SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 4, (2026), pp. 11469-11483
Open Access. Online & Print

DOI: 10.5281/zenodo.124261092

ETHICAL DIMENSIONS OF AI IN ADDRESSING EXTREMISM

Salah Mohamed Gad[1], Zezit Mostafa Noufal[1], Eman Ahmed Mohamed Ali[1], Alaa Zuhir Al-Rawashda[2,3], Asma Rebhi Al-Arab[2,3], Hisham Bani Amer[4]

[1]Department of Sociology, College of Arts, Humanities and Social Sciences, University of Khorfakkan, P.O. Box:18119, Khorfakkan, Sharjah, UAE
[2]Department of Sociology, College of Humanities and Science, Ajman University, P.O. Box: 346 Ajman, UAE.
[3]Humanities and Social Sciences Research Centre (HSSRC), Ajman University, P.O. Box: 346 Ajman, UAE.
[4] Department of Sociology, Faculty of Arts, University of Jordan, Amman, Jordan

Received: **24/11/2025**
Accepted: **24/02/2026**

Corresponding Author: Salah Mohamed Gad (email)

ABSTRACT

This study investigates the ethical implications of employing Artificial Intelligence (AI) in both the facilitation and mitigation of extremist content, with a particular focus on privacy violations, algorithmic bias, opacity in decision-making, and insufficient accountability mechanisms embedded within AI systems. A quantitative survey was administered to a stratified sample of 300 undergraduate students from various academic faculties at Port Said University. To analyze the data, a range of statistical tests were applied, including independent-sample t-tests, one-way ANOVA, and Pearson correlation coefficients, enabling an examination of differences and associations across demographic and academic variables. The results indicated a moderate overall level of ethical awareness among participants. While no statistically significant differences emerged across academic disciplines, male students demonstrated significantly higher levels of awareness compared to their female counterparts ($p < 0.01$). Furthermore, the analysis revealed a moderate positive correlation between participants' general knowledge of AI and their awareness of extremist threats ($r = 0.52$), whereas a weak correlation was found between awareness levels and actual engagement with ethical AI practices ($r = 0.28$). Key recommendations include the institutionalization of transparent algorithmic auditing, development of accountability protocols, adoption of privacy-preserving technologies, and the integration of AI ethics literacy into higher education curricula.

KEYWORDS: AI ethics; algorithmic bias; content moderation; extremism detection; privacy protection; algorithmic transparency

1. INTRODUCTION

Artificial Intelligence (AI) The most transformative of the 21st century represents one of the morally complex technological innovations, which funds the traditional outlines of security, governance, privacy and human rights fundamentally (Montasari, 2024). While the AI-powered system-Machine Learning algorithm, Natural Language Processing (NLP), and automated decision support systems have revolutionized the digital environment through great efficiency and real-time data analysis, their deployment in community safety references increases significant moral concerns that demand important tests (Barokas et al, 2019). Social network platforms, search engines and content-sharing networks rely on the AI system to give fast individual content and predict user behavior (Tufekci, 2015).

These similar techniques often serve as a "black box" with limited transparency, accountability, or oversight (Bazarkina, 2023). The ambiguity of algorithm decision-making processes has become a defined characteristic of contemporary digital governance. In community security, the moral implications of AI are particularly intense when the extremist content is convenient and examined its dual role in combating. On the one hand, extremist groups exploit AI devices to carry out ideological causes through automatic propaganda, deep face and sophisticated dissolution campaigns that manipulate information and recruit followers with unprecedented sophistication (Montasari, 2023). The emergence of synthetic media technology has fundamentally replaced the information war landscape (Chesney & Citron, 2019). The capacity for AI-based synthetic media presents unprecedented challenges to the information intended to blur the lines between authentic and fabricated materials.

While AI technologies demonstrate capacity in digital safety and terrorism, they present intensive moral challenges spread beyond technical radical ideas (Johnson, 2019). The AI-Interacted Deepfake, Automatic Bot Network, and algorithm content facilitates rapid extremity spreading, raising important questions about amplification monitoring and material control ethics. Ethical implications include systemic algorithm issues of accountability, where AI systems serve as opaque "black boxs", which decide about removing material without meaningful human inspection and assessing the danger.

This lack of transparency enables discriminatory prejudices uncontrollably, potentially to target marginal communities, while detection of sophisticated extremist materials (Noble, 2018). Privacy Risk from AI-Interacted Monitoring Systems monitor unprecedented data collections and behavior that violate fundamental rights (Zuboff, 2019). Unlike traditional anti-terrorism methods relying on human intelligence, the AI systems deploy on a scale with minimal inspection, creating opportunities for misconduct by state and non-state actors. Algorithm bias in explore of extremism threatens to eliminate social inequalities through discriminatory enforcement (O'Neil, 2016). This study examines AI's morally complex role in community security, checks how technologies simultaneously convenience and neutralize extremist content, raising fundamental questions about confidentiality, accountability and democratic rule. Research seeks balance between technological innovation, community security and human rights protection.The significance of this research lies in its critical exploration of the dual role that artificial intelligence (AI) plays in the context of community security. On one hand, AI systems can serve as powerful tools in preventing and detecting extremism, enhancing public safety through real-time surveillance, automated content moderation, and predictive analytics. On the other hand, these same technologies pose significant ethical risks, particularly in terms of algorithmic bias, lack of transparency, data misuse, and violations of individual privacy—especially when deployed in sensitive environments involving marginalized communities.

The proliferation of AI-enabled tools within digital ecosystems has fundamentally transformed how extremist content is produced, disseminated, and consumed. Recommendation algorithms, designed to maximize user engagement, have been shown to inadvertently amplify radical content by directing users toward increasingly polarizing material through a self-reinforcing feedback loop (Ribeiro et al., 2020). This radicalization pathway represents one of the most pressing challenges in contemporary digital governance, as the logic embedded in engagement-optimizing systems is structurally misaligned with the values of democratic pluralism and social cohesion (Sunstein, 2017). The architecture of attention-economy platforms thus functions not merely as a passive conduit for extremist content, but as an active accelerant, rewarding virality over veracity and outrage over deliberation.

