

Chapter

The Use of Big Data to Address Refugee Victimization: A Scoping Review

Gabriela Mesquita Borges and Laura Neiva

Abstract

This scoping review maps the empirical landscape of how Big Data has been used in the context of refugee victimization. Drawing on six empirical studies published between 2021 and 2022, identified through a systematic search across four electronic databases, the review examines the ways Big Data has been used to detect, represent, and respond to violence, exclusion, and neglect affecting refugees. The findings reveal that, while Big Data offers significant potential to enhance visibility, inform protective measures, and guide institutional responses, it also raises critical ethical challenges related to privacy, representativeness, algorithmic bias, and digital exclusion. Through a descriptive-analytical approach, the review underscores the dual role of Big Data: as both a means of addressing vulnerability and a mechanism that can inadvertently generate new forms of harm or reinforce existing inequalities. By offering a critical synthesis of the empirical literature, this review contributes to ongoing debates in refugee studies, highlighting the urgent need for context-sensitive, participatory, and ethically robust approaches to the use of Big Data in this context.

Keywords: Big Data, refugee victimization, digital exclusion, algorithmic bias, hate speech, surveillance

1. Introduction

In recent years, the global refugee crisis has intensified, with millions of individuals forcibly displaced due to armed conflict, political persecution, environmental degradation, and systemic instability. This escalation reflects not only the growing number of global crises but also the compounded vulnerabilities faced by displaced populations [1–12].

According to the 1951 Refugee Convention, refugees are outside their country of nationality or habitual residence and are unable or unwilling to return due to a well-founded fear of persecution based on race, religion, nationality, membership in a particular social group, or political opinion. This legal definition, while foundational, does not fully capture the range of hardships and risks refugees face throughout their displacement [13]. As refugees navigate complex and often

prolonged journeys – from initial displacement through transit, asylum-seeking, and resettlement – they encounter overlapping forms of victimization [13]. These include not only physical and psychological violence but also symbolic and structural harms such as discrimination, exclusion, exploitation, and institutional neglect [3].

This form of victimization is not limited to moments of crisis; it is frequently embedded within the very systems intended to provide protection. Inadequate access to healthcare, education, and legal support exacerbates this precarity [4, 7, 14, 15]. Scholars have drawn attention to how both structural and symbolic violence permeate migration systems, often concealed within humanitarian frameworks and institutional policies [4, 13, 16]. Media narratives also contribute to this marginalization by reinforcing racialized images of refugees as threats or outsiders [17, 18].

Institutional actors – governments, humanitarian agencies, and international organizations – play a pivotal role in shaping the refugee experience [5]. Yet, their responses remain fragmented, reactive, and often limited by data deficits, bureaucratic inefficiencies, and policy inertia [1]. While frameworks such as the Global Compact on Refugees promote responsibility-sharing and rights-based protection, systemic gaps persist [19]. A growing focus on border securitization further erodes institutional commitments to equitable, consistent, and sustainable support [7, 20].

Concurrently, the digital transformation of social life has created new arenas – both threatening and empowering – for refugees. Among the most prominent developments within this transformation is the growing use of Big Data to capture, process, and interpret refugee-related information across multiple digital environments.

As used in this paper, Big Data refers to the techniques, practices, and systems employed to collect, cross-reference, analyze, and process large volumes of structured and unstructured data from sources such as social media platforms, institutional records, electronic health data, and participatory digital environments [21]. These methods – including sentiment analysis, machine learning, topic modeling, and geospatial mapping – are increasingly utilized to identify patterns and correlations related to refugee victimization. Under this definition, a wide range of digital systems and practices, from social media monitoring to institutional data processing and participatory platforms, constitute forms or sources of Big Data. While these tools offer powerful means to make refugee experiences more visible, assess risks, and inform interventions, they also carry the potential to reproduce violence, exclusion, or neglect when applied uncritically or without participatory and ethical safeguards.

This dual capacity of Big Data is perhaps most visible in the role digital platforms play in shaping refugee representation, discourse, and exposure to violence. Digital platforms enable xenophobic discourse, surveillance, and systemic exclusion; on the other hand, they provide tools for visibility, advocacy, and community building. Studies show that far-right actors have weaponized digital platforms to spread antirefugee disinformation across Europe, particularly during high-profile crises such as the Evros border standoff and the Moria camp fire [22]. These hostile narratives often extend to NGOs and activists, undermining humanitarian advocacy through coordinated online harassment and discrediting campaigns [23].

At the same time, digital technologies can act as catalysts for cultural preservation and political participation [24]. Refugees and other marginalized groups are

increasingly using digital platforms to document their experiences, share narratives, and build transnational communities [23, 25–28]. These “cybercultures” offer spaces for identity negotiation and belonging. However, digital participation is also stratified. Access to digital infrastructure and literacy remains uneven, particularly for women and young people, who face layered vulnerabilities in displacement contexts [29–31]. Online spaces are also shaped by corporate and hegemonic cultural logics that often marginalize refugee voices or commodify their experiences.

This marginalization extends into migration governance. Digital technologies such as biometric registration, facial recognition, and mobile phone data extraction are now central to refugee management systems [24]. As Amnesty International [32] notes, these tools often operate without transparency or accountability and risk reinforcing racialized surveillance regimes. While such technologies may offer efficiencies and enhanced monitoring capabilities, they also pose significant threats to refugee privacy, due process, and protection [26].

Participatory platforms and social media channels hold promise for real-time feedback and localized response [33], but they also generate risks. Without strong data governance and community control, these systems may amplify inequality and exclusion, undermining the very rights they purport to protect [28]. Ethical concerns around consent, data representativeness, and algorithmic bias remain under-addressed [33].

