AI and Ethics
https://doi.org/10.1007/s43681-024-00639-8

REVIEW

Learning about AI ethics from cases: a scoping review of AI incident repositories and cases

Simon Knight[1,2,3] · Cormac McGrath[3,5] · Olga Viberg[3,4] · Teresa Cerratto Pargman[3,6]

Received: 24 August 2024 / Accepted: 24 November 2024

Abstract
Cases provide a practical resource for learning regarding the uses and challenges of AI applications. Cases give insight into how principles and values are implicated in real contexts, the trade-offs and different perspectives held regarding these contexts, and the—sometimes hidden—relationships between cases, relationships that may support analogical reasoning across contexts. We aim to (1) provide an approach for structuring ethics cases and (2) investigate existing case repository structures. We motivate a scoping review through a conceptual analysis of ethics case desirable features. The review sought to retrieve repositories, (sometimes known as observatories, catalogues, galleries, or incident databases), and their cases, for analysis of their expression of ethics concepts. We identify n = 14 repositories, extracting the case schema used in each, to identify how this metadata can express ethical concepts. We find that most repositories focus on harm-indicators, with some indicating positive impacts, but with little explicit reference to ethical concepts; a subset (n = 4) includes no structural elements addressing ethical concepts or impacts. We extract a subset of cases from the total cases (n = 2000) across repositories addressing education (n = 100). These are grouped by topic, with a structured content analysis provided of ethical implications from one sub-theme, offering qualitative insights into the ethical coverage. Our conceptual analysis and empirical review exemplify a model for ethics cases (shorthanded as Ethics-case-CPR), while highlighting gaps both in existing case repositories and specific examples of cases.

Keywords Design principles · Incidents · AI ethics · Case-based learning · Education

✉ Simon Knight
Simon.knight@uts.edu.au

Cormac McGrath
cormac.mcgrath@edu.su.se

Olga Viberg
oviberg@kth.se

Teresa Cerratto Pargman
tessy@dsv.su.se

1 Centre for Research on Education in a Digital Society and Transdisciplinary School, University of Technology Sydney, Sydney, Australia

2 UCL Knowledge Lab, University College London, London, UK

3 Digital Futures, Stockholm, Sweden

4 Division of Media Technology and Interaction Design, KTH, Stockholm, Sweden

5 Department of Education, Stockholm University, Stockholm, Sweden

6 Department of Computer and Systems Sciences, Stockholm University, Stockholm, Sweden

1 Introduction

The growing public discourse around artificial intelligence (AI) and its potential has been paralleled by increasing focus on the ethical impacts of AI in society, and the responsibilities of companies, developers, regulators, and the public in understanding AI and promoting responsible (dis)engagement with it. Ethical reasoning can be conceptualised as involving the application of principles to specifics of particular cases, supported by analogical reasoning that supports inferences from one case, applied to similar ones [1]; guidelines thus offer a resource to support such reasoning.

Over 100 guidelines and sets of principles for ethical AI have emerged (see §7.1), alongside significant public and policy discourse providing perspectives on potential areas of application for AI, including in the media (§7.2), and a number of repositories of cases of both positive and harms-oriented AI impact (§7.3). Such cases are often provided in guideline documents. *Cases* are examples of particular situations arising, in this paper they are taken to include

Published online: 09 January 2025

relatively brief expressions of events or incidents that have occurred, alongside relatively more detailed reporting. Use of cases is a common tool across contexts to support learning regarding the application of particular principles or theories to practical, concrete examples.

As we elaborate below, AI ethics cases may offer information to help people to learn about ethical concepts as they apply to particular technologies, applications, or implementations thereof, using the shorthand 'Ethics-case CPR' (Concepts, Perspectives, Relations):

1. Ethical concepts: such as the principles and values that are implicated, providing a lens to view the characteristics of the case;
2. *Ethical perspectives*: that provide a lens onto the dilemmas or trade-offs in the application of ethical principles or values, and the potential for stakeholders to hold different perspectives and the role of positionality in that;
3. *Ethical relations*: that help demonstrate the connections between cases and their nuanced similarities and differences. This supports analogical reasoning across cases, and shared discourse regarding types of cases.

To ground our analysis of the significance of Ethics-case CPR, we provide elaboration and justification of these desirable features in §7, with each subsection concluding with an overview of existing resources that address these considerations and gaps thereof. This conceptual analysis forms a part of the contribution of the paper. Our *first* feature for cases is important because *understanding key ethical concepts* is critical both to share and develop a language around issues for discussion and navigation of those issues, and because of the critical role that language plays in developing a sensitivity to these concepts. The *second* feature is key because ethics should not be thought of as a means to identify the correct outcome or rules, but rather, as the process through which we navigate implications, where there are often trade-offs and tensions in any given decision (e.g., the free speech principle that suggests one might publish other's private lives and privacy principles that might constrain that speech). The *third* feature is important because through identifying connections between cases, we can see the relationship between the specific (within the case), and the more general. These connections may be hidden for example, because cases involve the use of similar digital technologies in different implementation domains, or because they relate to similar or even identical AI uses, where one implementation is deemed effective and another harmful. As such, connections help us understand what could be done differently, and to develop a model or shared language for what may appear quite disparate things.

Given the range of resources now available around AI ethics—including principles, which should enshrine key ethical concepts; and incident repositories, which should both instantiate those concepts into cases, and draw attention to tensions within and connections between those cases—it is important to understand how these resources might support learning about ethical issues. To address this concern, this paper analyses instances of one form of case collection—*incident repositories*—through the following question:

RQ1 How do case repositories represent the ethical concepts implicated in reported cases? Specifically, how do repositories represent the Ethics-case-CPR elements, (a) ethical concepts, values and principles; (b) ethical perspectives; and (c) ethical relations.

To provide a deeper analysis and consideration of the kinds of issues covered in repositories, we sample cases from a single domain (applications of AI in education contexts), addressing our second question:

RQ2 *What ethical concepts are reflected in cases focused on educational applications?* Through RQ2, we seek to understand how the content of cases can inform understanding of the issues at stake, providing an example analysis, and resources for this analysis, that may be applied in other domains.

The remainder of the paper first (§7) highlights the role of cases in developing ethical reasoning against our desirable features, with each subsection mapping this discussion to example resources. Our empirical review approach is described (§8), with analysis addressing the research questions above (§9). The paper discussion (§10) closes with a proposal for a case model.

2 Background: role of cases in developing ethical reasoning

2.1 Cases to share and develop our ethical concepts and principles

How should we provide cases that help people understand how ethical values and concepts relate to particular situations? Recent reviews of tools for developing and assessing AI-based systems, and impact assessments, identified 352 [2] and 38 [3] respectively. There have been four recent reviews[1] of the myriad AI ethics principles and guidelines,

[1] Since writing a further review of 200 guidelines has been published (based on automated title and publisher comparisons with a manual check, 108 of which do not appear in earlier reviews) [4]

to which we might turn for guidance regarding ethics principles. These have focused on systematic review of scholarly literature (27 articles) for principles and challenges (identifying 22 and 15 respectively) [5]; public, private, and NGO sector AI policies (with 112 documents, 25 countries, and 25 ethical topics identified) [6, 7]; a purposive sample of 'prominent' AI principles documents in a range of languages (36 documents, identifying 47 principles under 8 key themes) [8]; and a scoping review of international guidelines on ethical AI from a range of sources (84 identified, with 11 ethical principles commonly identified, and 18 'impact' characteristics) [9]. In addition, Algorithm Watch set up a repository of AI guidelines (167 documents) [10], while the Ethics Codes Collection—which is a repository of ethics codes from across contexts—collates 68 documents under 'Artificial Intelligence and Robotics' [11, 12] (both repository counts as of late January 2023). A further repository attempts to "[map] the ecosystem of guidelines, principles, codes of ethics, standards and regulation being put in place around artificial intelligence", mapping 18 high-level guidelines, 13 processes and checklists, 11 interactive and practical tools, 4 industry standards, 5 online courses, 4 newsletters, and 23 policies from 6 regions [13].

However, to date, these resources and guidelines, "primarily focused on principles—the 'what' of AI ethics (beneficence, non-maleficence, autonomy, justice and explicability)—rather than on practices, the 'how'" [14 p. 2141]. Guidelines thus tend to be broad, with little guidance regarding how they might be applied across different practices, actors, and locations, or over time [15], leading to them operating "at a maximum distance from the practices [they] actually seek to govern" [16 p. 9]. Concerningly, guidelines may reify the inevitability of AI's integration into our lives, framing the innovative potential of AI as an imperative, while rarely pointing to the possibility of not using AI [17], and thus framing the ethical imperative in terms of, "'better building' […] because no statement offers 'not building' as an alternative" [17 p. 2128]. These concerns have motivated calls for development of worked examples and "a common language" [14 p. 2160] for AI ethics and corresponding business models and practices [14, 18, 19].