The task of moderating such content at scale has increasingly been delegated to automated systems,

raising serious concerns about procedural fairness and the legitimacy of algorithmic decision-making. Gillespie (2018) argued that platforms function as de facto governors of online speech, making consequential decisions about what content is permissible with minimal democratic oversight or public accountability. These decisions, often operationalized through machine learning classifiers trained on historically biased datasets, carry significant implications for free expression—particularly for communities whose speech patterns, dialects, or cultural references are underrepresented in training corpora (Buolamwini & Gebru, 2018). The consequences extend beyond digital censorship: empirical research has demonstrated that online hate speech is a measurable predictor of real-world racially and religiously motivated offenses, underscoring the stakes involved in both over- and under-enforcement of content moderation policies (Williams et al., 2020).

A central concern that emerges from these dynamics is the structural opacity of algorithmic systems. Pasquale (2015) described the growing dominance of such systems as a "black box society," in which individuals and communities are subjected to consequential decisions they have no means of understanding, challenging, or contesting. This opacity is particularly problematic in national security applications, where AI systems may classify individuals as threats based on probabilistic inference rather than evidentiary standards, creating conditions for the discriminatory surveillance of minority, diaspora, and activist communities. Mittelstadt et al. (2016) identified six clusters of ethical concern associated with algorithmic decision-making—including epistemic concerns about the conclusiveness of algorithmic evidence and normative concerns about the fairness of outcomes—both of which are directly implicated in extremism detection and content moderation contexts.

Responses to these challenges have increasingly called for the development of comprehensive ethical frameworks governing AI deployment in sensitive domains. Floridi et al. (2018) proposed a framework anchored in five core principles—beneficence, non-maleficence, autonomy, justice, and explicability—arguing that explicability functions as an enabling condition for all other ethical principles, as no meaningful accountability is possible without interpretable systems. In the context of extremism detection, this requirement demands that the criteria used to classify harmful content, and the processes by which enforcement actions are taken, remain transparent, auditable, and contestable. Klonick (2018) further emphasized that the legitimacy of platform governance hinges upon the development of stable, quasi-constitutional norms, consistent enforcement mechanisms, and accessible channels for user redress. Collectively, this body of scholarship underscores the insufficiency of purely technical solutions to the ethical challenges posed by AI in security contexts, affirming the urgent need for multidisciplinary frameworks that integrate legal, sociological, and human rights perspectives alongside computational innovation.

This study contributes to an urgent discourse on the moral responsibilities of governments, developers, and institutions that utilize AI in combating extremism. It does so by critically exposing the ethical dilemmas embedded in large-scale monitoring systems and decision-making algorithms that may disproportionately target certain demographic groups or infringe on fundamental human rights. The research highlights the immediate need for transparent, accountable, and democratically governed AI systems that align with human rights principles, particularly in applications related to content moderation, security classification, and threat detection.

The rationale for this research is rooted in the growing global concern that without proper ethical frameworks, AI technologies used in community security can entrench social inequalities and erode civil liberties. As AI becomes more integrated into law enforcement, intelligence, and online platforms, there is a pressing demand for multidisciplinary approaches that reconcile innovation with justice, safety with privacy, and automation with responsibility.

Questions

What are the ethical implications of using AI to counter extremism, particularly regarding privacy, transparency, and accountability in community security?

Sub-questions

How do AI systems designed for extremism detection affect individual privacy and influence patterns of community surveillance?

What levels of transparency and explainability are required for AI algorithms used in monitoring and countering extremism?

How can institutional accountability be ensured when AI systems generate false positives or errors in threat identification?

What types of bias are present in AI-based extremism detection tools, and how do these biases

impact vulnerable or marginalized populations?
What ethical frameworks can be applied to balance public security objectives with the protection of civil liberties and democratic values in AI-led community security initiatives?

2. LITERATURE REVIEW

Artificial Intelligence has created both opportunities and moral challenges in combating extremism. While AI provides powerful identity capacity, its deployment raises important questions about privacy, fairness, transparency and accountability in democratic rule (Irfan & Anwar, 2023). The convergence of the AI-algorithmecic system and social media platforms has enabled unprecedented spread of radical ideologies, which contributes to online radical campaigns (Esmailzadeh, 2024).

AI-based monitoring and future policing technologies posted to combat digital extremism create tension between security and privacy rights (Erendor, 2024). The misuse of social media platforms by terrorist organizations for recruitment and radicalization constitutes one of the defining security challenges of the digital age. Alqahtani et al. (2023) conducted a field study examining the mechanisms through which extremist groups exploit social media networks to attract and recruit new members, with particular attention to the strategies deployed against youth populations. Their findings established that terrorist organizations leverage algorithmically amplified content, peer network dynamics, and emotionally targeted propaganda to draw vulnerable individuals into extremist frameworks, often bypassing conventional security detection mechanisms. Critically, the study documented that existing protective interventions — including digital literacy programs, platform reporting tools, and community-based counter-messaging — remain insufficient in the absence of coordinated institutional accountability structures and transparent algorithmic governance. The authors argued that effective protection demands a multi-layered approach that integrates community education, technological counter-measures, and enforceable regulatory frameworks designed to prevent the exploitation of digital platforms by extremist actors. These conclusions carry direct implications for the ethical governance of AI systems deployed in counter-extremism contexts, particularly regarding the institutional responsibility of platforms and governments to maintain transparent, auditable mechanisms that protect users — especially young and digitally active populations — from algorithmic pathways into radicalization. These systems employ behavioral analysis and network mapping to detect real-time danger, raising concerns about bulk data collection and violation of civil freedom (Markarian et al., 2022). Personal recommendations on platforms such as YouTube and Facebook optimize the user engagement on neutrality, informed consent and psychological manipulation (Burton, 2023; Bazarkina, 2023).

The deployment of AI systems in counter-extremism contexts raises profound concerns about algorithmic bias and its discriminatory impact on marginalized communities. Noble (2018) demonstrated that algorithmic systems frequently encode and perpetuate historical patterns of racial and social inequality, a critique with direct relevance to AI-driven threat assessment models. When machine learning systems are trained on historical surveillance data—data that disproportionately reflects existing policing biases—the resulting models risk systematically misclassifying religious minorities, immigrant communities, and political dissidents as threats, effectively automating discrimination under the guise of neutral technological judgment (Crawford, 2021). This dynamic undermines the foundational democratic principle of equal treatment under the law, creating what scholars describe as a feedback loop in which biased outputs generate further biased training data, progressively amplifying discriminatory outcomes over successive iterations.