While refugee studies have acknowledged the increasing role of digital systems, no scoping review to date has systematically examined how Big Data has been used in the context of refugee victimization. Although individual studies have examined specific applications or ethical concerns, there remains a need for a broader synthesis of how Big Data contributes to identifying vulnerability – or, conversely, how it may shape new forms of harm and reinforce existing challenges. Considering this, this review, by focusing specifically on Big Data and refugee victimization, and drawing on six empirical studies published between 2021 and 2022 – identified through a systematic search across four electronic databases – offers an original contribution to the fields of refugee and migration research, as well as Big Data.

Adopting a descriptive-analytical approach [34, 35], it highlights the dual role of Big Data: while it holds significant potential to enhance refugee visibility, inform protective measures, and support institutional responses, it simultaneously raises critical ethical concerns related to privacy, representativeness, algorithmic bias, and digital exclusion. It moves beyond a merely technical or celebratory view of innovation and instead offers a critical, interdisciplinary synthesis that foregrounds ethical, political, and methodological concerns. The findings call for more reflexive, context-sensitive, and inclusive approaches to Big Data – ensuring that these tools serve not only to detect harm but also to advance protection, inclusion, and justice for displaced populations.

2. Methods

To develop this scoping review, we followed the guidance for descriptive reviews by Levac and colleagues (2010) and adopted the methodological framework for scoping studies proposed by Arksey [34].

2.1 Stage 1: Identifying the research question

The research question of this scoping review is: *How has Big Data been used in the context of refugee victimization?*

2.2 Stage 2: Identifying relevant studies

A search of the publications on three electronic databases (Web of Science, Scopus, and Annual Reviews) and via Google Scholar was undertaken in April 2025, with no restriction set for the time of publication. We used the following search expressions to identify relevant studies on the topic: “Refugees” AND “Big Data” AND “Victimization.” In addition, this search was followed by backward reference tracking, examining the references in the selected publications based on full-text assessment.

2.3 Stage 3: Study selection

The inclusion criteria allowed only original empirical studies written in English, reporting data on refugees, victimization, and Big Data. Nonoriginal full-length studies (commentaries, editorials, notes, newspaper articles, conference proceedings, and reports) were excluded. The titles of 330 records were retrieved. After removing duplicates, 219 articles were screened and assessed for eligibility. The first two authors independently screened all the papers based on the title and abstract, and subsequently based on the full text. This was cross-checked and discussed in both phases, and perfect agreement was achieved. The screening process is shown in **Figure 1**. Based on the title and abstract assessments, 214 records were excluded because they were not related to the research question ($n = 208$) or were not empirical studies ($n = 6$). Of the 5 fully read articles, 5 met the inclusion criteria. One article was included following backward reference tracking, and the final scoping review was composed of 6 articles. The search and screening process is shown in **Figure 1**.

2.4 Stage 4: Charting the data from the primary research

We developed a structured data charting form to systematically organize and synthesize relevant information from each study included in this scoping review. This form was designed to capture key descriptive characteristics of the empirical literature, including author(s), year of publication, country of study origin, study aims, data collection methods, participants, sample size, and the specific technologies or analytical tools used. This comprehensive mapping provided a clear overview of the landscape of Big Data applications in refugee research and allowed for consistency in comparative analysis.

Following the “descriptive–analytical” framework proposed by Arksey [34] and further refined by Levac [35], we approached the data extraction not simply as a technical inventory but also as a step toward critically assessing the results. Particular attention was paid to how each study conceptualized and operationalized **Big Data** – including its sources (e.g., social media platforms, institutional datasets, participatory platforms), analytical methods (e.g., machine learning, sentiment analysis, geospatial mapping), and stated or implied objectives related to refugee victimization.