There is thus a need for resources that support learning about ethics with respect to our practices, contexts, and range of perceived impacts [20–22]. That is, resources that help people learn to navigate their own activity and related responsibility, in particular contexts, and across values which may be perceived and prioritised differently in different settings. For example, exploration of how values of privacy and autonomy could be perceived differently across cultures [23], or concern for responsibility in implementing AI in education might act as a synonym for data governance issues for institutional leadership, and for challenges to an unequal education system by other stakeholders [24, 25]. This recognition of the significance of context or setting for understanding cases reflects that our decisions at a micro-level come into interaction with macro-contexts in which and on which technologies act and interact [26] to produce both direct or hard impacts, and indirect or soft impacts [27].

2.1.1 Ethical concepts through cases: the need, and state of practice

One means to support people in navigating the practical issues of learning about, and how to apply, ethical concepts to practical contexts is through engagement with cases [1], that provide nuanced examples "for discussion, learning and analysis" [28 p. 2]. Cases provide opportunity to express the significance of ethics-in-action [29], and to connect our ethical concepts including values and practices to, "the routines of daily practice" [30 p. 939], and indeed the positionality of those engaged in those routines [31].

There is some evidence that cases are used in wider learning practice [32–34]. However, it is not clear *what* cases are used, or *how* they are presented and investigated. Specifically, there is some concern that case use reflects parallel concerns regarding operationalisation of principles, in that use of cases mostly targets computer-science students, through a small sample of cases materials demonstrating harms, and largely focuses on technical 'solutions' to these harms [32–34]. Thus cases, and common case repositories (§7.3), may not provide the conceptual resource required to address our desirable features with respect to rich learning regarding ethical concepts, and their tensions (§7.2), and relations across case content (§7.3). Although cases exist, it is not clear if they provide information required to support individuals' learning about AI ethics.

2.2 Cases to promote ethical reasoning and understanding of ethical tensions and dilemmas

How do we use cases to help foreground tensions between ethical concepts and values, to understand their complexity and inherent trade-offs? A further feature of cases is that they should help us to navigate practical concrete tensions in our work, for example in issues such as fair allocation of teaching support or moderation of assessment grades [22, 35–37]. In practical cases, we are often faced with incommensurable aims, and thus decisions where no options can address all desired outcomes or principles, such as the common trade-off between free speech and privacy (or: what is the threshold for breaching someone's privacy/free speech?). 'Strategies' in such cases may involve applying 'solutions' (methods to apply rules or principles without addressing the underlying dilemma) while varying in the

degree to which specific information about the case feeds into the strategy [38]. Cases act as a stimulus, with real case studies having the advantage of being authentic, while hypothetical or counterfactual cases help us probe anticipations of the future and imagine alternatives [39]. Cases can provide a helpful method for learning about practical dilemmas, where abstracted theories are unlikely to support learning or help us directly solve such dilemmas in fields such as education [40], application of human rights principles such as the right to 'dignity' [41], and indeed the ethical imperatives to positive impact in our work, including in potential use of AI towards the Sustainable Development Goals (SDGs) [41].

Cases can provide a resource to support navigating issues, and there are repositories of candidate cases in AI [§2.3, and notably, 42, 43]. However, as reported elsewhere [30, 37], these cases tend to illustrate clear examples of maleficence, are relatively unnuanced in their controversiality, or/and typically focus on particular features of an incident that led to reporting, rather than the background context of the incident system. Cases can provide more nuanced insight into perspectives on issues, for example by unpacking perspectives on what 'error' has occurred in application of algorithms and AI we can navigate systemic versus more localised or one-off challenges [44, 45]. As Annany [44] notes, using a case analysis of an e-proctoring system that has a bias in facial recognition, different framings of this 'error' and 'solutions' foreground different features of the system. If the 'error' is framed in terms of dataset bias, the 'solution' is to remove that bias, and to frame the issue in terms of the representativeness of that initial dataset. What such framing fails to recognise is the potential harms in that response, i.e., that inclusions in data can be harmful, and that datasets may never be complete. Crucially, "it leaves little room to ask whether the system should exist at all" [44 p. 18], or what the ongoing harms—what Ehsan et al. [46] call 'algorithmic imprints'—may be, even after systems are removed.

2.2.1 Ethical perspectives and tensions through cases: the need, and state of practice

The rich approach to foreground perspectives and tensions in navigating AI ethics that we d47escribe, presents a challenge for "mainstream AI ethics" [47 p. 446], in part due to its composition of largely technically focused researchers whose "ethical methods resemble machine learning practices" [47 p. 446], in making predictions, seeking to generate patterns, and applying these to wider cases. This may lead to attempts to formalise ethical concepts, values and strategies (described, §7.1) in ways that do not reflect situated context. For example, one case—regarding a model for bail decisions—has generated "an eye-popping 21 separate definitions of fairness in the attempt to account for everything circulating through fairness and solidarity ethics" [47 p. 446]. For Brusseau, this reflects the complexity of ethical reasoning based exclusively on principles, and the need for 'disorientation', to understand that there are no clear-cut answers in these cases. Cases such as this make clear again the inadequacy of narrow construal of bias or fairness, calling for a "situated fairness where descriptions of how an algorithmic system is constituted and its consequences are grounded in accounts of people who live with it." [46 p. 1307]. Moreover, such examples reflect the potential of foregrounding how principles or concepts come into tension, and the significance of differing perspectives on these tensions in understanding the issues at stake.

2.3 Cases to provide lenses across the specific and general and the role of observatories

How then should we provide cases that can foreground key ethical concepts and their tensions, and help people to navigate issues across cases to support learning?

The normative case study has been proposed as one approach to support navigation of public values [48, and in education, 49]. These are case studies that aim to outline real situations and the values underpinning choices that might be made, through an assessment of the normative features of those cases. As Thacher [48] outlines, these cases can draw attention to thick ethical concepts, such as 'courage' or perhaps more salient in the case of AI, 'creepy'. What distinguishes thick ethical concepts like creepiness, is that they are both descriptive (they describe an ethical standard, perhaps of privacy and autonomy in this case), and evaluative (they ascribe a particular judgement in their application). When we say a technology is creepy, we are doing more than just describing the applicable ethical-legal principles of privacy or autonomy, we are also making a particular kind of evaluative judgement regarding the technology. The normative case study is intended to describe cases as a way to bring into focus the ways they are imbued with values, and in this way to help us probe, navigate, and perhaps change these values. Cases, then, can help us through their deliberate selection for challenging our existing assumptions, perhaps through similar cases (an analogical approach), and close attention to context. These normative case studies are real examples of "living dilemmas that polarise society" [49 p. 7], and provide opportunity for learning through their exposure of ethics concepts that help navigate the problem.

Repositories or registries of such cases have potential to help us learn, precisely because they can provide both detail about specific cases, but also tools to navigate across cases. In a report investigating public sector deployments of AI

that had been cancelled, and the background to these cancellations, the potential was highlighted of registries that contain details of where and how systems are being used, and any audits or other materials regarding them, to "*provide a centralised and verified space to hold [automated decision systems] information, allowing not just greater oversight of how, when and why these systems are deployed but also facilitate greater learning.*" [50 p. 13].

In considering how repositories might facilitate this navigation of relations among cases, a range of approaches have been reported. For example, Mclaren's [51] early work set out a model for mapping principles to cases, with structured indications of case relevance and principle use in navigating cases. Both Kitto and Knight [22] and Bjørgen et al. [52] describe structures that reflect both the principles at play and provide space for description of the core dilemmas or tensions in the case, while Scott and Yampolskiy [53] draw on risk models to classify AI failure cases.

Another model is for incident databases that capture real world cases of harms and near-misses, as exists in other industries, to support collective intelligence for "design, development, and deployment of intelligent systems" [43 p. 1]. Specifically, the AI Incident Database provides a, "systematized collection of incidents where intelligent systems have caused safety, fairness, or other real world problems" [43 p. 1], through a faceted search over archived incident reports. The intent of this database is to answer "the question, 'what can go wrong when someone deploys this system'?" [43 p. 1]. The systems aims to provide easy ways to make harms visible, and provides a system architecture that allows for multiple taxonomies to be developed and applied by different users, with the intent to "*express the full range of viewpoints represented by the [Partnership on AI] partner community by allowing partners to build their own taxonomies, taxonomy documentation, and data summaries […] This avoids the challenge of developing a single shared universal ontology for AI incidents and instead allows for multiple viewpoints on the data to develop and compete for mindshare*." [43 p. 3].

2.3.1 Ethical relations through cases and their repositories: the need, and state of practice

In work [54] seeking to develop a taxonomy of ethics issues, 150 practical cases reported in the AI Incident Database were reviewed with respect to issues, and the principles and themes from published AI guidelines [specifically, those presented by 55, see §2.1]. However, as Wei and Zhou [54] highlight, there is not enough clarity regarding how to apply the principles into practice, with the incident database—and their own analysis—providing only some insight into what happens when guidelines are *violated* (i.e., failures of some variety, rather than issues arising via intended operation, see also §7.2) [54]. In addition, while the rationale for aligning the incidents with principles is clear, this approach, and potential bias in the sourcing of incidents, may create a bias in emphasising particular kinds of techno-centric issues, over wider social concerns.