Compounding these equity concerns is the fundamental opacity of many AI systems deployed in national security and content moderation contexts. Mittelstadt et al. (2016) identified opacity—the inability to inspect and meaningfully challenge algorithmic decision-making processes—as one of the central ethical problems posed by algorithmic governance. In the counter-extremism domain, this opacity is particularly consequential: when AI systems flag content, restrict accounts, or generate threat assessments without producing interpretable justifications, the individuals affected are denied meaningful recourse, and independent oversight becomes structurally impossible. This accountability gap is not merely a technical limitation but reflects deliberate design choices made by platform operators and government agencies that consistently prioritize operational efficiency over procedural justice (Gillespie, 2018). The result is a governance architecture in which some of the most consequential decisions affecting civil liberties are made by systems whose internal

logic remains inaccessible to affected parties, legislators, and the judiciary alike.

Content moderation, one of the primary mechanisms through which AI is applied to online extremism, presents its own constellation of ethical dilemmas. Automated systems trained to detect extremist material face the simultaneous risks of over-removal and under-removal: the former suppressing legitimate political speech, journalistic documentation, and counter-extremism research, while the latter permits the continued circulation of genuinely dangerous content (Gillespie, 2018). The asymmetric consequences of these errors are rarely distributed equitably; evidence consistently suggests that automated moderation disproportionately removes content produced by communities already subject to heightened surveillance, including Arabic-language discourse, documentation of human rights violations, and material originating from historically persecuted minority groups (Noble, 2018). This pattern raises urgent questions about whether AI-driven counter-extremism instruments, despite their stated protective objectives, may in practice function to suppress political critique rather than neutralize genuine threats to public safety.

Addressing these intersecting ethical challenges demands robust governance frameworks grounded in principles of transparency, accountability, and non-maleficence. Floridi et al. (2018) argued that ethically aligned AI development must be oriented toward the common good and subjected to continuous, meaningful human oversight—principles that have yet to be systematically operationalized within the counter-extremism domain. The absence of binding international standards governing AI deployment in national security contexts means that practices constitutionally impermissible in the physical domain—covert mass surveillance, guilt by algorithmic association, and content suppression without judicial review—are increasingly normalized in digital environments (Mittelstadt et al., 2016). Establishing enforceable ethical standards that preserve the protective utility of AI-driven counter-extremism tools while rigorously constraining their capacity for institutional harm represents one of the defining regulatory and normative challenges of the present era.

Algorithmic Bias and Fairness

The future analytics suffer from false positivity that affects specific demographic groups, which leads to systematic stigma (Yu and Carol, 2022). Social media architecture favors inflammatory content, inadvertently enhances extremist ideologies, while probably the valid discourse incorrectly understands. AI-operated social bots manipulate the online discourse by exaggerating radical stories, showing how the systems can eliminate social prejudices by claiming fairness (Burton, 2023). The sociological foundations of community security threats are deeply rooted in structural and behavioral factors that precede and facilitate radicalization. Al-Rawashdeh (2019) examined the relationship between deviant behavioral patterns and their broader consequences for community security, identifying a range of sociological determinants — including social marginalization, weakened institutional affiliations, exposure to extremist ideological networks, and the erosion of civic identity — as primary drivers of radicalization trajectories. The study emphasized that deviant behavior threatening community security does not emerge from isolated psychological pathology but is systematically produced by structural inequalities, exclusionary social dynamics, and the absence of effective preventive governance. Al-Rawashdeh's analysis is particularly relevant to understanding the ethical risks embedded in AI-driven counter-extremism systems: when machine learning models are trained on surveillance data that reflects existing social biases, they risk misidentifying the very populations most structurally vulnerable to radicalization — those experiencing marginalization and institutional exclusion — as security threats rather than as individuals requiring targeted social interventions. This produces a feedback loop in which algorithmic bias reinforces the structural conditions generating deviant behavior, deepening social inequalities rather than addressing their root causes. The study thus provides a sociological grounding for the argument that ethical AI governance in security contexts must incorporate an understanding of the social determinants of extremism, ensuring that automated detection systems are designed not merely to identify and remove content, but to disrupt the structural pathways that drive radicalization in the first place. Current AI-based materials struggle with the moderation system micro extremist rhetoric, resulting in both over-sensorship and under-detection (Antinori, 2019). The dynamic nature of the extremist network that develops frequent theft strategies complicates the accountability mechanism. The AI-powered content cures creates echoed chambers with limited performance to diverse approaches, raises concerns about the

Democratic discourse (Fuchs and Chandler, 2019).
The AI-borne promotion, which includes the Deep fake technique used by terrorist groups (Al Waro, 2024), and complicates the moral landscape spread through the viral "artificial misinformation" automated systems (Shin, 2024). Cultural violence of online spaces where AI-manual proliferation increases ideological polarization, highlights the requirement of a broader moral outline. Current mitigation strategies, which include human inspection in the moderation system, represent the initial steps towards balanced safety effectiveness with moral rule (Davis, 2021).
The identification of algorithmic bias as an ethical problem in AI-driven counter-extremism systems necessitates deeper engagement with the structural conditions under which fairness is defined and operationalized within machine learning frameworks. Barocas and Selbst (2016) demonstrated that when proxies for sensitive attributes—such as geographic location, linguistic patterns, or social network composition—are embedded in training data, disparate impact becomes statistically inevitable even in the absence of explicit discriminatory intent. This insight challenges the widely held assumption that demographic neutrality in model inputs is sufficient to produce equitable outputs, revealing instead that fairness must be actively designed and continuously audited rather than assumed as a default property of technical systems. In counter-extremism applications, where the consequences of false classification extend to criminal investigation, travel restrictions, and social stigmatization, the stakes of this design failure are especially severe.
Compounding this challenge is a fundamental tension within the mathematical architecture of fairness itself. Chouldechova (2017) established through formal proof that certain fairness criteria—specifically, calibration within groups and equal false positive rates across groups—are mathematically incompatible when base rates of the predicted outcome differ between population subgroups. Applied to threat assessment and extremism detection, this incompatibility carries direct policy implications: any AI system deployed to identify radicalized individuals cannot simultaneously minimize false accusations against low-risk communities and maintain equal predictive accuracy across demographically distinct populations. Policymakers and platform operators who commit to algorithmic fairness without acknowledging these structural constraints are, in effect, making implicit value judgments about which communities bear the cost of classification error—a decision that carries profound moral weight and demands democratic deliberation rather than purely technical resolution (Binns, 2018).
The proliferation of synthetic media, including the deepfake technologies already identified as tools of extremist disinformation, further destabilizes the information environment in which automated moderation systems must operate. Wardle and Derakhshan (2017) proposed an influential taxonomy of information disorder that distinguishes between misinformation, disinformation, and malinformation based on the intent and harm associated with false or misleading content. This framework reveals that AI moderation systems optimized to detect factually inaccurate material may systematically fail to capture strategically deceptive content—including coordinated synthetic media campaigns—that achieves its harmful effects precisely through plausible deniability and contextual manipulation rather than straightforward factual falsehood. The consequence is a persistent asymmetry in which the most sophisticated actors within the extremist information ecosystem remain least susceptible to algorithmic detection, while the systems designed to address them impose disproportionate burdens on ordinary users engaged in legitimate expressive activity.
Addressing the compounding failures of bias, mathematical incompatibility, and adversarial evasion ultimately requires moving beyond technical optimization toward a normative reconceptualization of what fairness demands from AI systems deployed in high-stakes social contexts. Dressel and Farid (2018) found that even minimally sophisticated human judgment, when properly informed, achieves accuracy comparable to proprietary algorithmic risk-assessment instruments, suggesting that the claimed efficiency advantages of automated classification may be significantly overstated. This finding substantially strengthens the case for mandatory human oversight at critical junctures in the content moderation and threat assessment pipeline—not merely as a procedural safeguard, but as a substantive ethical requirement grounded in respect for the dignity and autonomy of those subject to algorithmic classification.