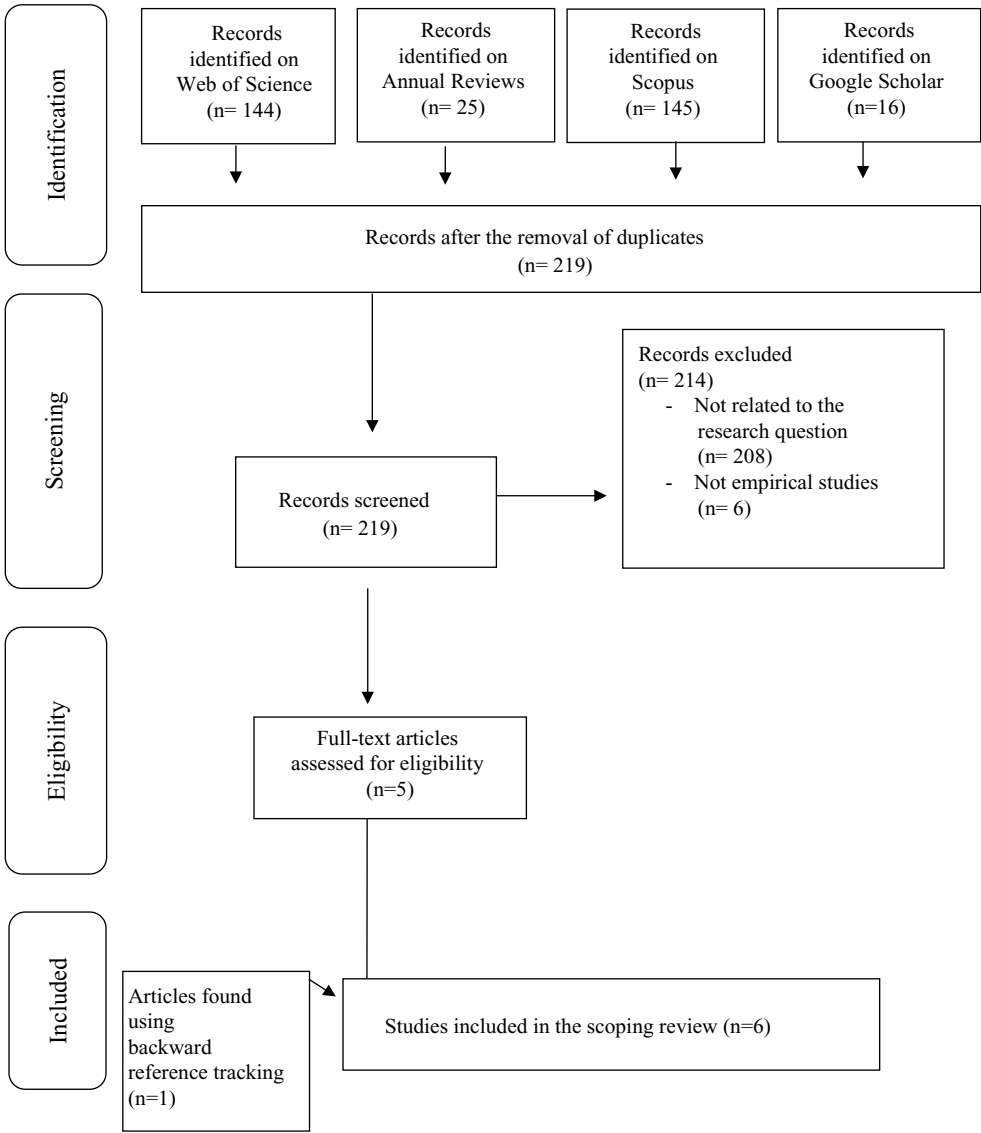


Figure 1.
Flowchart detailing the search and screening process of the scoping review.

Rather than organizing findings by methodological type or chronology, we opted for a conceptual structure that highlights the dual role of Big Data: (1) its capacity to respond to refugee victimization, and (2) its potential to reproduce or intensify vulnerability through bias, fragmentation, or exclusion.

2.5 Stage 5: Collating, summarizing, and reporting the results

Table 1 presents information about the main characteristics of the six included studies, ordered by year of publication. The main empirical achievements of the studies are discussed in the second part of the Results section.

Author(s)	Year	Country of origin	Study aim	Data collection methods	Participants	Sample size
Arcila-Calderón, C.; Blanco-Herrero, D.; Frías-Vázquez, M.; Seoane-Pérez, F.	2021	Spain	To analyze expressions of hate speech and sentiments on Twitter in Spain during the reception of the boat Aquarius.	Manual coding and computational analysis of tweets using sentiment analysis and topic modeling	Twitter users posting in Spanish using #Aquarius.	24,254 tweets
Huang, L. S.	2021	Canada	To identify and analyze the language-learning needs of adult Syrian refugees in Canada and realign language instruction accordingly.	Mixed methods: needs assessment surveys, semi-structured interviews, and analysis of learners' oral production using IELTS and NVivo.	17 LINC instructors and 14 adult Syrian refugee learners.	31 participants (survey), 17 interviews (instructors and learners), oral data from 14 learners.
Koukoulis, K. & Koukopoulos, D.	2021	Greece	To propose and evaluate crowdsourced collective memory management services using interactive refugee route maps to support cultural identity and integration in smart cities.	System development, user interaction testing, scenario-based evaluation.	Volunteers (refugees and descendants), specialists, and general users of the platform.	Not specified quantitatively; the system was tested via a case study and comparative evaluation.
Nyanchoga, M. M.; Lee, P.; Barbery, G.	2021	Australia	To explore electronic health records to estimate the extent of catch-up immunization and assess factors associated with under-immunization among refugees and asylum seekers in South-East Queensland.	Secondary analysis of electronic health records using statistical software (SPSS).	Refugees and asylum seekers referred to a specialized clinic in South East Queensland (aged 2 + years).	397 individuals

Author(s)	Year	Country of origin	Study aim	Data collection methods	Participants	Sample size
Sánchez-Holgado, P.; Amores, J. J.; Blanco-Herrero, D.	2022	Spain	To explore whether online hate speech on Twitter can be used as a predictor of social acceptance toward immigrants at the provincial level in Spain.	Analysis of 97,710 geolocated tweets using machine learning for hate speech detection; secondary data analysis from INE and CIS for immigration statistics and attitudes.	Geolocated tweets about migrants and refugees (2015–2020); Spanish citizens surveyed in the CIS 2017 study.	97,710 tweets; secondary survey data from CIS (exact number not specified).
Vrysis, L., Vryzas, N., Kotsakis, R., Saridou, T., Matsiola, M., Veglis, A., Arcila-Calderón, C., and Dimoulas, C.	2021	Greece (primary affiliation – first part of the study developed), with international collaboration (Spain, based on Arcila-Calderón – second part of the study developed).	To develop and present a web-based platform (PHARM) for the detection and analysis of hate speech against refugees and migrants across social media and online content.	Web scraping from multiple platforms (Twitter, Facebook, YouTube, news websites); sentiment and topic analysis; machine learning classification; geolocation tools.	Not applicable (no human subjects interviewed; data collected from publicly available digital content).	Multiple corpora are used; precise numeric sample is not given, but data is sourced continuously from multiple social platforms and web sources in real time.

Table 1.
 Overview of the five empirical studies included in the article.