Aliman et al. [56] seek to address this concern in a call for an alternative form of Transdisciplinary AI Observatory. They do this through a taxonomy-based observatory that both (1) analyses cases with respect to instantiated risks (i.e., historic), while also (2) analysing counterfactuals (i.e., a prospective, what could plausibly have happened), both those events that would have resulted in better (upward), and worse (downward) counterfactuals. This approach aims to develop future-oriented regulation, while also drawing attention to the usefulness of connections between cases, where these connections may assist in analogical reasoning for upward and downward counterfactuals. However, this call has not yet been met, nor is it clear how existing repositories might contribute to, or address, such a repository.

3 Methods

3.1 Review approach

A systematic approach akin to a scoping review was taken to identifying repositories, and cases of relevance within them. This model is appropriate for our context because the intent is not to systematically gather or assess all evidence on a topic, but rather to understand the commonly available resources and the nature of different cases and repositories available. The approach adopted allowed us to recognise the grounding in our existing practice and knowledge, while taking a systematic approach to identify additional repositories and cases. It also recognises that the different repositories have different features which makes a standardised approach to, for example, search, challenging to develop. To structure our reporting, and support replication we provide a modified PRISMA flow diagram, and adapt checklist reporting items for a scoping review [57–59]; the review was not pre-registered.

3.2 Search strategy

3.2.1 Overarching approach

3.2.1.1 Eligibility criteria overview Our primary focus is grey literature case repositories in which (1) there are multiple cases that are navigable (i.e., not compiled into a single PDF); and (2) the intended aim is to convey features of the impact of AI relating to ethical implications, i.e., relating

to consequential impact (positive or negative), ethical concepts, or other normative features. In some cases, repositories indicated they contained cases, but these were either found to be (a) examples or registers of use, but without any content regarding consequential impact (positive or negative); or (b) examples of regulatory context or implementation, but without discussion of clear cases and the connection of normative (i.e., values) to particular (i.e., case details and specific implications).

3.2.1.2 Search strategy The repository search was informed by our prior knowledge of a number of case repositories. In addition, repositories were identified through their use of a shared codebase, presence in a list of case repositories (or registries), and through screening of search snippets from a regular google search. This approach is not intended to be exhaustive, and we recognise that the choice of terms and targeting of particular forms of representation (e.g. in a website database, not a compiled PDF) is likely to have an impact on results identified.

3.2.1.3 Quality appraisal No quality assessment was made of sources retrieved.

3.2.2 Venues/sources

Sources were identified through:

1. A review of our teaching materials on the topic, a non-systematic search in conducting a narrative review regarding cases, and through use of shared code-bases in the repositories
2. A review of search snippets from a regular Google search
3. A review of the list at https://www.aiethicist.org/ethics-cases-registries
4. Link for- and backward chasing.

To sample cases for addressing RQ2, within each identified repository, a separate search was conducted for all cases regarding 'education' or related; in many repositories the cases were tagged by field, and thus the nearest cognate profession or discipline was selected for filtering.

3.2.3 Inclusion and exclusion criteria

In identifying repositories:

Inclusion criteria:

1. Contained multiple cases, relating to AI (broadly construed)
2. Presented consequential elements of AI (positive or negative impacts)
3. Provided a mechanism (e.g., links) to navigate between cases

Excluded if they:

1. Did not discuss consequential aspects of AI (e.g., were examples of use, without consideration of impacts)
2. Were not presented as 'cases' (e.g., discussed regulation or strategies, but without discussing application to specific contexts)
3. Were not presented in a structured way (e.g., news articles, single documents with discussions of examples but without a clear structured navigation)
4. Was not available at the point of search, a follow-up visit to the site, or via search for an archival copy

Non-English language repositories/cases were filtered (see results), but included in our initial searches. A subset of cases was extracted from the repositories for RQ2, with the initial inclusion criterion "tagged 'education'" or proximal; on investigation some such cases were not clearly related to education or learning, and were thus excluded from further analysis for RQ2.

3.2.4 Searches conducted

The following process was adopted, depicted in the modified PRISMA Flow diagram [60], Fig. 1.

1. The resources described in Venues/sources were reviewed for relevance
2. A Google query was conducted using the string: "artificial intelligence" AND ("incident database" OR "observatory" OR "catalogue" OR "repository" OR "gallery") AND (ethic* OR impact) AND (cases OR incidents OR reports OR "case studies"). This query returned 165 results (note, on the first page the search reports "About 9,170,000 results", clicking through to later pages reveals this is not the case). All snippets from the 165 results were screened for references to repositories or lists of case studies of AI impact and ethics.
3. Relevant links from both these web resources, and scholarly pieces we were aware of, were reviewed for links to repositories.

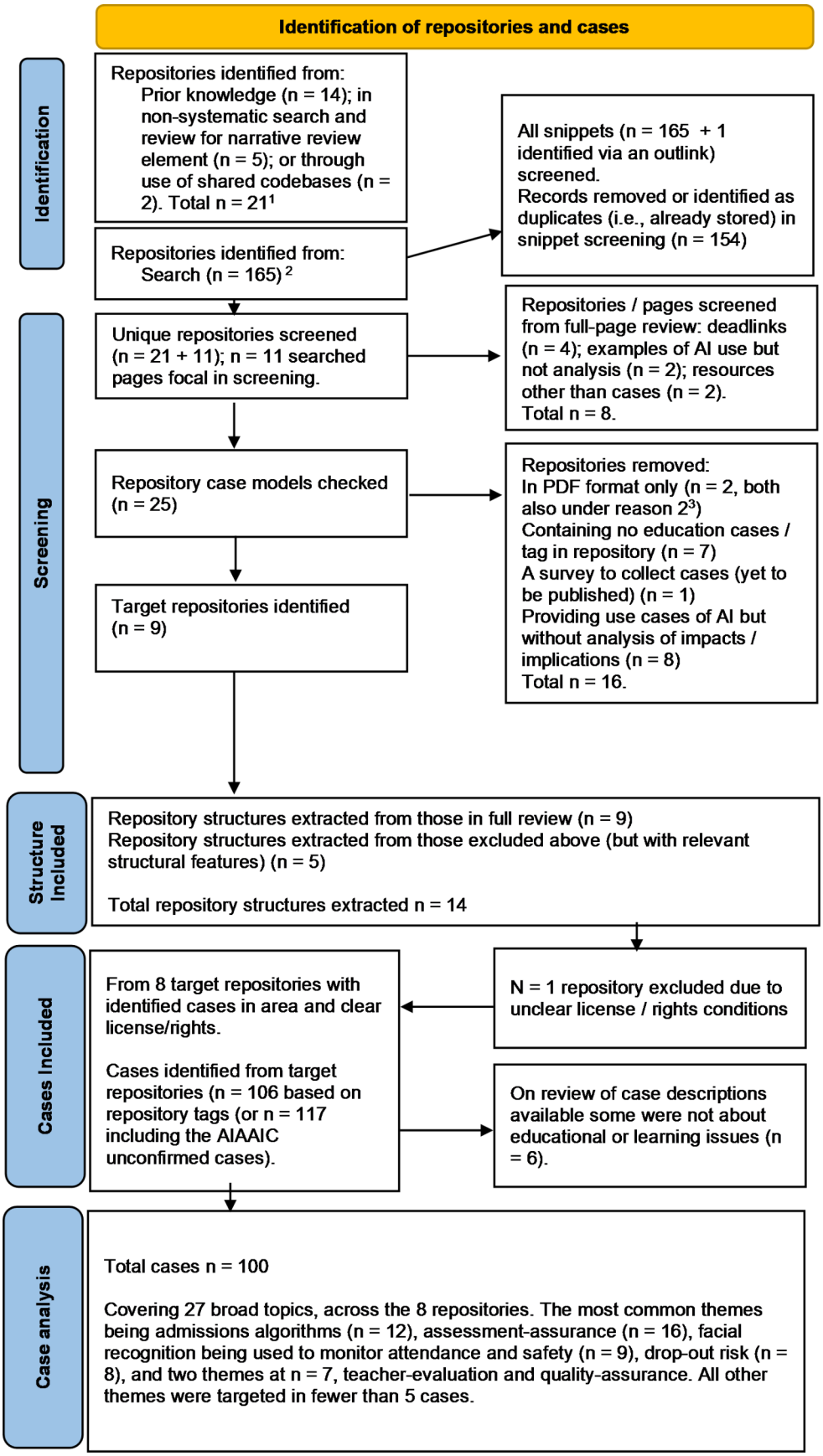

Fig. 1 Modified PRISMA Flow showing identification of AI ethics repositories (top) and cases (bottom)

3.3 Data management and processing

A procedure (see Supplement: Data management procedure) was developed to search, track, manage, and annotate items, largely within the open source Zotero reference manager [61]. Items were created for all repositories identified, with their structure or schema extracted. Some repositories explicitly provide this structural information, (e.g. the incidentdatabase.ai) while in others it was inferred through analysis of (1) the site search/browse functions; (2) the case structures presented, and (3) in some cases, investigation of either the page html, or the website codebase (e.g. ai-observatory.in's github contains its case schema). To address RQ2, and sample cases for deeper analysis, items were created for cases relating to education, using tags or filters from the repository to identify this sample.