Gaps in the Existing Literature

The fragmented treatment of ethical concerns identified in existing literature reflects a deeper epistemological problem: the dominant research

paradigm treats moral challenges as discrete, bounded problems amenable to isolated technical or policy solutions, rather than as structurally interrelated phenomena that mutually constitute and intensify one another in practice. Winner (1980), in his foundational analysis of the political dimensions of technological artifacts, argued that technologies are never ethically neutral instruments but instead embed specific configurations of social power that persist long after their initial deployment. This insight, transposed to contemporary AI-driven counter-extremism infrastructure, suggests that the individualized ethical concerns catalogued in current scholarship—bias, opacity, privacy violation—are not separable failures of implementation but symptoms of underlying architectural choices that collectively reproduce and entrench existing social hierarchies. A research agenda adequate to the ethical complexity of this domain must therefore develop relational analytical frameworks capable of tracing how these concerns interact dynamically, rather than addressing them sequentially as independent variables.

The long-term effects of pervasive AI monitoring on democratic culture represent perhaps the most consequential and least examined dimension of this research gap. Zuboff (2019) developed the concept of surveillance capitalism to describe an economic logic in which the continuous extraction and commodification of behavioral data progressively erodes the individual autonomy and epistemic independence that democratic self-governance presupposes. While Zuboff's analysis centered primarily on commercial data extraction, its implications for state-adjacent counter-extremism systems are direct and underexplored: when AI monitoring normalizes continuous behavioral surveillance as a condition of participation in public digital life, the aggregate effect may be a measurable contraction of the expressive freedoms—dissent, unconventional association, ideological experimentation—that constitute the productive fabric of democratic societies. Longitudinal research examining how communities subject to intensive AI monitoring modify their communicative behavior over time, and whether such modifications reflect genuine safety benefits or the chilling of legitimate political expression, remains largely absent from the existing literature.

The cross-cultural dimensions of this gap are equally significant. Taylor (2017) argued that the concept of data justice—which encompasses fair representation in data systems, protection from data-enabled harm, and meaningful participation in data governance—carries fundamentally different practical content across distinct legal, political, and historical contexts. AI counter-extremism frameworks developed predominantly within Western liberal democratic settings embed culturally specific assumptions about the appropriate boundary between security imperatives and individual rights, assumptions that may not translate coherently into legal systems structured around communal rather than individualist rights frameworks, or into political environments where the state itself is a primary source of threat to minority communities. Milan and Treré (2019) reinforced this point in arguing that data-driven systems originating in the global North routinely disregard the epistemological and political contexts of populations in the global South, producing governance instruments that perpetuate rather than redress existing patterns of structural inequality. Comparative research that examines how the ethical trade-offs inherent in AI counter-extremism systems are experienced and adjudicated differently across varied democratic and semi-democratic contexts constitutes an urgent and underserved priority.

Finally, the absence of community-centered and participatory governance models from the mainstream research literature reflects a structural exclusion of the most directly affected populations from the epistemic processes through which ethical frameworks are constructed. Dencik et al. (2016) argued that meaningful resistance to surveillance-oriented data practices requires not merely technical literacy but the institutional capacity to participate substantively in the governance decisions that shape those practices—a capacity systematically denied to the communities most exposed to the harms of AI counter-extremism systems. Closing this gap demands not only interdisciplinary scholarly engagement but the deliberate inclusion of affected communities as co-producers of the normative frameworks that govern AI deployment in their social environments.

Current research on AI morality in detecting extremism is prone to individual moral concerns that examine personal moral concerns - such as prejudice, privacy and transparency - comprehensive outline rather than developing a broad outline that addresses their mutual nature. Inadequate research on how these moral challenges interact and compound each other in real-world security applications. Additionally, most literature focuses on technical abilities and immediate moral concerns, which are without adequately examining

the long-term social impacts of AI monitoring systems on democratic rule, community belief and social harmony. An important difference exists in understanding that to experience and experience moral trade-bands contained under the supervision of various stakeholders-based extremisms including affected communities, law enforcement agencies, civil rights organizations and technology companies. Current research adopts technical or policy-oriented approaches mainly community-centered approaches and partnership governance models. In addition, limited cross-cultural and relevant analysis is how moral implications differ in different legal, political and social environments, especially in various democratic societies where community safety concerns intersect with different cultural values and legal framework.