3. Results

The studies span diverse geographic and thematic contexts, including the real-time monitoring of antirefugee sentiment through social media mining and automated analysis in Spain, the deployment of a web interface for multiplatform hate speech detection and analysis across Southern Europe, the use of participatory digital platforms to reconstruct disrupted narratives and document resilience in Greece, the exposure of immunization access disparities via institutional Big Data analysis in Australia, and the role of learner-generated data in shaping inclusive language programs for Syrian refugees in Canada.

4. Study characteristics

The included studies were published between 2021 [36–39] and 2022 [40]. All studies were conducted in single countries: three in Spain [39, 40], one in Canada [36], one in Greece [37], and one in Australia [38].

The methodological approaches varied: four studies employed computational or statistical methods using large datasets, particularly social media analysis and machine learning [39, 40] and electronic health records (EHRs) [38]; one study used a mixed-methods design, incorporating surveys, interviews, and language production analysis [36], while another focused on system development and scenario-based user evaluations [37].

Participants ranged from Twitter users [39, 40], social media users, and digital content creators across multiple platforms [39], to adult Syrian refugees and Language Instruction for Newcomers to Canada (LINC) instructors in Canada [36], and from healthcare clients and professionals involved in immunization programs [38] to volunteers and specialists engaged in digital heritage and identity-building platforms [37]. Sample sizes also varied, with studies analyzing 50,000–60,000 tweets and online posts [39]: 24,254 tweets [40]; 97,710 tweets [39]; multiplatform corpus (precise figures not reported), health records of nearly 400 individuals [38], and qualitative input from over 30 participants [36].

5. The use of big data to address refugee victimization

The review highlights the diverse applications of Big Data in addressing refugee victimization, revealing both its transformative potential and the critical challenges that accompany its use. Spanning applications in digital discourse analysis, hate speech detection, narrative preservation, healthcare data integration, and structured educational assessments, the six studies reviewed illustrate the multifaceted roles Big Data can play in making refugee victimization more visible – while also exposing critical gaps in coverage, interpretation, and inclusivity.

The study by Vrysis [39] provides empirical evidence on the application of Big Data to examine refugee victimization through public digital discourse. Analyzing 24,254 Twitter posts related to the reception of the Aquarius rescue ship in Spain, the authors employed sentiment analysis and topic modeling to detect patterns of online hostility. Their findings show that 27% of the tweets expressed hate or rejection toward refugees, while 9.9% were explicitly categorized as hate speech. These tools

enabled large-scale detection and emotional analysis of discourse, highlighting the capacity of Big Data to identify targeted hostility and track shifts in public sentiment. A temporal analysis revealed a paradox: while hate speech increased following the Spanish Prime Minister's decision to accept the Aquarius, overall sentiment became slightly more positive. This underscores the ability of Big Data to reveal layered and sometimes contradictory emotional trends in refugee-related discourse.

Vrysis [39] extends this potential through the development of the PHARM platform, a multiplatform web interface designed to detect hate speech against refugees and migrants across Twitter, Facebook, YouTube, and digital news outlets. By integrating machine learning, sentiment lexicons, and geospatial mapping, the platform enables real-time monitoring of xenophobic discourse. The study illustrates the role of Big Data infrastructures in consolidating fragmented digital sources and supporting the early detection of refugee-directed hostility. However, it also exemplifies the limitations of such tools. The reliance on automated classifiers and lexicon-based analysis presents challenges in interpreting nuanced content, such as sarcasm, coded language, or multilingual posts. These limitations are compounded by reliance on public data, which introduces sampling biases and excludes offline or less visible forms of victimization.

Sanchez-Holgado [40] offers a further example of Big Data use in monitoring refugee-targeted hostility. Using a dataset of 97,710 geolocated tweets from 2015 to 2020, the authors applied a machine learning classifier to identify hate speech directed at refugees in Spain. They correlated these findings with demographic data and public opinion surveys to assess whether digital hate discourse could predict local attitudes toward forced displacement. Although no significant statistical relationship was found, the study underscores both the promise and the interpretive risks of using digital trace data. The authors caution that online discourse is shaped by complex sociopolitical dynamics and that algorithmic tools are vulnerable to misclassification, particularly in relation to irony and culturally contingent expressions. This highlights the need for contextual and qualitative triangulation when using Big Data to assess exclusionary sentiment.

Koukoulis [37] presents Crowdpower, a participatory platform that applies principles associated with Big Data – such as metadata tagging, modular system design, and scalable content aggregation – to preserve refugee migration narratives. Although the platform does not employ predictive analytics or real-time data mining, it enables refugees and their descendants to contribute personal stories, images, videos, and geotagged content assembled into interactive maps. Migration routes are conceptualized as dynamic data containers, linking narrative fragments that reconstruct the emotional and cultural dimensions of displacement. The system affirms refugee agency and digital storytelling as tools for empowerment and cultural preservation. However, the lack of real-time monitoring and limited predictive capacity pose challenges for broader institutional adoption. Questions of data validation, long-term scalability, and systemic impact remain open.

Nyanhoga [38] demonstrates how Big Data is applied in healthcare settings to reveal refugee victimization through institutional neglect. Analyzing the EHRs of 397 refugee clients in South East Queensland, the study identified significant gaps in vaccine coverage. This use of Big Data exposes structural neglect that might otherwise be obscured in aggregated statistics. It also underscores the potential for digital integration across sectors – such as health, immigration, and settlement – to improve service delivery. However, the study also notes that fragmentation between

systems can limit the effectiveness of data infrastructures, leading to delayed or inadequate care. Even where Big Data exists, its underutilization or lack of coordination can reinforce vulnerability.

