responsible for explaining, in comprehensible terms, what predictive use has been made of a client's data since the previous consent renewal — a task that is only possible if the transparency obligation described in Section 8.2 is also met.

Policy implication. Professional bodies such as NASW and IFSW should incorporate a renewable-consent standard into their codes of ethics, rather than treating consent as satisfied by a single intake-stage disclosure.

8.2 Mandated Algorithmic Transparency Toward Practitioners

Principle. Transparency obligations should be directed toward the caseworker using a tool, not only toward the public or the regulator, on the grounds established in Section 2.2: a practitioner who cannot see which variables produced a risk score cannot exercise the professional judgment her licence requires of her.

Implementation. Agencies procuring algorithmic decision-support tools should require, as a contractual condition of purchase rather than a negotiated extra, vendor disclosure of the model's input variables, any known limitations or documented disparities, and a record of material updates to the model since deployment. This is deliberately a floor requirement rather than full model interpretability, which even the XAI literature acknowledges remains technically unresolved for many model classes (Langer et al., 2021).

Institutional responsibility. Agencies should treat this transparency requirement as a procurement standard, comparable to the technical documentation obligations already required of Significant Data Fiduciaries under India's DPDP Rules and of high-risk system providers under the EU AI Act.

Practitioner responsibility. Practitioners retain the professional obligation to exercise independent judgment rather than deferring automatically to an algorithmic score, a obligation that current NASW and IFSW codes already imply but do not state explicitly in relation to algorithmic tools.

Policy implication. Regulatory bodies overseeing social work licensure should require, as a condition of an agency's use of algorithmic decision-support, evidence that this practitioner-facing transparency standard is met, independent of whatever public-facing transparency obligations may separately apply under AI-specific legislation.

ISSN Online 3108-026X, ISSN Print 3139-4833
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 9, September 2026 | DOI 10.17148/IMRJR.2026.030902

8.3 Institutional Co-Responsibility for Automated Decisions

Principle. Professional accountability structures should expand to name the procuring agency and, where relevant, the vendor as co-responsible parties when an algorithmic tool materially shapes a casework decision, rather than treating the individual practitioner as the sole locus of professional liability.

Implementation. This requires renegotiating, rather than abandoning, individual practitioner accountability: the practitioner remains accountable for the decision she makes, but the agency becomes accountable for the adequacy of the transparency and consent infrastructure surrounding the tool she is required to use, and the vendor becomes contractually accountable for the accuracy of its documented limitations.

Institutional responsibility. Agencies should establish an internal review function — distinct from, and in addition to, individual practitioner supervision — specifically responsible for auditing algorithmic tool performance and disparate impact on an ongoing basis, mirroring the Data Protection Impact Assessment obligations already required of Significant Data Fiduciaries under India's DPDP Rules.

Practitioner responsibility. Practitioners should have a documented, low-friction mechanism for flagging suspected algorithmic error or bias without that flag being treated as a challenge to their own competence — addressing directly the score-manipulation workaround Saxena and Guha (2024) documented, which arose in the absence of any legitimate channel for practitioners to contest a score they judged inaccurate.

Policy implication. This layered co-responsibility model roughly mirrors the deployer and provider obligations already established under the EU AI Act's high-risk framework, once its delayed provisions take effect, and social work's own professional codes should adopt an equivalent structure rather than waiting for AI-specific legislation to arrive in every jurisdiction in which the profession operates.

8.4 Relationship to Existing Governance Models

This framework is deliberately narrower in scope than general-purpose AI governance instruments such as the NIST AI Risk Management Framework or the Council of Europe's 2024 Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law, both of which address algorithmic accountability across all sectors rather than within a single licensed profession. Its distinguishing feature is that it is addressed to a professional body — NASW, IFSW, and their international and national counterparts — that already possesses licensing and disciplinary authority over its members, and can therefore operationalise these principles through existing ethics codes and continuing-education requirements rather than through new legislation. Where the Cobbe, Lee, and Singh (2021) "reviewable" automated decision-making framework specifies what a record must preserve for external auditors, the framework proposed here specifies what a practitioner must be able to see and act upon in real time, and where Bovens' (2007) accountability framework identifies actor, forum, and obligation to explain, this paper's contribution is to specify who occupies each of those three roles within the particular institutional configuration of contemporary social work practice.