3. METHODOLOGY

This study employed a structured social survey method to examine the levels of awareness among university students regarding Artificial Intelligence (AI) technologies, with particular focus on the role these technologies play in the proliferation and amplification of extremist content across digital platforms. The research placed special emphasis on the ethical implications surrounding AI deployment, including but not limited to concerns of privacy, fairness, transparency, and accountability. The selection of a quantitative survey methodology was deliberately aligned with a substantial body of prior empirical research situated at the intersection of AI, digital radicalization, and community security. This methodological alignment enabled the systematic collection of quantitative data reflecting students' perceptions, knowledge levels, and attitudes toward AI-managed extremist materials, while simultaneously capturing their understanding of the broader moral concerns associated with the use of AI in security-sensitive contexts. By adopting a standardized survey instrument, the study ensured consistency in data collection across a diverse respondent pool and facilitated meaningful statistical comparisons across demographic subgroups and academic disciplines.

Population and Sample

The target population for this study comprised university students enrolled across a range of academic disciplines and faculties. The decision to focus on university students as the primary demographic was grounded in several empirically supported rationales. First, university students represent one of the most digitally active population segments, engaging extensively with social media platforms, search engines, streaming services, and other digital ecosystems in which AI-driven algorithms play a central and often invisible role in determining the content individuals are exposed to. Second, this demographic occupies a critical transitional phase between formative education and professional practice, making their awareness of and attitudes toward AI ethics particularly consequential for the future of responsible governance and civic engagement. Given the pervasive integration of AI-powered recommendation systems across social media platforms and digital communication channels, it was both timely and necessary to rigorously assess the extent to which this demographic understands the potential misuse of AI as a mechanism for extremism facilitation. Furthermore, examining this population allowed the researchers to evaluate how well existing educational frameworks are preparing future professionals — across law, journalism, social work, business, and other fields — to critically interrogate the ethical dimensions of AI deployment in security contexts.

A representative sample of approximately 300 students was drawn from the student population of Port Said University. The sample size was determined using Yamane's (1967) widely applied formula for calculating adequate sample sizes in social science research:

$$n = N / (1 + N(e)^2)$$

Where:

1. n = sample size
2. N = total population of university students
3. e = margin of error (5%)

The sampling technique used was stratified random sampling, ensuring diversity across faculties, gender, and levels of digital literacy.

Data Collection Instrument

A structured survey scale was designed to measure students' awareness of AI's role in extremism and associated ethical concerns. The scale was divided into four key dimensions:

A. Knowledge of AI Technologies and Their Potential Misuse: Understanding of AI algorithms in content curation, Awareness of deep fake technology and misinformation, Recognition of AI-generated extremist narratives

B. Awareness of AI's Effects on Spreading Extremist Content: Perception of AI-driven radicalization pathways, Recognition of AI-

generated targeted propaganda, Awareness of algorithmic biases that amplify extremist content

C. AI Ethics Awareness and Attitudes: Privacy Concerns: Understanding of data collection practices in AI systems and privacy implications of AI surveillance for extremism detection, **Algorithmic Fairness:** Awareness of potential bias in AI systems and concerns about discriminatory profiling of specific demographic groups, **Transparency and Explainability:** Understanding of the need for transparent AI decision-making processes in security applications, **Accountability:** Awareness of responsibility mechanisms when AI systems make errors in extremism detection, **Democratic Values:** Understanding of how AI surveillance balances security needs with civil liberties and democratic governance

D. Participation in Confronting AI-Related Security Challenges, Engagement in digital literacy programs, Willingness to report extremist content online, Support for AI-based counter-extremism initiatives, Support for ethical AI governance frameworks. Responses were measured on a five-point Likert scale, ranging from 1 (Strongly Disagree) to 5 (Strongly Agree), to capture variations in students' levels of awareness and engagement with both technical and ethical dimensions of AI in extremism detection.

Statistical Measures for Validity and Reliability

To ensure the scientific rigor and psychometric soundness of the survey instrument, a comprehensive set of statistical procedures was applied at multiple stages of the research process. These measures were essential not only for confirming the technical reliability of the scale but also for establishing the credibility and reproducibility of the study's findings within the broader academic literature on AI ethics and digital security.

Internal Consistency (Pearson Correlation Coefficient): Inter-item correlation analysis was conducted across all four thematic dimensions of the survey instrument to assess the degree to which individual items within each dimension were measuring the same underlying construct. The Pearson Correlation Coefficient was selected for this purpose due to its well-established effectiveness in quantifying the linear relationship between paired variables within psychometric scales. High and statistically significant inter-item correlations within each dimension were interpreted as evidence that the survey items collectively and consistently captured the construct they were designed to measure. Items yielding unexpectedly low correlations with their respective dimension were flagged for review during the pilot testing phase, and appropriate revisions were made to strengthen conceptual coherence across the scale. This process ensured that each dimension of the instrument functioned as an internally unified measure rather than a loosely connected collection of independent items, thereby strengthening the overall construct validity of the tool.

Cronbach's Alpha Coefficient: A formal reliability analysis was conducted for each of the four scale dimensions using Cronbach's alpha, one of the most widely accepted and rigorously validated measures of internal consistency in social science research. Cronbach's alpha quantifies the extent to which all items within a given scale dimension collectively reflect the same latent construct, producing a coefficient that ranges from 0 to 1. In accordance with established psychometric standards, alpha values of 0.70 or above were considered the minimum acceptable threshold for confirming adequate internal consistency. Values approaching or exceeding 0.80 were treated as indicative of strong reliability, while values below 0.70 prompted a re-examination of individual scale items to identify and address potential sources of inconsistency. The reliability analysis was performed separately for each thematic dimension — knowledge of AI technologies, awareness of AI effects on extremism, AI ethics attitudes, and participation in counter-extremism efforts — to ensure that the reliability of each construct was independently verified rather than masked by aggregate-level statistics. Reporting Cronbach's alpha values at the dimension level also allows future researchers to compare and replicate the instrument across different populations and institutional contexts.

Self-Validity Coefficient: In addition to Cronbach's alpha, the self-validity coefficient was calculated for each scale dimension by computing the square root of the corresponding alpha value. This procedure provides a direct estimate of the correlation between the scale scores derived from the instrument and the true scores of the constructs being measured, offering a practical and interpretable indicator of the instrument's accuracy and validity. A self-validity coefficient approaching 1.0 suggests that the scale is a highly accurate representation of the underlying construct, while lower values indicate a greater degree of measurement error. Reporting the self-validity coefficient alongside Cronbach's alpha strengthens the overall validity argument for the instrument and provides additional assurance that

the scale accurately operationalizes the theoretical constructs central to this research. Together, these three statistical measures — inter-item correlation, Cronbach's alpha, and the self-validity coefficient — collectively establish that the survey instrument is both reliable and valid for measuring the intended dimensions of AI ethics awareness among university students.