Huang [36], a mixed-methods study of adult Syrian refugees in Canada's LINC program, does not deploy Big Data technologies directly, but it draws on structured datasets analyzed with NVivo software and IELTS descriptors. The study highlights how institutional barriers – such as rigid curriculum standards, language assessments, and digital literacy gaps – mediate refugee vulnerability in education. Although based on qualitative tools, the structured nature of the data collection and analysis parallels certain aspects of Big Data methodology. However, the study also reveals how such structured systems can marginalize already vulnerable learners when they fail to reflect diverse socio-cultural realities.

Together, these studies present a nuanced picture. Big Data offers powerful capabilities for enhancing refugee visibility, detecting targeted hostility, and informing more inclusive policies. Yet, these same systems face critical limitations: algorithmic misclassification, data representativeness issues, lack of contextual sensitivity, and structural fragmentation can all reduce their effectiveness or even reinforce the exclusions they aim to redress. Moreover, the adoption of Big Data tools without participatory design or ethical safeguards risks depoliticizing refugee experiences and overlooking localized forms of victimization. For Big Data to contribute meaningfully to refugee protection, its use must be context-aware, ethically grounded, and integrated with broader institutional reforms that center refugee voices and address systemic forms of neglect.

6. Discussion

This scoping review highlights the diverse applications of Big Data in addressing refugee victimization, examining its potential to reveal patterns of violence, exclusion, and neglect, as well as its role in generating new forms of harm and reinforcing existing challenges. However, it is crucial to underscore that these insights are drawn from only six empirical studies meeting the strict inclusion criteria. This small corpus reflects the nascent and exploratory stage of research at this intersection. Consequently, the findings discussed here should be interpreted as indicative of emerging trends rather than as a comprehensive or generalizable account of the field.

The selected studies reflect a growing methodological shift in refugee research toward the adoption of large-scale, algorithmic, and automated approaches aimed at detecting patterns of exclusion, hostility, and neglect. While these technologies offer the potential to enhance visibility and responsiveness, they also raise critical ethical, methodological, and epistemological challenges.

Across the reviewed literature, a dual pattern emerges: on one hand, Big Data is shown to reveal hidden dimensions of refugee victimization that traditional methods may overlook – such as digital hate speech, service disparities, and emotional undercurrents in public sentiment. On the other hand, the same tools can introduce new risks, including algorithmic misclassification, contextual oversimplification, and the reproduction of institutional blind spots. The studies analyzed illustrate this ambivalence clearly.

Vrysis [39] and Sanchez-Holgado [40] demonstrate how sentiment analysis, topic modeling, and machine learning can detect hate speech against refugees on

social media platforms, revealing correlations between hostility and negative emotional tone, especially during politically sensitive moments. Vrysis [39] goes further by developing PHARM, a real-time platform that consolidates data from multiple digital sources to monitor xenophobic discourse. Their use of geospatial mapping and automated classifiers exemplifies how Big Data can support large-scale tracking and visualization of refugee-directed hostility. Meanwhile, Nyanhoga [38] analyzes EHRs to expose vaccination gaps among refugee populations, showing how institutional datasets – when mobilized effectively – can identify forms of structural neglect otherwise invisible in aggregated national statistics. Huang [36]'s study, while based on structured qualitative data rather than Big Data per se, reinforces the value of systematic data use in revealing institutional barriers in refugee education and integration processes.

However, these technological applications are accompanied by clear limitations. Vrysis [39] and Sanchez-Holgado [40] warn against the interpretive risks of relying on sentiment metrics alone, noting that irony, coded language, or cultural nuance often elude algorithmic detection. These limitations highlight a broader concern: that Big Data, despite its capacity for scale, may lack the cultural calibration and contextual sensitivity necessary to fully grasp the complexity of refugee victimization. This echoes broader critiques of how algorithmic systems – especially in migration governance – may reflect and reproduce symbolic and structural violence embedded in state and humanitarian practices [13, 16, 4].

Vrysis [39] raises similar concerns, pointing to classifier bias and content ambiguity that undermine the generalizability and ethical use of automated detection tools. These risks are intensified in securitized digital environments, where technologies such as biometric registration, facial recognition, and mobile phone surveillance often operate without transparency or accountability, reinforcing racialized forms of monitoring and social control [24, 26, 32].

Nyanhoga [38] shows that even institutional Big Data – such as health records – can be underutilized or poorly integrated across systems. Fragmented data governance and lack of cross-sector coordination reduce the proactive potential of these tools, allowing structural neglect to persist. Huang [36]'s study deepens this critique by demonstrating how educational institutions, despite having access to detailed learner data, may reinforce exclusion through rigid assessment frameworks, limited inclusivity, and inadequate attention to digital literacy.

Koukoulis [37] offers a contrasting vision of Big Data use through Crowdpower, a participatory platform that centers refugee agency in the co-production of migration narratives. While not predictive or real-time, the platform embodies key principles of Big Data – such as metadata tagging and scalable aggregation – while subverting its extractive tendencies. By prioritizing ethical storytelling, cultural preservation, and collaborative knowledge-making, this approach challenges dominant digital paradigms that often erase refugee perspectives. This addresses the problem of epistemic exclusion, as noted by Berg [29, 30], where refugee voices are marginalized by design in dominant digital systems – either through lack of access, algorithmic bias, or platform governance structures [28, 31].