Analysis of this extracted data drew on conventional content analysis in two forms, in both cases primarily inductive approach [62]. One analysis was of the characteristics of the repository metadata, the other of the case materials themselves, with both aiming to provide a heuristic overview of the ways the repositories make ethics concepts available for use by readers.

To address RQ1, regarding structures addressing the Ethics-case-CPR elements, data charting drew on repository metadata, with the repository schema used to create standardised structure labels and their possible values, and the themes addressed by cases in repositories. The following themes were identified through an inductive approach: descriptive (i.e., describing the factual features of the case), technology ownership-focus (i.e., describing who owned the technology or its use context), issue-focus (i.e., describing issues—positive or negative—relating to the case or its emergence as an incident), scale-focus (i.e., focusing on the scale of use or impact), use-focus (i.e., focusing on the context of use), technology nature focus (i.e., focusing on the underlying technology). Notably, for the purpose of our analysis, the Theme 'issues' captured a number of facets that directly capture ethical concerns (e.g., harms, 'AI ethics risk'), alongside those that indirectly captured expression of the 'issues' at stake (e.g., what the media trigger for the 'incident' was, which generally expressed features of harm).

To address RQ2, the sample cases (those relating to education) were analysed, using the case report in the repository alongside links to external sources. At least two researchers reviewed each case. A further subset of these cases was identified (relating to the sub-theme of teacher evaluation) for closer analysis. This theme was selected as the materials related to a number of incidents (rather than a single event reported in multiple locations), across countries and repositories. For this sub-theme, we (1) conducted an a priori analysis of the ethical concepts of relevance to the topic drawing on their expert knowledge, in which we tabulated which principles might be of relevance to the case topic with descriptive text (2) content analysis, in which we used the same principles in a close reading of the case material, to identify if and where the case materials identify these principles reviewed each case report in the topic (n = 2 reviewers for each case material) (3) conducted a comparative analysis, in which we first summatively assessed the materials against our a priori analysis (and vice-versa), and then collated the extracted case elements to develop a narrative comparison.

4 Results

4.1 Overview of repositories

Repositories, catalogues, observatories, galleries, registers or libraries of cases, were identified as described in the PRISMA flow diagram and §8.2.2 above.

From an initial check of n = 25 repositories, the structural elements of n = 14 (see Table 1) were extracted for further analysis (RQ1). To address RQ2, n = 8 repositories were analysed for their case contents (n = 100 relevant cases identified using each repository's own tagging for 'education' related cases).

Of the n = 8 repositories with cases in our target area, the 'about' pages were checked to ensure that the stated intent of the repositories was consistent with the aims we identified above regarding the sharing of cases to support learning (Table 2).

4.2 RQ1: How do case repositories represent the ethical concepts implicated in reported cases?

4.2.1 RQ1a: How do case repositories represent ethical values and principles?

Repository structures (fields or categories) that expressed 'issues' were extracted, to investigate the set of repository facets capturing the issues at stake—broadly construed—in each case. As noted in methods, this includes facets that, "*directly capture ethical concerns (e.g., harms, 'AI ethics risk'), alongside those that indirectly captured expression of the 'issues' at stake (e.g., what the media trigger for the 'incident' was, which generally expressed features of harm).*"

There are 10 repositories (71.43%) with coverage of issues (positive and negative outcomes and actions) as part of their structures, and 4 without (28.57%). For three of these ten with coverage of issues, that coverage is a significant emphasis; two of these are repositories (meeting

Table 1 Repositories in analysis

N	Repository	License	Locale	Launch	Updated
1	AI Observatory [63, 64]	cc-by	India	2019–20	Nov-21
2	AI Watch [65, 66]	Github repo lists The European Union Public Licence (EUPL)	Europe	2019	2021?
3	AIAAIC [42]	cc-by-sa	International (UK based)	2021	Feb-23
4	Algoritmos Publicos [67]	cc-by-sa	Chile	2021	unclear
5	AI Incident database [68, 69]	Github repo provides Apache License, V 2.0	International (USA based)	2019	Feb-23
6	The Observatory of Algorithms with Social Impact (OASI) [70]	cc-by	International (Spain based)	2021	Feb-23
7	ODImpact [71]	cc-by-sa	International (USA & UK based)	2016	2016
8	Fairlac [72]	cc-by	Latin America	Jul-20	Unclear
9	Data Scores [73, 74]				
10	CDEI and techUK AI Assurance Case Survey [75]				
11	African Observatory on Responsible AI [76]				
12	Awful-AI Repository [77, 78]				
13	Fujitsu *AI Ethics Impact Assessment Casebook* [79]				
14	Bias in AI: Examples Tracker [80]				

*Rows 9–14 are resources from which structures were extracted, but which were excluded from further analysis (see Method)

our inclusion criteria) while the third—the Fujitsu AI Ethics Impact Assessment Casebook—is the only structure to refer explicitly to ethics or ethical principles.

The original facets from each individual repository, and a set of facets derived into a composite list are available in the supplementary files [81]. Across these facets, while explicit ethical principles or descriptions of impacts with relation to ethics are not common, a set of possible harms and the broad space of positive impacts is a feature in at least some of the repositories, where these facets at times expressed concerns relating to ethics generally, or specific features including harms and transparency. Such facets include those addressing:

- Assurance of benefits, limitations in any assurance model, and principles (and evidence) underpinning any assurance
- Types of harm, intent to harm, likelihood of harm, and its severity
- Whether bias has been addressed and audits conducted
- Impacts on autonomy
- Populations impacted
- Legal context
- What brought the case to become an incident (e.g. legal intervention, media coverage, etc.).
- Positive impacts
- Transparency.

These features are reflected through a mix of controlled vocabulary lists (e.g., a list of possible population categories), binary classes (e.g., if an audit has been conducted or not), links to external resources, and free text. In some repositories it is not clear how data-values, particularly on controlled vocabularies, are selected (e.g., why a particular harm type or magnitude has been chosen as a label, rather than another).

4.2.2 RQ1b: How do case repositories represent ethical perspectives?

In the issues facets, facets relating to the issue of perspectives or relationships across cases included those describing stakeholders, including who implemented the system, who was targeted by the system, and who developed the system.[2] However, none of the databases appeared to indicate the perspective of this variety of stakeholders on the issues, with detailed discussion of those issues largely left to the external links. Nor are indications provided of how principles might come into tension, or the presence of dilemmas in cases in any repositories. This information could be provided through structural features such as semantic relationships between concepts (e.g., principle x is "in tension with" principle y), or/and through indications of different views on intended/actual outcomes.

As a result, in navigating cases it would be challenging to understand why different stakeholders might hold, or have

[2] A set of these focus on ownership: "Manner-of-Procurement, Implemented by/business owner, Implemented by/business owner, Developed For or Requested By, Developed For or Requested By, Public/private, Implemented by/business owner, Public/private, Financing, Region, Assurance-technique-recipient, Implemented by/business owner, Public/private, Stakeholders, Financing, Public/private"; a second group focus on the context of use including the stakeholders and domain into which the tool was being deployed: "Application-type, General domain/sector using Main government division (COFOG I level) class, Specific domain/sector using Government group (COFOG II level), Target-population, Stakeholders"; additionally, a third cluster centred on the AI or Model Type being used.

Table 2 Summary of repository purposes and creators

N	About (excerpts)	Who
1	Critique narratives of AI; document uses and their social, political and technical contexts; document actual and potential harms to individuals and groups; provide information for stakeholders to understand and mitigate harms	Independent India based research and design team with initial funding from Mozilla Foundation
2	Collates AI uses across public services and their geographic coverage, and social and policy uptake	EU funded team
3	Detail incidents and controversies relating to AI. Identify key risks and the: sectors, countries, and technologies most exposed to incidents, and the form of these incidents, their public triggers	AIAAIC is an independent, non-partisan, public interest initiative […] AIAAIC was founded in 2019 by Charlie Pownall. […] An RSA Fellow, he is a member of the European Commission's European AI Alliance and digital rights advocacy organisation the Open Rights Group. In former lives, Charlie was an EU official, speechwriter, journalist, and reputation risk and communications consultant
4	Make AI visible in Chilean public sector to encourage innovation	Chilean researchers
5	Index harms/near misses related to AI in order to mitigate the risks of these	Responsible AI Collaborative, an organization chartered to advance the AI Incident Database. The governance of the Collaborative is architected around the participation in its impact programming. […] Funded by partnership on AI, mix of corporate and academic directors
6	Collate AI used by government and companies globally to understand the risks and challenges they pose	Eticas Foundation is a nonprofit associated with the Eticas Group. We promote research, forecasting, awareness, advocacy and training on the interaction between technology, data and society. Based in Spain
7	Collate impact of open data for actionable insight by policy makers	Under the leadership of Andrew Young and Stefaan Verhulst, and in close collaboration with Laura Bacon of Omidyar Network, The Global Impact of Open Data initiative saw the development of 19 case studies of open data projects launched around the world, and a key findings paper articulating key findings across the case studies. In addition, Becky Hogge contributed six case studies focused on the United Kingdom
8	Map initiatives in Latin America relating to AI and impact in social policy	"diverse network of professionals and experts who want to promote an ethical application of AI in Latin America and the Caribbean from academia, government, civil society, industry, and the entrepreneurial sector"

held, different perspectives on issues, or what underlying tensions there may be between ethical principles for which stakeholders may hold different priorities.