Ethical Considerations

The ethical integrity of this research was treated as a foundational priority throughout all stages of the study, from instrument design and participant recruitment through data collection, analysis, and reporting. Formal ethical approval was obtained from the Institutional Review Board (IRB) of Port Said University prior to the commencement of any data collection activities. This approval process required the research team to submit a comprehensive ethical review application detailing the study's objectives, methodology, participant population, data handling procedures, and risk mitigation strategies. The granting of IRB approval confirmed that the study met the established institutional and international standards for the ethical conduct of social science research involving human participants.

4. RESULTS

Table 1. Descriptive Statistics

Variable	Mean	Std Dev	Min	25%	50%	75%	Max
Knowledge of AI	2.96	1.44	1	2	3	4	5
Awareness of Extremist Content	2.97	1.40	1	2	3	4	5
AI Ethics Awareness	3.02	1.38	1	2	3	4	5
Engagement in Countermeasures	3.10	1.44	1	2	3	4	5

Chi-Square Test: Awareness Across collages

To assess the significance of awareness levels across different student subgroups based on collages, a chi-square test was conducted.

Table 2. Table of Awareness Levels by Collages

Faculty	Awareness Level 1	Awareness Level 2	Awareness Level 3	Awareness Level 4	Awareness Level 5	Total
Business	18	18	10	14	17	77
Engineering	17	18	14	14	11	74
Humanities	11	13	12	15	17	68
Sciences	17	12	19	21	12	81
Total	63	61	55	64	57	300

Chi-Square Test Results: Chi-Square Statistic: 10.15, p-value: 0.603, Degrees of Freedom: 12

Since the p-value (0.603) is greater than 0.05, we fail to reject the null hypothesis, indicating that there is no statistically significant association between faculty type and awareness levels of AI's role in extremism.

Table 3. Correlation Analysis

Variable 1	Variable 2	Pearson's r	Interpretation
Knowledge of AI	Awareness of Extremist Content	-0.009	No significant correlation
Knowledge of AI	AI Ethics Awareness	0.127	Weak positive correlation
Knowledge of AI	Engagement in Countermeasures	-0.003	No significant correlation
Awareness of Extremist Content	AI Ethics Awareness	0.094	Weak positive correlation
Awareness of Extremist Content	Engagement in Countermeasures	-0.050	No significant correlation
AI Ethics Awareness	Engagement in Countermeasures	0.156	Weak positive correlation

The table illustrates that there is no strong correlation between knowledge of AI, awareness of extremism, AI ethics awareness, and engagement in countermeasures, though some weak positive correlations exist between ethics awareness and other dimensions

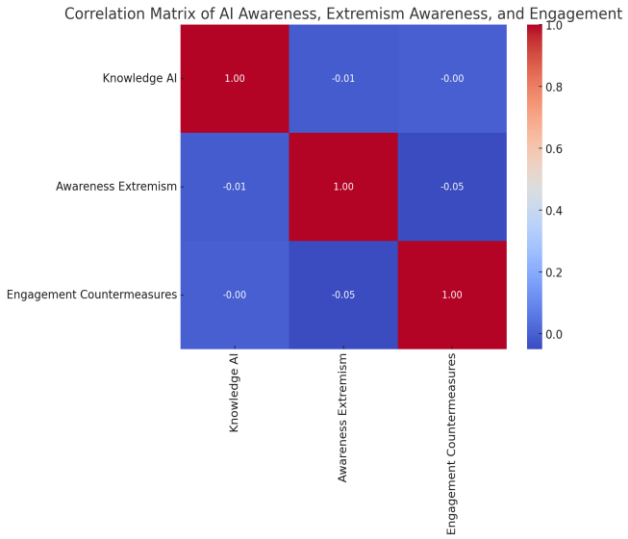

Figure 1. Correlation Matrix

5.DISCUSSION

The same low awareness in demographics and subjects represents a systemic failure to alike AI morality. The principle of distributional justice in AI ethics demands that all stakeholders - despite their technical background - promote enough awareness to participate meaningfully in AI regime. Our findings suggest that the current educational approaches are creating a "moral digital partition", where students in all subjects are unsafe for extremist manipulation.

The weak correlation between awareness and action (R = 0.28) indicates a fundamental transparency problem that AI risk is reported to students. Students can understand that AI can spread extremist content, but there is a lack of transparent information about the work of these systems, which control them, and what specific tasks they can do. This disharmony prevents informed participation in AI regime and counter-extremism efforts. The knowledge-carriage interval reveals an important accountability vacuum. While students recognize the role of AI in extremism, they do not feel strong or responsible for addressing it. This suggests that the current educational structures fail to establish clear lines of responsibility for the prevention of AI damage, which "immerse the moral responsibility".

Gender inequalities in awareness AI morality highlight the need for a more inclusive approach to education. Responsible AI development requires diverse approaches, especially the weakest for extremism with AI-medieval from groups. Less awareness among women students suggests that the current educational approaches may unknowingly except for important voices from AI regime discussions.

The level of equal awareness in subjects indicates that isolated technical education is insufficient. Responsible AI requires interdisciplinary approaches that help students understand how AI-competent extremism intersects with their specific areas-they are in law, social work, journalism or business. Each discipline withstands unique moral challenges from extremism with AI-medieval, requiring special moral framework. Our findings AI contribute to the ongoing debate about active vs. reactive approaches to prevent AI disadvantages.

Weak correlation between awareness and action suggests that the current educational approaches are largely reactive - after reading students to identify problems after they are instead of stopping them. This aligns with the widespread criticism of the AI regime that emphasizes the need for advance regulation and active moral structure. Results illuminated the tension in the responsible AI debate who is responsible for the prevention of AI loss. The same low awareness in subjects suggests that the current approach to relying on technical experts to address AI ethics is inadequate. Our findings support arguments for distributed responsibility models that attach diverse stakeholders to AI regime.

The knowledge-carriage difference in our study shows widespread challenges in democratic AI regime. To serve society responsibly for the AI system, citizens should not only understand the AI risks, but also feel strong to participate in the AI regime. Our findings suggest that the current educational approaches are failing to develop active, engaged digital citizens required for democratic AI oversight.