These findings underscore the importance of combining the scale of Big Data with the depth of qualitative insight. While Big Data offers valuable tools for detecting emerging threats – such as hate speech, misinformation, or systemic service gaps – its effectiveness depends on ethical reflexivity, contextual interpretation, and methodological pluralism. Yet, given the very limited empirical base, caution is

warranted in extrapolating these results to broader contexts. The evidence available provides a useful snapshot of methodological and ethical tendencies, but the absence of longitudinal, comparative, and cross-regional analyses limits the robustness of general claims. Technological optimism must, therefore, be tempered by critical awareness of power asymmetries embedded in data design, access, and application – especially when dealing with displaced and vulnerable populations.

This review also raises questions about the institutional responsibilities shaping data governance in refugee contexts. As noted by McAdam [19] and [20], policy approaches often prioritize border control and bureaucratic efficiency over long-term protection. These tendencies are mirrored in digital contexts, where securitized logics, fragmented systems, and limited accountability continue to undermine refugee rights and dignity [1, 4].

Furthermore, the reviewed literature remains geographically imbalanced. Most studies are based in high-income Western countries, with limited empirical research emerging from the Global South, where the majority of the world's refugee populations reside. Countries such as Lebanon, Turkey, Kenya, and Bangladesh are underrepresented despite their critical relevance. This contributes to global knowledge hierarchies that prioritize Western data regimes and overlook the socio-political dynamics of refugee-hosting states [4].

In addition, the small number of studies ($n = 6$) reviewed here highlights both the novelty and the fragility of the empirical evidence base. The scarcity of studies constrains the ability to draw general conclusions but underscores the value of this synthesis as a first step toward systematizing a still-emerging research domain. Future research should build on this foundation by expanding the scope to include underrepresented regions, diverse data sources, and interdisciplinary methodologies that better capture local and cultural specificities.

Finally, a notable gap persists in the inclusion of refugee voices in the development, implementation, and evaluation of Big Data tools. Few studies engage refugees as active participants or cocreators of knowledge. This reinforces epistemic inequalities and raises concerns about who defines victimization, who controls data narratives, and how those narratives are operationalized in policy [3, 4, 13].

In sum, this review affirms the analytical potential of Big Data in refugee research but cautions that the current empirical base remains too limited to support broad generalizations. Its application must, therefore, be ethically grounded, context-sensitive, and participatory. Without these safeguards, data-driven innovation risks reproducing the very exclusions it seeks to redress.

7. Limitations

This scoping review has several important limitations that should be acknowledged to appropriately contextualize its findings.

First, the empirical base is extremely limited: only six studies met the strict inclusion criteria. This reflects the novelty of research at the intersection of Big Data and refugee victimization, but it also constrains the generalizability of the conclusions. The review should, therefore, be read as a preliminary mapping of an emerging field rather than as a comprehensive synthesis.

Second, all included studies were published between 2021 and 2022 and primarily originated from high-income, Western contexts. This temporal and geographic

concentration limits the diversity of perspectives represented, particularly from the Global South, where the majority of the world's refugees reside. As such, the insights generated here may disproportionately reflect Western data infrastructures, institutional frameworks, and socio-political contexts.

Third, the methodological heterogeneity of the included studies – ranging from machine learning applications and institutional data analyses to participatory and mixed-method designs – complicates direct comparison across findings. This variability, while enriching the descriptive scope, precludes systematic assessment of efficacy, validity, or replicability.

Finally, although the scoping review followed recognized methodological guidelines [34, 35], publication bias and language restrictions (English-only inclusion) may have led to the exclusion of relevant nonindexed or regional studies.

Future research should broaden both the linguistic and geographic scope of systematic searches and encourage empirical studies that integrate ethical, participatory, and cross-sectoral dimensions of Big Data use in refugee contexts.

8. Conclusion

As Big Data technologies become increasingly embedded in the study and governance of refugee populations, this scoping review provides an initial descriptive and analytical mapping of their diverse applications in addressing refugee victimization – highlighting both their potential to uncover patterns of violence, exclusion, and neglect, and their capacity to generate new forms of harm and reinforce existing challenges. Based on only six empirical studies, the evidence base remains limited and uneven, warranting cautious interpretation of the patterns identified. The findings presented here should, therefore, be read as indicative of emerging directions rather than conclusive trends.

The six studies reviewed reveal a fundamental tension between the promise of enhanced visibility, responsiveness, and institutional insight on one hand, and the risks of exclusion, misrepresentation, and neglect on the other. Big Data has enabled researchers and institutions to detect hate speech, identify health disparities, visualize large-scale hostility, and coconstruct refugee narratives with unprecedented scale and detail. These capacities offer the potential to surface marginalization that may go unnoticed through conventional methods. When used critically and responsibly, Big Data can contribute to a more proactive and informed response to the complex realities of forced displacement – particularly when deployed across platforms and sectors with integrative, real-time functionalities.

However, the relatively small number of studies included in this review underscores both the novelty and the fragility of the empirical field. This limitation does not diminish the contribution of the review; rather, it highlights its role as a foundational effort to consolidate a scattered and emerging area of inquiry. Future research must expand this evidence base through longitudinal, comparative, and cross-regional analyses, particularly in the Global South, to ensure that insights are contextually grounded and globally representative.

Nonetheless, the limitations of current Big Data practices remain significant. Algorithmic classifiers continue to struggle with nuance, irony, and culturally embedded expressions, while fragmented or siloed data systems restrict the ability to act on insights in a coordinated and effective way. Moreover, refugee voices are often

excluded from the design, validation, and interpretation of digital tools, reinforcing epistemic inequalities and undermining participatory justice. The depoliticizing effect of abstraction – where complex experiences of displacement are reduced to patterns or data points – also raises concerns about the ethical desirability of such approaches when used uncritically.