8.5 Implications for Policy and Practice

For professional bodies, the immediate implication is that NASW's and IFSW's ethics codes should be revised to state explicitly how confidentiality and consent apply to algorithmic tools, rather than leaving practitioners to infer this from principles drafted before such tools existed. For agencies, the implication is that procurement decisions should treat practitioner-facing transparency and renewable consent as non-negotiable contractual terms rather than optional vendor features. For regulators, the implication is that the eighteen-month gap created by the EU AI Act's Digital Omnibus deferral, and comparable gaps in jurisdictions without equivalent legislation at all, should not be read as licence to delay institutional accountability measures that professional bodies can adopt independently of statute. Table 1 summarises the framework's core recommendations by stakeholder.

Table 1. Practice Implications by Stakeholder

Stakeholder	**Recommended Actions**
Social workers	Explain, in plain language, what predictive use has been made of a client's data since the last consent renewal; exercise independent professional judgment rather than deferring automatically to an algorithmic score; use the documented channel for flagging suspected algorithmic error.
Agencies	Build renewable-consent cycles into data infrastructure and vendor contracts; make practitioner-facing model transparency a contractual condition of

ISSN Online 3108-026X, ISSN Print 3139-4833
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 9, September 2026 | DOI 10.17148/IMRJR.2026.030902

	procurement, not a negotiated extra; establish an internal audit function for ongoing disparate-impact review.
Vendors	Disclose model input variables, known limitations or documented disparities, and a record of material updates since deployment, as a standard term of the procurement contract.
Regulators / professional bodies	Revise NASW, IFSW, and equivalent national codes to state explicitly how confidentiality and consent apply to algorithmic tools; require evidence of practitioner-facing transparency as a condition of an agency's licensure to use algorithmic decision-support.

9. LIMITATIONS

This paper is conceptual in nature, and its claims should be read accordingly. It relies on secondary sources rather than original empirical data; the authors' own fieldwork on related questions in the Kerala welfare context has not yet commenced, and the framework proposed here has not been empirically validated against practitioner experience. The empirical literature reviewed in Section 5 is concentrated on child welfare systems in the United States, with more limited representation from the Global South generally and from India specifically; the regulatory analysis addresses India's DPDP Rules directly, but the empirical and ethnographic literature on algorithmic tools in Indian welfare delivery remains thin relative to the North American and European literature, and this asymmetry should be read as a limitation of the available evidence base rather than an assumption that the dynamics described here transfer without modification. Finally, the accountability framework developed in Section 8 requires empirical testing — through practitioner interviews, agency case studies, or pilot implementation — before its practical feasibility, as distinct from its conceptual coherence, can be established. We regard this as a natural next stage of research rather than a deficiency to be resolved within a conceptual paper of this scope.

This paper has moved from a diagnosis to a remedy: from the ways algorithmic mediation disrupts casework, through the three structural gaps that disruption opens in social work's ethical codes, to the three principles we propose to close them. Figure 2 summarises that whole trajectory in one view, including the point of the argument that matters most for professional practice: closing these gaps is not a matter of adding a fourth, AI-specific rule alongside the existing ones. It requires revising what confidentiality, consent, and accountability already mean within the profession's own codes, so that each is anchored to who actually has the capacity to discharge it rather than to who signed the intake form.