4.2.3 RQ1c: How do case repositories represent ethical relations?

Based on analysis of the repository structures, two fields—*links* and *relations*—were identified as pertaining to relationships or perspectives on cases, although these fields could also be used simply to provide further detail on the cases rather than to facilitate cross-case inferential reasoning. Of the repositories, 12 included links to further detail or related cases. That is 11 contained links, 4—relations, and 3—both. One of these (the incidentdatabase.ai) has the relationship 'variant': "*an incident that shares the same causative factors, produces similar harms, and involves the same intelligent systems as a known AI incident*", as well as 'similar incidents' identified via textual similarity. This repository (incidentdatabase.ai), with variants, similar incidents, and sophisticated and navigable taxonomy, includes the most structural features to represent ethical relations, although these are largely left implicit.

Across features related to ethical relations, there were no identified structural features to identify how cases might differ or be similar in important ways. For example, none contain features that would help counterfactual analogical reasoning, i.e., to consider "what went wrong" in one case, by understanding features (positive or negative) of another case or its context. Some features could be used to infer such relations. For example, the AIAAIC's "triggers", indicates the trigger for a case becoming an incident (e.g., media coverage, or a legal intervention), also flagging related "research, audits, investigations, inquiries, litigation".

These features could be used to consider different contextual features of related cases, although such inferences are non-trivial.

4.3 RQ2: What ethical concepts are reflected in the subset of cases focused on educational applications?

4.3.1 Overview of themes in coverage of AI in educational contexts

To address RQ2, we sampled cases (see 3.3) in order to undertake closer analysis within a coverage of a particular domain. N = 2000 cases were identified across eight repositories included. Repositories each categorised cases by domain, with filters or text queries indicating the education sector. Of the total number of cases, n = 111 (5.55%) of cases were in the education category, ranging from 0.89% of cases (incidentdatabase.ai) to 32.08% (fairlac.iadb.org).

Table 3 Overview of Education case foci and broad area of application and impact within the area of education

Case focus	Application/impact	n
Assessment-assurance	Outcome	15
Admissions-algorithm	Outcome	12
Facial-recognition-for-dropout-or-attendance-and-safety-monitoring	Process-proxy	9
Dropout-risk	Process-proxy	8
Quality-assurance-targeting	Process-proxy	7
Teacher-evaluation	Process-proxy	7
Outcome-prediction	Outcome	5
Biometric-identity-checks	Administrative	4
Automated-assessment-models	Outcome	4
Engagement-emotion-classification	Process-proxy	4
Jobs-or-course-pathways-or-recommenders	Outcome	3
Risk-surveillance-wellbeing-and-violence	Societal	2
Wellbeing-mental-health-chatbot	Societal	2
Student-profiling	Process-proxy	2
Automated-child-linked-data	Administrative	2
Q&A-chatbot	Administrative	2
Real-time-measurement-and-presentation-to-support-user-interpretation-and-learning	Learning	2
Mining-internal-knowledge-base	Administrative	2
Adaptive-learning	Learning	1
Unemployment-risk	Societal	1
Deepfake-based-legalcase	Societal	1
Curriculum-pedagogy-management	Administrative	1
Weaponised-drones-for-security-USA	Societal	1
Price-targeting	Societal	1
Operational-efficiency-planning-eg-timetabling	Administrative	1
Survey-analysis	Administrative	1
Grand Total		100

Informed by the AIAAIC labels, these cases were categorised by their area of application within education. At this stage, six cases were excluded as not being about education or educational institutions (these were cases in which an incident had occurred at an education site, but the primary focus was something else, for example food delivery). This overview of areas addressed, summarised in Table 3, provides some insight into the scope of coverage in cases of issues in the application of AI to education and learning. Themes largely focus directly on *administrative processes* such as timetabling (n = 13), or on administrative aspects of teaching and learning including *outcome* indicators (n = 39) such as admissions algorithms, and *process-proxies* of learning (n = 37) such as teacher-evaluation based on data reporting, with some relating to broader societal concerns (n = 8) such as use of chatbots to support mental health. Few cases (n = 3) related to specialised pedagogic applications of AI such as use of adaptive-learning tools in teaching or dashboards for teacher insight into their learners, nor was there any clear coverage of the broader concerns around algorithmic imprints and errors discussed in the Background section.

4.3.2 Overview of values and principles in coverage of AI in educational contexts

To investigate how ethical concepts were instantiated in these cases at a finer grain, and drawing on our domain knowledge to investigate the specific application in education, we reviewed case material, which we exemplify through discussion of cases (n = 7) addressing the sub-theme *teacher evaluation* (see §8.3 rationale for theme selection). We first reviewed all collated case materials, using this review to develop analysis materials. Based on our initial review, we developed an analysis tool (ECAT, see supplement) to express the ways case topics might relate to a set of common ethical principles (synthesised from relevant recent resources).

As Table 4 indicates, we identified a connection to the range of ethics principles and concepts we analysed, while the case materials themselves reflected a subset of these, with no materials reflecting the full set. These differences are discussed in more detail in the next section.

4.3.3 Comparative analysis of ethical concepts in coverage of AI in educational contexts

Our comparative analysis of the case materials on the sub-theme (teacher evaluation) treated the compiled materials as the unit of analysis, to probe how the repositories—en masse—represented the ethical concepts at stake in the case topic. A detailed overview can be found in the supplementary

Table 4 Comparison of case material coverage of ethics concepts with our a priori analysis

Source	Equity, fairness and diversity	Human autonomy and dignity, including agency and accountability	Respect for democracy, justice, the rule of law, and environmental and societal wellbeing	Prevention of harm	Privacy and data governance	Technical robustness and safety	Transparency and explicability
1	X	X	X	–	–	X	X
2	X	X	X	X	–	X	X
3	–	X	–	X	–	–	X
4	–	X	X	–	X	X	–
5	–	–	X	–	–	–	X
6	X	X	–	X	–	–	X
7	X	–	–	X	–	–	X
Across case material	4/7	5/7	4/7	4/7	1/7	3/7	6/7
Issues identified in our a priori analysis	X	X	X	X	X	X	X

https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/washington-dc-schools-teacher-value-added-scoring;

https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/houston-isd-teacher-performance-evaluation-opacity

https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/sheri-g-lederman-nyc-teacher-effectiveness-assessment

https://incidentdatabase.ai/cite/9/

https://incidentdatabase.ai/cite/96/

https://airtable.com/appkg5eQBvy3zcosd/shrsAN2oTf68kM6O9/tblG2604tSoMOcwWX/viwoQmFjcPVR6VOSE/reczhhDlQoR4sMAWG?backgroundColor=teal&viewControls=on (links directly to "SAS| Algorithm to determine teachers' effectiveness")

https://airtable.com/appkg5eQBvy3zcosd/shrsAN2oTf68kM6O9/tblG2604tSoMOcwWX/viwoQmFjcPVR6VOSE/recabY8axKYDpZytu?backgroundColor=teal&viewControls=on (links directly to "Buona Scuola| Algorithm to determine allocation of school teachers")

files [81], here we will note key issues that were represented, and draw out central features that were not.

With respect to equity, fairness, and diversity, the general potential for unfairness in decisions made with automated evaluation (and lack of transparency in this regard, particularly with proprietary systems), was noted by materials. The materials did also note, although on checking only in one report, that the idea of using contextual indicators (sometimes called 'value added scores')—indicators that account for demographic characteristics of the students—is intended as a fairness measure, demonstrating one representation of our second desirable feature. The differing working conditions of teachers was mentioned with respect to challenges of particular schools, although not with respect to equity for teachers.

Human autonomy and dignity, including agency and accountability was reported largely with respect to teacher employment decisions made without recourse (a breach of dignity) and limits to teacher autonomy through pressure to teach to the test. Some limited discussion noted the issue of transparency in how scores were created, although these largely did not reflect on the potential of such information to support teacher professional development.

Respect for democracy, justice, the rule of law, and environmental and societal wellbeing was mostly discussed with respect to pending legal cases and concerns for due process in making employment decisions. Impacts of these tools on relationships between stakeholders were sometimes mentioned (e.g., negative impacts of systems without input from parent-and-learner-stakeholders who have positive views of a given teacher).

Prevention of harm was largely discussed with respect to individuals who had been negatively impacted. Privacy and data governance were mostly discussed with respect to making evaluations public, in such a way that teachers can be directly impacted. There was little other discussion of wider concerns for data governance. Technical robustness and safety were implied in discussion of teachers who received poor evaluations, with one report noting concerns regarding value added scores; this concern was mostly presented with respect to transparency and concerns about how robustness might be evaluated given poor transparency. The wider concern regarding evaluation of systems with respect to their purpose (i.e., if they actually support improvement of teaching quality) was absent. Transparency and explicability were significant features of coverage, including regarding concern about lack of transparency in use of proprietary algorithms.