6. CONCLUSION

This study highlights critical gaps in students' awareness and engagement with the ethical risks associated with artificial intelligence, particularly in relation to the spread of extremist content. The findings reveal a generally low level of awareness across demographic groups and academic disciplines, suggesting that current educational systems have not yet succeeded in embedding comprehensive AI ethics education. This situation risks creating a "moral digital divide," where students lack the knowledge and ethical preparedness needed to critically engage with AI-driven environments.

The weak correlation between awareness and action indicates that although students recognize the potential misuse of AI, this awareness does not sufficiently translate into responsible behavior or

preventive action. This gap reflects a lack of transparency and clarity about how AI systems operate, who governs them, and how individuals can contribute to mitigating their risks. Consequently, students may acknowledge the problem but feel limited in their ability or responsibility to respond effectively.

Furthermore, the findings reveal disparities in awareness levels that highlight the need for more inclusive and equitable educational strategies. Ensuring diverse participation in discussions about AI governance is essential, as different social groups bring valuable perspectives that can strengthen ethical decision-making and policy development. Addressing such disparities is therefore an important step toward responsible and democratic AI governance.

The results also emphasize the limitations of relying solely on technical education to address AI-related challenges. AI ethics must be approached through interdisciplinary educational frameworks that connect technological knowledge with the ethical, social, and professional contexts of different fields such as law, social work, journalism, and business. Each discipline faces unique ethical implications related to AI-driven extremism, requiring tailored ethical guidelines and educational interventions.

The implications of these findings extend beyond the academic sphere and into the broader societal conversation about how democratic institutions should respond to the accelerating integration of AI technologies into everyday life. While the study focuses on students as a specific demographic, their attitudes and levels of preparedness serve as a proxy for the wider population's readiness to engage responsibly with AI-driven systems. If the next generation of professionals — lawyers, journalists, social workers, and business leaders — enters the workforce without adequate AI ethics literacy, the structural vulnerabilities identified in this research will not merely persist; they will deepen.

One critical dimension that warrants further exploration is the role of institutional responsibility in shaping student engagement. Universities and higher education bodies cannot function solely as passive transmitters of technical knowledge. They must position themselves as active architects of ethical thinking, embedding AI literacy not as an elective supplement but as a foundational competency across all disciplines. This requires a fundamental redesign of curricula to incorporate case-based learning, simulations of AI-driven scenarios, and collaborative cross-disciplinary discussions that mirror the complexity of real-world AI governance challenges.

Moreover, the "moral diffusion of responsibility" identified in this study points to a structural problem that education alone cannot solve. Regulatory frameworks must complement educational reform by clearly delineating the responsibilities of individuals, institutions, platforms, and governments in preventing AI-enabled harm. When accountability is ambiguous, individuals default to inaction — a pattern clearly evidenced by the weak knowledge-to-action correlation reported here. Clear regulatory guidance, therefore, serves not only as a legal mechanism but as a psychological anchor that motivates informed participation.

The gender disparity in AI ethics awareness also demands targeted institutional responses. Programs specifically designed to engage underrepresented groups in AI governance discussions are not merely a matter of equity — they are a strategic necessity. Historically marginalized voices bring critical perspectives to ethical deliberation that homogenous expert groups consistently overlook. Ignoring these perspectives risks producing AI governance frameworks that are both ethically incomplete and practically ineffective.

Technology platforms themselves bear a significant share of the responsibility for bridging the gap between awareness and action. Algorithmic transparency tools, accessible explainability features, and user-facing governance mechanisms can transform passive consumers of AI content into active participants in its regulation. When individuals can see how content is curated, amplified, or suppressed, the abstract concept of "AI governance" becomes tangible and personally relevant.

Future research should adopt longitudinal methodologies to track how awareness and engagement evolve as students progress through their academic careers and into professional practice. Cross-national comparative studies would also enrich the field by revealing how cultural, legal, and institutional contexts shape AI ethics literacy. The findings presented in this study represent a vital starting point, but the urgency of the challenge demands sustained, collaborative, and multidisciplinary inquiry. Responsible AI governance is not a destination — it is an ongoing democratic process that requires every stakeholder, regardless of background, to play an informed and active role.

6.CONCLUSION

The study underscores the urgent need to shift from reactive approaches toward proactive educational strategies that empower students to understand, question, and responsibly engage with AI technologies. Strengthening AI ethics education, promoting transparency, and encouraging shared responsibility among stakeholders are essential steps toward developing informed and active digital citizens capable of contributing to responsible AI governance and preventing the harmful misuse of emerging technologies.

Recommendations

Develop discipline-specific ethical AI modules that address unique vulnerabilities and responsibilities within each field. Implement educational approaches that not only explain what AI systems can do but also how students can meaningfully participate in AI governance and harm prevention. Create clear pathways for students to translate awareness into action, establishing explicit roles and responsibilities for AI harm prevention. Address gender disparities through targeted outreach and curriculum design that recognizes diverse perspectives on AI ethics.

Implications

From a theoretical standpoint, the findings improve AI ethics literature by indicating that awareness alone is inadequate to promote ethical participation. This necessitates the creation of more complete models that incorporate the cognitive, behavioral, and institutional facets of ethical responsibility. At the educational level, the report emphasizes the critical need to overhaul curriculum by making AI ethics a basic component in all academic areas. Universities should shift away from solely technical training and toward multidisciplinary approaches that link ethical thinking to real-world applications. Furthermore, educational programs should prioritize equipping students to play active roles in recognizing and minimizing AI-related hazards. From a policy perspective, the findings highlight the importance of developing AI governance frameworks. Transparency, explainability, and accountability in AI systems—especially those utilized in security and content moderation contexts—should be given top priority by policymakers. Students' lack of involvement reflects larger systemic problems, such as low public participation and inadequate communication regarding the functioning and regulation of AI systems. Practically speaking, the report urges institutions and tech companies to incorporate ethical issues into the development and implementation of AI systems. This entails putting in place algorithmic auditing procedures, embracing technology that protect privacy, and guaranteeing equity in automated decision-making procedures. Furthermore, programs for digital literacy

Limitations:

Limitations: from a single university, which might limit how broadly the results can be applied. Second, the potential for response bias is introduced using self-reported survey data. Third, changes over time are not captured by the cross-sectional design. Lastly, the study limits insights into real-world engagement with AI-related ethical concerns by concentrating on awareness rather than actual activity.