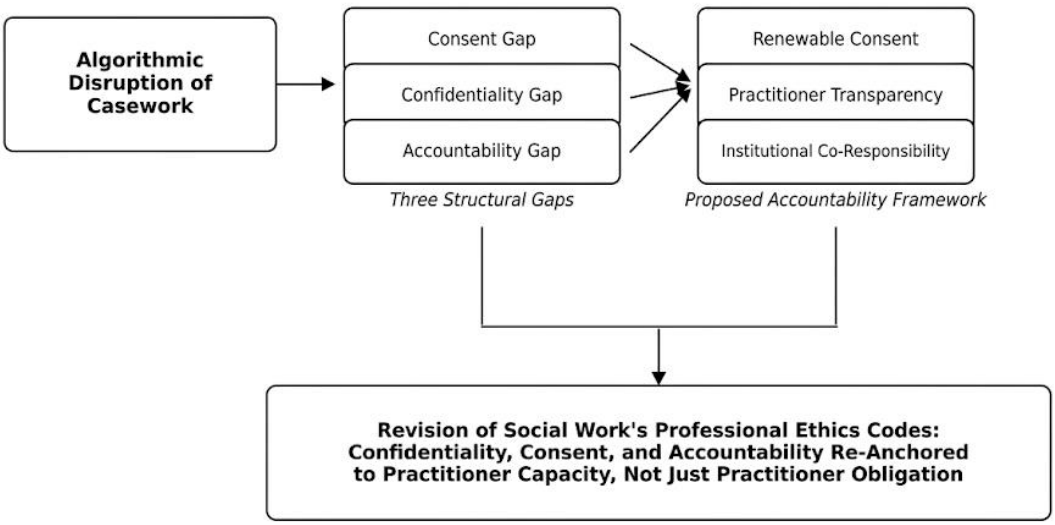

Figure 2. From algorithmic disruption, through the three structural gaps, to the proposed accountability framework and its resulting revision of professional ethics codes.

10. CONCLUSION

Social work's ethical codes were constructed for a professional environment in which a practitioner could observe the full trajectory of a client's information: where it travelled, who accessed it, and to what end it was used. That environment

has substantially changed. What has replaced it is a distributed system in which consent, confidentiality, and accountability have each become detached from the practitioner who remains, on paper, responsible for all three. Reducing algorithmic bias, however genuine an achievement, does not reattach them. Only a framework that follows the data itself — renewable consent, transparency toward the practitioners actually using these tools, and institutional co-responsibility for what the tools decide — offers a plausible route to doing so.

For future research, this analysis points toward several directions: empirical validation of the proposed framework through practitioner and client interviews; comparative study of algorithmic welfare governance across Indian states as DPDP implementation proceeds; and longitudinal tracking of the EU AI Act's Annex III provisions once they take effect in December 2027. For social work education, the implication is that curricula addressing professional ethics should incorporate algorithmic decision-making explicitly, rather than treating it as a specialised technology topic separate from core practice ethics. For policy, professional bodies need not wait for AI-specific legislation to converge across jurisdictions; the accountability structure proposed here can be adopted through existing codes of ethics and continuing-education requirements. The regulatory architecture is assembling, unevenly, across the EU, the OECD framework, and India's new data protection rules. Social work's own professional codes have not yet caught up. This paper has argued why they need to, and has proposed one plausible route by which they might.

REFERENCES

[1]. Bakiner, O. (2023). The promises and challenges of addressing artificial intelligence with human rights. *Big Data & Society, 10*(1). https://doi.org/10.1177/20539517231205476
[2]. Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal, 13*(4), 447–468. https://doi.org/10.1111/j.1468-0386.2007.00378.x
[3]. Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review, 81*(5), 825–836. https://doi.org/10.1111/puar.13293
[4]. Chaudhuri, B. (2022). Programmed welfare: An ethnographic account of algorithmic practices in the public distribution system in India. *New Media & Society, 24*(4), 887–902. https://doi.org/10.1177/14614448221079034
[5]. Cheng, H.-F., Stapleton, L., Kawakami, A., Sivaraman, V., Cheng, Y., Qing, D., Perer, A., Holstein, K., Wu, Z. S., & Zhu, H. (2022). How child welfare workers reduce racial disparities in algorithmic decisions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–22). Association for Computing Machinery.
[6]. Chouldechova, A., Benavides-Prado, D., Fialko, O., & Vaithianathan, R. (2018). A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. *Proceedings of Machine Learning Research, 81*, 134–148.
[7]. Cobbe, J., Lee, M. S. A., & Singh, J. (2021). Reviewable automated decision-making: A framework for accountable algorithmic systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 598–609). https://doi.org/10.1145/3442188.3445921
[8]. Cooper, A. F., Moss, E., Laufer, B., & Nissenbaum, H. (2022). Accountability in an algorithmic society: Relationality, responsibility, and robustness in machine learning. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 864–876). https://doi.org/10.1145/3531146.3533150
[9]. Council of Europe. (2024). *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*.
[10]. Dencik, L., Hintz, A., Redden, J., & Treré, E. (2022). *Data justice*. Sage.
[11]. DLA Piper. (2026, June). *The Digital AI Omnibus: Proposed deferral of high risk AI obligations under the AI Act* [Update]. DLA Piper GENIE.
[12]. Floridi, L. (2013). *The ethics of information*. Oxford University Press.
[13]. Gabriels, A. E., Kuiper, C. H. Z., & Harder, A. T. (2025). Algorithmic tools to help social workers in child protection service make well-informed decisions: A scoping review. *Child Abuse & Neglect, 169*(Pt 2), Article 107678. https://doi.org/10.1016/j.chiabu.2025.107678
[14]. Gibson Dunn. (2026, May 27). *EU AI Act Omnibus agreement — Postponed high-risk deadlines and other key changes* [Client alert].
[15]. International Federation of Social Workers. (2018). *Global social work statement of ethical principles*. IFSW.
[16]. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2
[17]. Kawakami, A., Sivaraman, V., Cheng, H.-F., Stapleton, L., Cheng, Y., Qing, D., Perer, A., Wu, Z. S., Zhu, H., & Holstein, K. (2022). Improving human-AI partnerships in child welfare: Understanding worker practices, challenges, and desires for algorithmic decision support. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–18). Association for Computing Machinery.
[18]. Khera, R. (2017). Impact of Aadhaar on welfare programmes. *Economic and Political Weekly, 52*(50), 61–70.