Importantly, the studies reviewed reflect a growing awareness of these risks and a willingness among scholars to move beyond simplistic views of technological progress. Participatory platforms like Crowdpower (as shown by Koukoulis [37]), critical analyses of algorithmic bias, and efforts to combine computational tools with qualitative insight suggest a promising path forward for recalibrating Big Data use in refugee research. These examples point to the value of hybrid, ethically grounded methods that center refugee agency, cultural specificity, and the need for institutional accountability.

Rather than positioning Big Data as inherently progressive or inherently problematic, this review advocates for interdisciplinary and reflexive frameworks that combine computational capacity with grounded, context-sensitive analysis. Such models are essential to ensure that technological innovation in refugee studies aligns with the core values of protection, inclusion, and social equity.

The implications of these findings extend to both research and policy. By identifying the capabilities and constraints of existing Big Data applications, this review can inform the development of ethical standards in refugee data governance, promote cross-sectoral integration, and support the creation of rights-based digital tools that respond to the actual needs and experiences of displaced populations. Policymakers, humanitarian actors, and technologists must work collaboratively to ensure that Big Data infrastructures not only expose vulnerabilities but actively contribute to meaningful protection and social transformation.

Ultimately, while Big Data continues to reshape humanitarian knowledge and decision-making, sustained attention to ethical governance, power asymmetries, and the lived realities of refugees remains essential. Given the limited and geographically narrow empirical foundation, these conclusions should be viewed as a starting point for deeper, more inclusive, and empirically grounded inquiry – one capable of preventing the reinforcement of the very exclusions these tools aim to address.

Conflict of Interest

The authors declare no competing interests.

Acknowledgments

The authors acknowledge the use of ChatGPT for language polishing of the manuscript.

Author details

Gabriela Mesquita Borges^{*1} and Laura Neiva²

1 Center for Legal, Economic, and Environmental Studies (CEJEIA),
Communication and Society Research Center (CESC), Lusíada University of Porto,
Porto, Portugal

2 Department of Sociology, Center for Legal, Economic, and Environmental Studies
(CEJEIA), Communication and Society Research Center (CESC), Lusíada University
of Porto, Institute for Social Sciences, University of Minho, Braga, Portugal

*Address all correspondence to: gabrielamborges@por.ulusiada.pt

IntechOpen

© 2025 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Aas KF. *The Crimmigrant Other: Migration and Penal Power*. London: Routledge; 2019
- [2] Ataç I, Rygiel K, Stierl M. The contentious politics of refugee and migrant protest and solidarity movements: Remaking citizenship from the margins. *Citizenship Studies*. 2016;**20**(5):527–544. DOI: 10.1080/13621025.2016.1182681
- [3] Borges GM. Journey of violence: Refugee women's experiences across three stages and places. *Journal of International Migration and Integration*. 2024a;**25**:673–693. DOI: 10.1007/s12134-023-01102-z
- [4] Borges GM. Refugee women's intersecting social locations: A contemporary perspective on the process of becoming a refugee woman. *International Criminology*. 2024b;**4**:279–292. DOI: 10.1007/s43576-024-00138-0
- [5] Canning V. *Gendered Harm and Structural Violence in the British Asylum System*. London: Routledge; 2017. DOI: 10.4324/9781315720975
- [6] Canning V. Corrosive control: State-corporate and gendered harm in bordered Britain. *Critical Criminology*. 2020;**28**:259–275. DOI: 10.1007/s10612-020-09509-1
- [7] Freedman J. Sexual and gender-based violence against refugee women: A hidden aspect of the refugee "crisis." *Reproductive Health Matters*. 2016;**4**(47):18–26. DOI: 10.1016/j.rhm.2016.05.003
- [8] Freedman J. Grand challenges: Refugees and conflict. *Frontiers in Human Dynamics*. 2019;**3**:1–3. DOI: 10.3389/fhumd.2019.00001
- [9] Ilcan S. The borderization of waiting: Negotiating borders and migration in the 2011 Syrian civil conflict. *Environment and Planning C: Politics and Space*. 2020;**40**(5):1012–1031
- [10] Ilcan S. The border harms of human displacement: Harsh landscapes and human rights violations. *Social Sciences*. 2021;**10**(123):1–16. DOI: 10.3390/socsci10040123
- [11] Reilly N, Sahraoui N, McGarry O. Exclusion, minimization, inaction: A critical review of Ireland's policy response to gender-based violence as it affects migrant women. *Frontiers in Human Dynamics*. 2021;**3**:1–17. DOI: 10.3389/fhumd.2021.642445:contentReference[oaicite:0]{index=0}
- [12] Tastsoglou E, Petrinioti X, Karagiannopoulou C. The gender-based violence and precarity nexus: Asylum-seeking women in the Eastern Mediterranean. *Frontiers in Human Dynamics*. 2021;**3**:1–17. DOI: 10.3389/fhumd.2021.660682:contentReference[oaicite:1]{index=1}
- [13] Borges GM, Faria R. Breathing under water: Gendering the violence against refugee women. In: Bošković M, Globalization and Its Impact on Violence against Vulnerable Groups. USA: IGI Global; 2020. p. 1–25. DOI: 10.4018/978-1-5225-9627-1.ch001
- [14] Farmer P. *Pathologies of Power: Health, Human Rights, and the New*

War on the Poor. California: University of California Press; 2005

[15] Horn R, Wachter K, Friis-Healy EA, Ngugi SW, Creighton J, Puffer ES. Mapping complex systems: Responses to intimate partner violence against women in three refugee camps. *Frontiers in Human Dynamics*. 2021;**3**:1–15. DOI: 10.3389/fhumd.2021.613792