In relation to our three desirable features for cases, then, while a range of ethical concepts are implied or directly addressed in the detail of case materials, these are very rarely found in the repository metadata itself, and often require navigating to external (generally news) websites. This makes it challenging to identify commonalities in cases, and to understand how systems might implicate particular ethical concerns. Across the case materials, reporting is largely framed in terms of the challenge or problem, i.e., a harm arising from teacher evaluation systems. Ironically, this can

make it hard to understand nuance in the cases, because initial rationales for the introduction of the system are not explored in depth, nor are different perspectives provided (news coverage often focuses on impacted individuals). That is, the perspectives and tensions at play are not well addressed. Finally, we saw no discussion across the materials analysed of relationships between cases, either across the different instances of teacher evaluation and what these multiple cases might inform us of, nor analogical examples of similar applications or tools applied in other contexts, or indeed the counterfactual—examples of places that have not implemented a tool, and how that has worked for them.

5 Discussion

5.1 Findings of the review

Our analysis of the structure of case repositories reporting on AI incidents indicates that ethical concepts are rarely directly addressed using explicit ethics language and concepts, addressing RQ1, i.e., whether case repositories represent the Ethics-case-CPR elements implicated in reported cases. Instead, ethical considerations are more frequently implied through discussion using concepts that relate to the consequences or impacts of AI and related incidents. In some repositories, we found neither direct nor indirect coverage of ethics. More universally, while differing perspectives or views may be present in some of the external links in repositories, these were not represented in the repository metadata itself, obscuring this important concern from view. Similarly, although links between cases were provided, these were generally instances of multiple reports of the same incident, or closely related incidents, without detail required to engage in analogical reasoning.

Our analysis indicates at a metadata level, a similar narrowness in focus, with relatively few cases investigating learning or pedagogic concerns, instead focusing on administrative issues, addressing RQ2 asking about what ethical concepts are reflected in illustrative cases focused on educational applications. Close analysis of the cases indicates a deeper engagement with ethics in some cases, particularly in externally linked resources such as new media, including AI-powered technology. However, here too, there are gaps with a relatively descriptive analysis provided, focused on individual incidents without connections across them or resources to explore the range of perspectives on those incidents. These cases did identify connections across incidents, but often in relation to a particular technocentric focus (e.g., a particular algorithm), rather than wider analysis of, for example, the counterfactual issue "what does this situation look like in places that have not implemented that tool this way?".

Overall, the narrow focus and descriptive nature of the cases limit their value for understanding and navigating ethical dilemmas in applications of AI. In the context of cases on education applications of AI, learning regarding the ethical issues at stake would require for example, a broader exploration of ethical issues (i.e. a shift from administrative concerns to include pedagogic and learning-focused applications and ethical issues) and their associated concepts (e.g., principles of fairness applied to these issues), multi-perspective expressions (i.e. providing resources and analyses that explore various stakeholder perspectives and the implications of ethical decisions), and linking of incidents (i.e. building connections across incidents to facilitate a deeper understanding of the ethical landscape),.

Across our analysis, the structures of repositories and their expression into cases are unlikely to foreground: key ethical concepts and their application to the substantive issues of AI; different perspectives on challenging dilemmic cases; and the ways that reasoning in one case may apply to another. This is a significant limitation that relates to learning towards ethical reasoning in the navigation of issues and their associated ethical concepts, and application to specific contexts of use of AI. These desirable features for cases are important features in learning to navigate AI ethics, both because engagement with the application of ethical concepts to practical cases is a means through which to learn about ethics [1, 82], and because cases have the potential to help people learn about AI ethics and the implications of use—or absence—of any particular AI tool in aspects of their day to day lives. Repository design may be developed to better support the application of ethical concepts to real-world issues, including the different perspectives stakeholders may have and the tensional between ethical principles.

5.2 Review scope and limitations

This paper set out to analyse repositories of cases. While providing a salient scope to the work, by their nature these repositories are varied in focus and content, not always easy to discover, and provide a limited set of cases reported. Other approaches to analyse would include content analysis of cases in the popular media, or self-report approaches with a range of stakeholders.

Within the repositories, the content is similarly constrained. The most common external source in the repositories is news media reporting, but of course a limited segment of concerns regarding AI is likely to be newsworthy. Compounding this, what gets transferred into the repositories is also mediated by those who run the repositories. This is of course also true of our own analysis, particularly

in our closer analysis of education cases; our interpretation of issues at stake, and our reading of the case material is informed by our position, a lens that both resources our analysis but may also constrain it in meaningful ways. These considerations are important, but not part of our analysis (nor part of the analysis in the repositories themselves).

However, the explicit intent of these repositories is to address the kinds of needs identified in the introduction regarding application of ethical concepts to issues in AI use and their navigation. Our findings indicate that the ways in which our core desirable features are operationalised in the repository structures and case reporting is generally limited. This suggests that structural changes could facilitate expression of ethical concerns in cases, to support learning about ethics. Learning activities that use these repositories [see, discussion of, e.g., 32–34, in §2.1.1], should take these limitations into account. Practitioners, including educators, should consider how to bridge the gaps underscored in our findings to enable effective uses of these repositories in supporting ethical understanding and decision-making.

5.3 Conclusions and implications

The ways cases are expressed can provide the conceptual resources to support learning about the ethical concepts at stake in applications of AI. While a range of useful repositories have emerged that collate examples of AI applications, focusing on both their positive and negative impacts, they may hitherto not be structured to support such learning.

The findings of this study demonstrate that while the expression of cases in repositories has the potential to support learning for ethical reasoning, current structures are often not designed in ways that lend themselves to such learning. Our analysis offers:

1. A Model for Repositories (Ethics-case-CPR): Grounded in our conceptual analysis, this model can be adopted to better represent ethical concerns.
2. Empirical Review: An analysis of existing repositories, their metadata, and example cases, providing an overview of variation in these resources and insight into the range of issues addressed by them.
3. Rationale for Adoption of our model: Based on both conceptual analysis and empirical review that demonstrates gaps, providing a strong case for restructuring repositories.
4. Analysis Tools: Tools to support experts in analysing and developing ethics cases.

Based on the resources developed, and analysis conducted, we further contribute an overarching provocation: To learn how to ethically (dis)engage with AI in its inception, design, implementation, and evaluation, stakeholders must learn about features of its ethical implications—this critical role of learning in consideration of ethical (dis)engagement with AI requires greater attention. Grounded in this claim, we suggest that applying a learning lens to resources created to promote awareness, register uses, or otherwise act as observatories motivates consideration of the features of such resources with respect to their potential to support learning. Existing case repositories have gaps, and this may shape what can be learnt from them. We propose Ethics-case-CPR is a useful overarching frame. Repositories might further consider additions to their schema to reflect these features (drawn from discussion §7.3).

1. What principles, ethical concepts, or legal issues are at play in the targeted context and in other contexts where the issue may occur?
2. What are the dilemmas, tensions, or predicaments at play?
    a. Why are these important in this case? How are they in tension, and is there consensus regarding that tension?
    b. What do different stakeholders think about how these values apply?
    c. What are the consequences of emphasising one or other set of values?
3. What other cases relate to this issue? Perhaps including:
    a. Similar populations impacted
    b. Similar impacts observed
    c. Similar resources—including ethical guidelines or discussion of ethical concepts, strategies, or other materials—that have, or might help, in navigating the issue
    d. Contra-applications observed, i.e., cases in which an application was made, but with different outcomes (or, in which an application was explicitly not made)

Addressing this provocation may assist repositories in enhancing their role in learning for ethical reasoning, helping stakeholders navigate the complex ethical landscape of AI applications.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s43681-024-00639-8.

Author contributions S.K., led all aspects of the work in collaboration with co-authors. Further contribution: Conceptualization: S.K., C.M., O.V. and T.C.; Data curation: S.K.; Formal analysis: S.K., C.M., O.V. and T.C.; Funding acquisition: S.K.; Investigation: S.K.; Methodology: S.K.; Project administration: S.K.; Resources: S.K.; Software:

S.K.; Validation: S.K., O.V. and T.C.; Writing—original draft: S.K., C.M., O.V. and T.C.; Writing—review & editing: S.K., C.M., O.V. and T.C. The Institute of Physiology CRediT generator was used to generate this list (https://www.fgu.cas.cz/en/articles/833-credit-generator).

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. The work was begun over the course of the lead author's sabbatical, which included periods at the University College London (UCL) Knowledge Lab and at KTH/the Swedish Digital Futures research centre. Digital Futures provided financial and intellectual support for the lead author's residency, as did the University of Technology Sydney. The lead author receives Australian Research Council funding (DE230100065, and DP240100602) and funding from the James Martin Institute (JMI), the latter two related to the ethics of AI, and the former to learning in the context of socioscientific issues. The views expressed herein are those of the authors and are not necessarily those of the Australian Government or Australian Research Council.