[19]. Kotsenas, A. L., Balthazar, P., Andrews, D., Geis, J. R., & Cook, T. S. (2021). Rethinking patient consent in the era of artificial intelligence and big data. *Journal of the American College of Radiology, 18*(1, Pt B), 180–184. https://doi.org/10.1016/j.jacr.2020.09.022

[20]. Langer, M., Oster, D., Speith, T., Hermanns, H., Kästner, L., Schmidt, E., Sesing, A., & Baum, K. (2021). What do we want from Explainable Artificial Intelligence (XAI)? A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. *Artificial Intelligence, 296*, Article 103473. https://doi.org/10.1016/j.artint.2021.103473

[21]. Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology, 6*, 175–183. https://doi.org/10.1007/s10676-004-3422-1

[22]. Ministry of Electronics and Information Technology, Government of India. (2025, November 14). *Digital Personal Data Protection Rules, 2025* [Gazette notification]. Press Information Bureau.

[23]. Moulaei, K., Akhlaghpour, S., & Fatehi, F. (2025). Patient consent for the secondary use of health data in artificial intelligence (AI) models: A scoping review. *International Journal of Medical Informatics, 198*, Article 105872. https://doi.org/10.1016/j.ijmedinf.2025.105872

[24]. National Association of Social Workers. (2021). *Code of ethics of the National Association of Social Workers*. NASW Press.

[25]. Nissenbaum, H. (1996). Accountability in a computerized society. *Science and Engineering Ethics, 2*(1), 25–42. https://doi.org/10.1007/BF02639315

[26]. Organisation for Economic Co-operation and Development. (2019). *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449). OECD Publishing.

[27]. Santoni de Sio, F., & Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy & Technology, 34*(4), 1057–1084. https://doi.org/10.1007/s13347-021-00450-x

[28]. Saxena, D., & Guha, S. (2024). Algorithmic harms in child welfare: Uncertainties in practice, organization, and street-level decision-making. *ACM Journal on Responsible Computing, 1*(1), Article 2. https://doi.org/10.1145/3616473

[29]. Saxena, D., Badillo-Urquiola, K., Wisniewski, P. J., & Guha, S. (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), Article 348. https://doi.org/10.1145/3476089

[30]. Tronto, J. C. (1993). *Moral boundaries: A political argument for an ethic of care*. Routledge.

[31]. Wang, Y., Federman, A., Wurtz, H., Manchester, M. A., Morgado, L., Scipion, C. E. A., Adjini, M., Williams, K., Wills, B., Helmly, V., & Arias, J. J. (2025). The evolving literature on the ethics of artificial intelligence for healthcare: A PRISMA scoping review. *Frontiers in Digital Health, 7*, Article 1701419. https://doi.org/10.3389/fdgth.2025.1701419

[32]. Widder, D. G., & Nafus, D. (2023). Dislocated accountabilities in the "AI supply chain": Modularity and developers' notions of responsibility. *Big Data & Society, 10*(1). https://doi.org/10.1177/20539517231177620

[33]. Wieringa, M. (2020). What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 1–18). Association for Computing Machinery. https://doi.org/10.1145/3351095.3372833

[34]. Yu, M.-H., & Rose, R. A. (2026). Algorithmic-assisted decision-making tools in child welfare practice: A systematic review. *Research on Social Work Practice*. Advance online publication. https://doi.org/10.1177/10497315251350933