[16] Chamberlain-Salaun J, Mills J, Usher K. Linking symbolic interactionism and grounded theory methods in a research design: From Corbin and Strauss' assumptions to action. *SAGE Open*. 2013;**3**(3). DOI: 10.1177/2158244013505757

[17] Bailey OG, Harindranath R. Racialised 'othering': The representation of asylum seekers in news media. In: Stuart A, editor. *Journalism: Critical Issues*. Maidenhead: Open University Press; 2005. p. 274–286

[18] Himmel R, Baptista MM. Migrants, refugees and othering: Constructing Europeaness. An exploration of Portuguese and German media. *Comunicação E Sociedade*. 2020;**38**:179–200

[19] McAdam J. The global compacts on refugees and migration: A new era for international protection? *International Journal of Refugee Law*. 2018;**30**(4):571–574. DOI: 10.1093/ijrl/eez004

[20] Mayblin L, Wake M, Kazemi M. Necropolitics and the slow violence of the everyday: Asylum seeker welfare in the postcolonial present. *Sociology*. 2019;**54**(1):107–123. DOI: 10.1177/0038038519862124

[21] Cukier KN, Mayer-Schoenberger V. The rise of Big Data. *Foreign Affairs*. 2013;**92**(3):28–40

[22] Institute for Strategic Dialogue. The Networks and Narratives of Anti-refugee Disinformation in Europe. 2021. Available from: <https://www.isdglobal.org/wp-content/uploads/2021/07/The-networks-and-narratives-of-anti-migrant-discourse-in-Europe.pdf> [Accessed: 2025-April-12]

[23] Alencar A, Kondova K, Ribbens W. The smartphone as a lifeline: An exploration of refugees' use of mobile communication technologies during their flight. *Media, Culture & Society*. 2019;**41**:828–844

[24] Berg M. Refugee integration in Germany: The interplay of othering, digital exclusion, and identity negotiation. *Journal of International Migration and Integration*. 2025;**26**:235–254. DOI: 10.1007/s12134-025-01238-0

[25] AbuJarour SA, Krasnova H, Diaz Andrade A, Olbrich S, Tan CW, Urquhart C, Wiesche M. Empowering refugees with technology: Best practices and research agenda. In: *Proceedings of the European Conference on Information Systems*. 2017

[26] Alsaleh A. The impact of technological advancement on culture and society. *Scientific Reports*. 2024;**14**:32140. DOI: 10.1038/s41598-024-83995-z

[27] Andrade AD, Doolin B. Information and communication technology and the social inclusion of refugees. *MIS Quarterly*. 2016;**40**:405–416

[28] Gillespie M, Osseiran S, Cheesman M. Syrian refugees and the digital passage to Europe: Smartphone infrastructures and affordances. *Social Media + Society*. 2018;**4**(1):2056305118764440

[29] Berg M. Women refugees' media usage: Overcoming information precarity

in Germany. *Journal of Immigrant & Refugee Studies*. 2021;**20**(1):125–139. DOI: 10.1080/15562948.2021.1907642

[30] Berg M. Information-precarity for refugee women in Hamburg, Germany, during the COVID-19 pandemic. *Information, Communication & Society*. 2022;1–17. DOI: 10.1080/1369118X.2022.2129271

[31] Kutscher N, Hüttmann J, Fujii MS, Engfer NP, Friedrichs-Liesenkötter H. Educational participation of young refugees in the context of digitized settings. *Information, Communication & Society*. 2022;**25**(4):570–586

[32] Amnesty International. Defending the Rights of Refugees and Migrants in the Digital Age. 2024. Available from: <https://www.amnesty.org/en/documents/pol40/7654/2024/en/> [Accessed: 2025-March-21]

[33] UNHCR. Using Social Media in Community-based Protection: A Guide. 2021. Available from: <https://www.unhcr.org/innovation/wp-content/uploads/2021/01/Using-Social-Media-in-CBP.pdf> [Accessed: 2025-May-13]

[34] Arksey H, O'Malley L. Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*. 2005;**8**(1):19–32. DOI: 10.1080/1364557032000119616

[35] Levac D, Colquhoun H, O'Brien KK. Scoping studies: Advancing the methodology. *Implementation Science*. 2010;**5**(1):69. DOI: 10.1186/1748-5908-5-69

[36] Huang LS. “I have big, big, big dream!”: Realigning instruction with the language-learning needs of adult Syrians with refugee experience in

Canada. *Refuge: Canada's Journal on Refugees*. 2021;**37**(2):141–163

[37] Koukoulis K, Koukopoulos D. Rediscovering refugee routes: On the creation of collective memory map services. *Procedia Computer Science*. 2021;**191**:41–48

[38] Nyanchoga MM, Lee P, Barbery G. Exploring electronic health records to estimate the extent of catch-up immunisation and factors associated with under-immunisation among refugees and asylum seekers in South East Queensland. *Vaccine*. 2021;**39**(42):6238–6244

[39] Vrysis L, Vryzas N, Kotsakis R, Saridou T, Matsiola M, Veglis A, Arcila-Calderón C, Dimoulas C. A web interface for analyzing hate speech. *Future Internet*. 2021;**13**(3):80. DOI: 10.3390/fi13030080

[40] Sanchez-Holgado P, Amores JJ, Blanco-Herrero D. Online hate speech and immigration acceptance: A study of Spanish provinces. *Social Sciences-Basel*. 2022;**11**(11):1–20