Data availability The author confirms that all data generated or analysed during this study are included in this published article, or in the supplementary materials [81]. All supplementary materials available via the data publication at Knight, et al., (2024). Knight, S., McGrath, C., Viberg, O., & Cerratto Pargman, T. (2024). Supplements to: Learning about AI ethics from cases: A scoping review of AI incident repositories and cases. figshare. https://doi.org/10.6084/m9.figshare.26420758.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. O'Mathúna, D., Iphofen, R.: Making a case for the case: an introduction. In: O'Mathúna, D., Iphofen, R. (eds.) Ethics, integrity and policymaking: the value of the case study, pp. 1–12. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-15746-2_1
2. Ortega-Bolaños, R., Bernal-Salcedo, J., Germán Ortiz, M., Galeano Sarmiento, J., Ruz, G.A., Tabares-Soto, R.: Applying the ethics of AI: a systematic review of tools for developing and assessing AI-based systems. Artif. Intell. Rev. **57**, 110 (2024). https://doi.org/10.1007/s10462-024-10740-3
3. Stahl, B.C., Antoniou, J., Bhalla, N., Brooks, L., Jansen, P., Lindqvist, B., Kirichenko, A., Marchal, S., Rodrigues, R., Santiago, N., Warso, Z., Wright, D.: A systematic review of artificial intelligence impact assessments. Artif. Intell. Rev. **56**, 12799–12831 (2023). https://doi.org/10.1007/s10462-023-10420-8
4. Corrêa, N.K., Galvão, C., Santos, J.W., Del Pino, C., Pinto, E.P., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., Terem, E., de Oliveira, N.: Worldwide AI ethics: a review of 200 guidelines and recommendations for AI governance. Patterns. **4**, 100857 (2023). https://doi.org/10.1016/j.patter.2023.100857
5. Khan, A.A., Badshah, S., Liang, P., Waseem, M., Khan, B., Ahmad, A., Fahmideh, M., Niazi, M., Akbar, M.A.: Ethics of AI: A Systematic Literature Review of Principles and Challenges. In: The International conference on evaluation and assessment in software engineering 2022. pp. 383–392. ACM, Gothenburg Sweden (2022). https://doi.org/10.1145/3530019.3531329.
6. Schiff, D.: AI Ethics Global Document Collection, https://ieee-dataport.org/open-access/ai-ethics-global-document-collection-0, (2020).
7. Schiff, D., Borenstein, J., Biddle, J., Laas, K.: AI ethics in the public, private, and NGO sectors: a review of a global document collection. IEEE Trans. Technol. Soc. **2**, 31–42 (2021). https://doi.org/10.1109/TTS.2021.3052127
8. Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., Srikumar, M.: Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI. SSRN Electron J (2020). https://doi.org/10.2139/ssrn.3518482
9. Jobin, A., Ienca, M., Vayena, E.: Supplementary file the global landscape of AI ethics guidelines. Nat Mach Intell. **1**, 389–399 (2019). https://doi.org/10.1038/s42256-019-0088-2
10. Algorithm Watch: About: AI Ethics Guidelines Global Inventory, https://inventory.algorithmwatch.org/about. Accessed 26 Jan 2023
11 Laas, K., Davis, M., Hildt, E. (eds.): Codes of ethics and ethical guidelines: emerging technologies. Changing Fields. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-030-86201-5
12. Study of Ethics in the Professions: The Ethics Codes Collection, http://ethicscodescollection.org/search?topic=Artificial%20Intelligence%20and%20Roboticsfilter=true. Accessed 14 Jan 2023
13. EthicalML: Awesome AI Guidelines, https://github.com/EthicalML/awesome-artificial-intelligence-guidelines, (2023).
14. Morley, J., Floridi, L., Kinsey, L., Elhalal, A.: From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. Sci. Eng. Ethics **26**, 2141–2168 (2020). https://doi.org/10.1007/s11948-019-00165-5
15. Rességuier, A., Rodrigues, R.: AI ethics should not remain toothless! A call to bring back the teeth of ethics. Big Data Soc. **7**, 2053951720942541 (2020). https://doi.org/10.1177/2053951720942541
16. Hagendorff, T.: The ethics of AI ethics: an evaluation of guidelines. Mind. Mach. **30**, 99–120 (2020). https://doi.org/10.1007/s11023-020-09517-8
17. Greene, D., Hoffmann, A.L., Stark, L.: Better, nicer, clearer, fairer: a critical assessment of the movement for ethical artificial intelligence and machine learning. (2019).
18. Gill, K.S.: Ethical encounters. AI Soc. **36**, 1–7 (2021). https://doi.org/10.1007/s00146-020-01135-3
19. Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M., Floridi, L.: Operationalising AI ethics: barriers, enablers and next steps. AI & Soc. (2021). https://doi.org/10.1007/s00146-021-01308-8
20. Gray, C.M., Boling, E.: Inscribing ethics and values in designs for learning: a problematic. Educ. Tech. Res. Dev. **64**, 969–1001 (2016). https://doi.org/10.1007/s11423-016-9478-x
21. Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S.B., Santos, O.C., Rodrigo, M.T., Cukurova, M., Bittencourt, I.I., Koedinger, K.R.: Ethics of AI in education: towards a community-wide framework. Int. J. Artif. Intell. Educ. (2021). https://doi.org/10.1007/s40593-021-00239-1

22. Kitto, K., Knight, S.: Practical ethics for building learning analytics. Br. J. Edu. Technol. **50**, 2855–2870 (2019). https://doi.org/10.1111/bjet.12868
23. Viberg, O., Jivet, I., Scheffel, M.: Designing culturally aware learning analytics: a value sensitive perspective, http://arxiv.org/abs/2212.09645, (2022).
24. McGrath, C., CerrattoPargman, T., Juth, N., Palmgren, P.J.: University teachers' perceptions of responsibility and artificial intelligence in higher education—an experimental philosophical study. Comput. Educ. Artif. Intell. **4**, 100139 (2023). https://doi.org/10.1016/j.caeai.2023.100139
25 CerrattoPargman, T., McGrath, C., Viberg, O., Knight, S.: New vistas on responsible learning analytics: a data feminism perspective. J. Learn. Anal. (2023). https://doi.org/10.18608/jla.2023.7781
26. Johnson, D.G., Verdicchio, M.: The sociotechnical entanglement of AI and values. AI & Soc. (2024). https://doi.org/10.1007/s00146-023-01852-5
27 Swierstra, T.: Identifying the normative challenges posed by technology's 'soft' impacts. Etikk Praksis. Nord. J. Appl. Ethics **1**, 5–20 (2015). https://doi.org/10.5324/EIP.V9I1.1838
28 Stahl, B.C., Schroeder, D., Rodrigues, R.: The ethics of artificial intelligence: an introduction. In: Stahl, B.C., Schroeder, D., Rodrigues, R. (eds.) Ethics of artificial intelligence: case studies and options for addressing ethical challenges, pp. 1–7. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-17040-9_1
29. Guillemin, M., Gillam, L.: Ethics, reflexivity, and "ethically important moments" in research. Qual. Inq. **10**, 261–280 (2004). https://doi.org/10.1177/1077800403262360
30. Bezuidenhout, L., Ratti, E.: What does it mean to embed ethics in data science? An integrative approach based on microethics and virtues. AI & Soc. **36**, 939–953 (2021). https://doi.org/10.1007/s00146-020-01112-w
31 D'Ignazio, C., Klein, L.F.: Data feminism. MIT Press, Cambridge (2020)
32. Fiesler, C., Garrett, N., Beard, N.: What do we teach when we teach tech ethics? A syllabi analysis. In: Proceedings of the 51st ACM technical symposium on computer science education. pp. 289–295. Association for Computing Machinery, New York, NY (2020). https://doi.org/10.1145/3328778.3366825
33. Slavkovik, M.: Teaching AI Ethics: Observations and Challenges. Norsk IKT-konferanse for forskning og utdanning. (2020)
34. Tuovinen, L., Rohunen, A.: Teaching AI ethics to engineering students: reflections on syllabus design and teaching methods. In: Proceedings of the conference on technology ethics. CEUR (2021).
35 Forster, D., Maxwell, B.: Using codes of professional ethics and conduct in teacher education: pitfalls and best practice. In: Eaton, S.E., Khan, Z.R. (eds.) Ethics and integrity in teacher education, pp. 25–42. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16922-9_3
36. Gulson, K., Benn, C., Kitto, K., Knight, S., Swist, T.: Algorithms can decide your marks, your work prospects and your financial security. How do you know they're fair?, http://theconversation.com/algorithms-can-decide-your-marks-your-work-prospects-and-your-financial-security-how-do-you-know-theyre-fair-171590 (2021)
37. Knight, S., Shibani, A., Buckingham Shum, S.: A reflective design case of practical ethics in learning analytics. Br. J. Edu. Technol. (2023). https://doi.org/10.1111/bjet.13323
38. Robinson, P.: Moral disagreement and artificial intelligence. AI Soc. (2023). https://doi.org/10.1007/s00146-023-01697-y
39. Ekberg, M.E.: Exploring the design, delivery and content of a 'bioethics for the biosciences' module: an empirical study. J. Acad. Ethics **14**, 103–114 (2016). https://doi.org/10.1007/s10805-015-9246-2
40 Levinson, M., Fay, J.: Dilemmas of educational ethics: cases and commentaries. Harvard Education Press, Cambridge (2019)
41 Stahl, B.C., Schroeder, D., Rodrigues, R.: Dignity. In: Stahl, B.C., Schroeder, D., Rodrigues, R. (eds.) Ethics of artificial intelligence: case studies and options for addressing ethical challenges, pp. 79–93. Springer International Publishing, Cham (2023). https://doi.org/10.1007/978-3-031-17040-9_7
42. AIAAIC: AIAAIC—AI, algorithmic and automation incident and controversy repository, https://docs.google.com/spreadsheets/u/0/d/1Bn55B4xz21-_Rgdr8BBb2lt0n_4rzLGxFADMlVW0PYI/htmlview# (2021)
43. McGregor, S.: Preventing repeated real world AI failures by cataloging incidents: The AI incident database. https://doi.org/10.48550/arXiv.2011.08512 (2020)
44. Ananny, M.: Seeing like an algorithmic error: what are algorithmic mistakes, why do they matter, how might they be public problems? https://law.yale.edu/sites/default/files/area/center/isp/documents/ananny.pdf (2022)
45. Jones, K.M.L., Rubel, A., LeClere, E.: A matter of trust: higher education institutions as information fiduciaries in an age of educational data mining and learning analytics. J. Am. Soc. Inf. Sci. **71**, 1227–1241 (2020). https://doi.org/10.1002/asi.24327
46. Ehsan, U., Singh, R., Metcalf, J., Riedl, M.: The Algorithmic Imprint. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. pp. 1305–1317. Association for Computing Machinery, New York, NY (2022). https://doi.org/10.1145/3531146.3533186
47. Brusseau, J.: Using edge cases to disentangle fairness and solidarity in AI ethics. AI Ethics. **2**, 441–447 (2022). https://doi.org/10.1007/s43681-021-00090-z
48. Thacher, D.: The normative case study. Am. J. Sociol. **111**, 1631–1676 (2006). https://doi.org/10.1086/499913
49 Gurr, S.K., Forster, D.J.: Using normative case studies to examine ethical dilemmas for educators in an ecological crisis. Educ. Philos. Theory (2023). https://doi.org/10.1080/00131857.2023.2169128
50. Redden, J., Brand, J., Sander, I., Warne, H.: Automating public services: learning from cancelled systems. Collective Wellbeing, Carnegie UK; Data Justice Lab; Western FIMS (2022).
51. McLaren, B.M.: Extensionally defining principles and cases in ethics: An AI model. Artif. Intell. **150**, 145–181 (2003). https://doi.org/10.1016/S0004-3702(03)00135-8
52. Bjørgen, E.P., Madsen, S., Bjørknes, T.S., Heimsæter, F.V., Håvik, R., Linderud, M., Longberg, P.-N., Dennis, L.A., Slavkovik, M.: Cake, death, and trolleys: dilemmas as benchmarks of ethical decision-making. In: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society. pp. 23–29. Association for Computing Machinery, New York, NY (2018). https://doi.org/10.1145/3278721.3278767
53. Scott, P.J., Yampolskiy, R.V.: Classification schemas for artificial intelligence failures. https://doi.org/10.48550/arXiv.1907.07771 (2019)
54. Wei, M., Zhou, Z.: AI ethics issues in real world: evidence from AI incident database. https://doi.org/10.48550/arXiv.2206.07635 (2022)
55. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nat. Mach. Intell. **1**, 389–399 (2019). https://doi.org/10.1038/s42256-019-0088-2
56. Aliman, N.-M., Kester, L., Yampolskiy, R.: Transdisciplinary AI observatory—retrospective analyses and future-oriented contradistinctions. Philosophies **6**, 6 (2021). https://doi.org/10.3390/philosophies6010006
57. Peters, M.D.J., Marnie, C., Tricco, A.C., Pollock, D., Munn, Z., Alexander, L., McInerney, P., Godfrey, C.M., Khalil, H.: Updated methodological guidance for the conduct of scoping reviews. JBI