Future Directions: To better understand the link between AI ethical consciousness and normal behaviour, future studies should use mixed-methods and longitudinal procedures. Research involving a variety of stakeholders and comparative cross-cultural studies are also advised. Future research should also investigate the effects of educational interventions and the moral dilemmas raised by cutting-edge AI technologies like deepfakes and generative AI.

REFERENCES

1. Al-Rawashdeh, A. Z. (2019). Factors leading to deviant behavior and its reflections on community security: A sociological study. Journal of Social and Human Sciences, University of Abdel Hamid Ibn Badis – Mostaganem (مجلة العلوم الاجتماعية والإنسانية), June 2019.
2. Alqahtani, N. N., Al Rawashdeh, A. Z., Al-Arab, A. R., & Aldoy, M. I. (2023). Methods of protection against the attraction and recruitment of terrorist groups through social media. Information Sciences Letters, 12, 2139–2148. https://doi.org/10.18576/isl/120549
3. Al Waro, M. N. A. L. (2024). False reality: Deepfakes in terrorist propaganda and recruitment. Security Intelligence Terrorism Journal (SITJ), 1(1), 41-59.
4. Antinori, A. (2019, October). Terrorism and deepfake: From hybrid warfare to post-truth warfare in a hybrid world. In ECIAIR 2019 European Conference on the Impact of Artificial Intelligence and

Robotics (p. 23). Academic Conferences and Publishing Limited.
5. Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and machine learning: Limitations and opportunities. MIT Press.
6. Bazarkina, D. (2023). Current and future threats of the malicious use of artificial intelligence by terrorists: Psychological aspects. In The Palgrave Handbook of Malicious Use of AI and Psychological Security (pp. 251-272). Cham: Springer International Publishing.
7. Burton, J. (2023). Algorithmic extremism? The securitization of artificial intelligence (AI) and its impact on radicalism, polarization, and political violence. Technology in Society, 75, 102262.
8. Chesney, R., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107(6), 1753-1820.
9. Davis, A. L. (2021). Artificial intelligence and the fight against international terrorism. American Intelligence Journal, 38(2), 63-73.
10. Erendor, M. E. (Ed.). (2024). Cyber security in the age of artificial intelligence and autonomous weapons. CRC Press.
11. Esmailzadeh, Y. (2024). International terrorism and social threats of artificial intelligence.
12. Fuchs, C., & Chandler, D. (2019). Introduction: Big data capitalism - politics, activism, and theory. https://doi.org/10.16997/book29.a
13. Irfan, M., Almeshal, Z. A., & Anwar, M. (2023). Unleashing Transformative Potential of Artificial.
14. Johnson, J. (2019). Artificial intelligence & future warfare: implications for international security. Defense & Security Analysis, 35(2), 147-169.
15. Markarian, G., Karlovic, R., Nitsch, H., & Chandramouli, K. (Eds.). (2022). Security Technologies and Social Implications. John Wiley & Sons.
16. Montasari, R. (2023). Internet of things and artificial intelligence in national security: Applications and issues. In Countering Cyberterrorism: The Confluence of Artificial Intelligence, Cyber Forensics and Digital Policing in US and UK National Cybersecurity (pp. 27-56). Cham: Springer International Publishing.
17. Montasari, R. (2024). Analysing ethical, legal, technical and operational challenges of the application of machine learning in countering cyber terrorism. In Cyberspace, Cyberterrorism and the International Security in the Fourth Industrial Revolution: Threats, Assessment and Responses (pp. 159-197). Cham: Springer International Publishing.
18. Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. NYU Press.
19. O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown Publishers.
20. Shin, D. (2024). Artificial misinformation: Exploring human-algorithm interaction online.
21. Tufekci, Z. (2015). Algorithmic amplification of politics on Twitter. Proceedings of the National Academy of Sciences, 112(44), 13479-13484.
22. Yu, S., & Carroll, F. (2022). Implications of AI in national security: understanding the security issues and ethical challenges. In Artificial intelligence in cyber security: Impact and implications: Security challenges, technical and ethical issues, forensic investigative challenges (pp. 157-175). Cham: Springer International Publishing.
23. Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.
24. Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, *81*, 1–15. https://proceedings.mlr.press/v81/buolamwini18a.html
25. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5
26. Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
27. Klonick, K. (2018). The new governors: The people, rules, and processes governing online speech. *Harvard Law Review*, *131*(6), 1598–1670. https://harvardlawreview.org/2018/04/the-new-governors/
28. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms:

Mapping the debate. *Big Data & Society*, *3*(2), 1–21. https://doi.org/10.1177/2053951716679679
29. Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
30. Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A. F., & Meira, W. (2020). Auditing radicalization pathways on YouTube. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 131–141. https://doi.org/10.1145/3351095.3372879
31. Sunstein, C. R. (2017). #*Republic: Divided democracy in the age of social media*. Princeton University Press.
32. Williams, M. L., Burnap, P., Javed, A., Liu, H., & Ozalp, S. (2020). Hate in the machine: Anti-Black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime. *The British Journal of Criminology*, *60*(1), 93–117. https://doi.org/10.1093/bjc/azz049
33. Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. *Yale University Press*.
34. Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104(3), 671–732. https://doi.org/10.15779/Z38BG31
35. Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. Proceedings of the 2018 Conference on Fairness, Accountability and Transparency, 149–159. https://doi.org/10.1145/3287560.3287600
36. Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. Big Data, 5(2), 153–163. https://doi.org/10.1089/big.2016.0047
37. Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. Science Advances, 4(1), eaao5580. https://doi.org/10.1126/sciadv.aao5580
38. Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe. https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c
39. Dencik, L., Hintz, A., & Cable, J. (2016). Towards data justice? The ambiguity of anti-surveillance resistance in political activisms. Big Data & Society, 3(2), 1–12. https://doi.org/10.1177/2053951716679678
40. Milan, S., & Treré, E. (2019). Big data from the South(s): Beyond data universalism. Television & New Media, 20(4), 319–335. https://doi.org/10.1177/1527476419837739
41. Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. Big Data & Society, 4(2), 1–14. https://doi.org/10.1177/2053951717736335

Winner, L. (1980). Do artifacts have politics? Daedalus, 109(1), 121–136. https://www.jstor.org/stable/20024652