Evid. Implement. **19**, 3 (2021). https://doi.org/10.1097/XEB.0000000000000277
58 Peters, M.D.J., Godfrey, C., McInerney, P., Khalil, H., Larsen, P., Marnie, C., Pollock, D., Tricco, A.C., Munn, Z.: Best practice guidance and reporting items for the development of scoping review protocols. JBI Evid. Synth. **20**, 953 (2022). https://doi.org/10.11124/JBIES-21-00242
59. Tricco, A.C., Lillie, E., Zarin, W., O'Brien, K.K., Colquhoun, H., Levac, D., Moher, D., Peters, M.D.J., Horsley, T., Weeks, L., Hempel, S., Akl, E.A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M.G., Garritty, C., Lewin, S., Godfrey, C.M., Macdonald, M.T., Langlois, E.V., Soares-Weiser, K., Moriarty, J., Clifford, T., Tunçalp, Ö., Straus, S.E.: PRISMA Extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann. Intern. Med. **169**, 467–473 (2018). https://doi.org/10.7326/M18-0850
60. Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., Moher, D.: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ (2021). https://doi.org/10.1136/bmj.n71
61. Center for History and New Media: Zotero Quick Start Guide, http://zotero.org/support/quick_start_guide.
62. Hsieh, H.-F., Shannon, S.E.: Three approaches to qualitative content analysis. Qual. Health Res. **15**, 1277–1288 (2005). https://doi.org/10.1177/1049732305276687
63. ai-observatory: ai-observatory/ai-observatory.github.io, https://github.com/ai-observatory/ai-observatory.github.io, (2021)
64. ai-observatory: ai-observatory, https://ai-observatory.in/database. Accessed 27 Jan 2023
65. European Commission Joint Research Centre: Organisation for Economic Co-operation and Development: AI watch, national strategies on artificial intelligence: a European perspective. Publications Office, LU (2021)
66. Vaccari, L., Perego, A.: Welcome to the AI Watch T6 Explorer!, https://github.com/AI-Watch/AI-watch-T6-X, (2022).
67. GobLab UAI: Algoritmos Públicos— GobLab UAI, http://www.algoritmospublicos.cl/. Accessed 27 Jan 2023.
68. Responsible AI Collaborative: Artificial Intelligence Incident Database (AIID), https://github.com/responsible-ai-collaborative/aiid, (2023).
69. Responsible AI Collaborative: Welcome to the Artificial Intelligence Incident Database, https://incidentdatabase.ai. Accessed 27 Jan 2023
70. Eticas Foundation: The Observatory of Algorithms with Social Impact—OASI—Eticas Foundation, https://eticasfoundation.org/oasi/. Accessed 27 Jan 2023
71. ODI: About Open Data's Impact Repository, http://odimpact.org. Accessed 27 Jan 2023
72. fAIrLAC: Observatory| fAIrLAC, https://fairlac.iadb.org/en/observatorio. Accessed 22 Feb 2023
73. Christo: Data Scores in the UK, https://github.com/critocrito/data-scores-in-the-uk, (2021)
74. Data Scores: Data Scores in the UK, https://data-scores.org/documents. Accessed 22 Feb 2023
75. CDEI, techUK: Centre for Data Ethics and Innovation (CDEI) and techUK: AI Assurance case studies, survey collection, https://www.surveymonkey.com/r/aiassurance?_cldee=FS8po3ln9-cJuQpaz0GyKGIBO8JTyWje0MR5-cR9MV-ShpuU-n4QFfkoirwjmhq6recipientid=contact-53fae05e4d45e811811b5065f38b5621-72283894730946148f6a25baa8eee283esid=30d8b6b2-814f-ed11-bba3-000d3adea432. Accessed 22 Feb 2023..
76. African Network on Responsible AI: African Observatory on Responsible Artificial Intelligence, https://www.africanobservatory.ai/resources. Accessed 21 Feb 2023.
77. Dao, D.: Awful AI, https://github.com/daviddao/awful-ai, (2023).
78. Dao, D., W, H., Walla, A.-A., VocalFan, Rubiel, E., Schreiber, F., Jaworski, J., Liu, L., Williams, N., Pawlowski, N., Ammanamanchi, P.S., Stadlmann, S., Kühne, S., Taiz, Diethe, T.: jvmncs, twsl: daviddao/awful-ai: Awful AI—2021 Edition, https://zenodo.org/record/5855972, https://doi.org/10.5281/zenodo.5855972 (2022)
79. Fujitsu: AI Ethics Impact Assessment Casebook, https://www.fujitsu.com/global/documents/about/research/technology/aiethics/fujitsu-AIethics-case_en.pdf (2022)
80. Berkley HAAS Center for Equity, Gender and Leadership: Bias in AI: Examples Tracker, https://docs.google.com/spreadsheets/d/1eyZZW7eZAfzlUMD8kSU30IPwshHS4ZBOyZXfEBiZum4/edit?usp=embed_facebook. Accessed 28 Jan 2023
81. Knight, S., McGrath, C., Viberg, O., Cerratto Pargman, T.: Supplements to: Learning about AI ethics from cases: a scoping review of AI incident repositories and cases, https://doi.org/10.6084/m9.figshare.26420758 (2024).
82 Stahl, B.C., Schroeder, D., Rodrigues, R.: Ethics of artificial intelligence: case studies and options for addressing ethical challenges. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-17040-9