Information Research - Vol. 30 No. iConf (2025)

Does artificial intelligence harm labour? Investigating the limitations of incident trackers as evidence for policymaking

Theodore Dreyfus Ledford

DOI: https://doi.org/10.47989/ir30iConf47296

Abstract

Introduction. From the point of view of public policy, artificial intelligence (AI) is an emerging technology with as-yet-unknown risks. AI incident trackers collect harms and risks to inform policymaking. We investigate how labour is represented in two popular AI incident trackers. Our goal is to understand how well the knowledge organization of these incident trackers reveals labour-related risks for AI in the workplace, with a focus on how AI is impacting and expected to impact workers within the United States.

Data and Analysis. We search for and analyse labour-related incidents in two AI incident trackers, the Organization for Economic Cooperation and Development's AI incidents monitor (OECD AIM) and the AI incident database (AIID) from the responsible AI collaborative.

Results. The OECD AIM database categorised workers as stakeholders for 600 incidents with 6,744 associated news reports. From the AIID, we constructed a set of 57 labour-related incidents.

Discussion and Conclusions. The AI incident trackers do not facilitate ready retrieval of labour-related incidents: they used limited existing labour-related terminology. AI incident trackers' reliance on news reports risks overrepresenting some sectors and depends on the news reports' framing of the evidence.

Introduction

To make regulatory decisions, policymakers need to define problems and priorities in a process called agenda setting (Baumgartner, 2015). Issues raised in the agenda setting phase inform the evidence to be collected from empirical research (Head, 2010; Pawson et al., 2011) to shape the space of policies envisioned (MacKillop & Downe, 2023; Schiff, 2024).

Artificial intelligence (AI), while long studied in the academic sphere, is, from the point of view of public policy, an emerging technology with as-yet-unknown risks. In the public consciousness, two vivid risks people envision from AI are existential threats to humanity (Cameron, 1984) and the risk of being replaced by machines (Autor, 2015; 2022). In the past few years, multiple groups have introduced AI incident trackers (Abercrombie et al., 2024; Hutiri et al., 2024; McGregor, 2021; OECD, 2024) and taxonomies (Arda, 2024; Cattell et al., 2024; Critch & Russell, 2023; Shelby et al., 2022; Weidinger, 2022) to analyse the potential harms and risks of AI.

In this paper, we focus on labour relations with AI (Arntz et al., 2016), sometimes called the 'future of work' (Laker, 2023). Specifically, we investigate how two popular AI incident trackers represent labour-related risks for AI in the workplace.

Background

AI and labour

Policymakers seek to mitigate the effects of emerging technology on labour displacement (Lane & Saint-Martin, 2021), deskilling and economic trade competition (OECD, 2022b). In 2020, US policymakers defined an AI as a '*machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments*' (ScienceIsUS, 2024). As early as the 20th century, governments and research experts forecasted the effects of automation on the workforce (Acemoglu and Restreppo, 2019). More recently, Frey & Osborne (2017) measured which occupations are at the most risk of automation.

Incident tracking

We identified a proliferation of AI incident trackers (e.g., Hutiri et al., 2024; McGregor, 2021; Rodrigues et al., 2023; Shrishak, 2023). These AI incident trackers are informed by earlier incident reporting strategies to address system failures and risks in aviation (NASA Aviation Safety Reporting System, n.d.; Reynard, 1986), healthcare (Kohn et al., 2000; Macrae, 2016), software development (Booth et al., 2013) and cybersecurity (van der Kleij et al., 2022). However, current AI trackers '*rely heavily on news coverage of AI incidents*' (Turri & Dzombak, 2023). There is no standard structure for incident tracking. Incidents begin as records of an event that are deemed worth reporting. Incident tracking depends on a predefined documentation procedure indicating what information may indicate failure or risk within a system and is '*worth*' knowing (Turri & Dzombak, 2023). Recording an event as an incident transforms it into labelled data. Experts standardise the vocabulary that guides how incident trackers can assess the implications of a technology or policy (Hoffmann & Frase, 2023; OECD, 2022a).

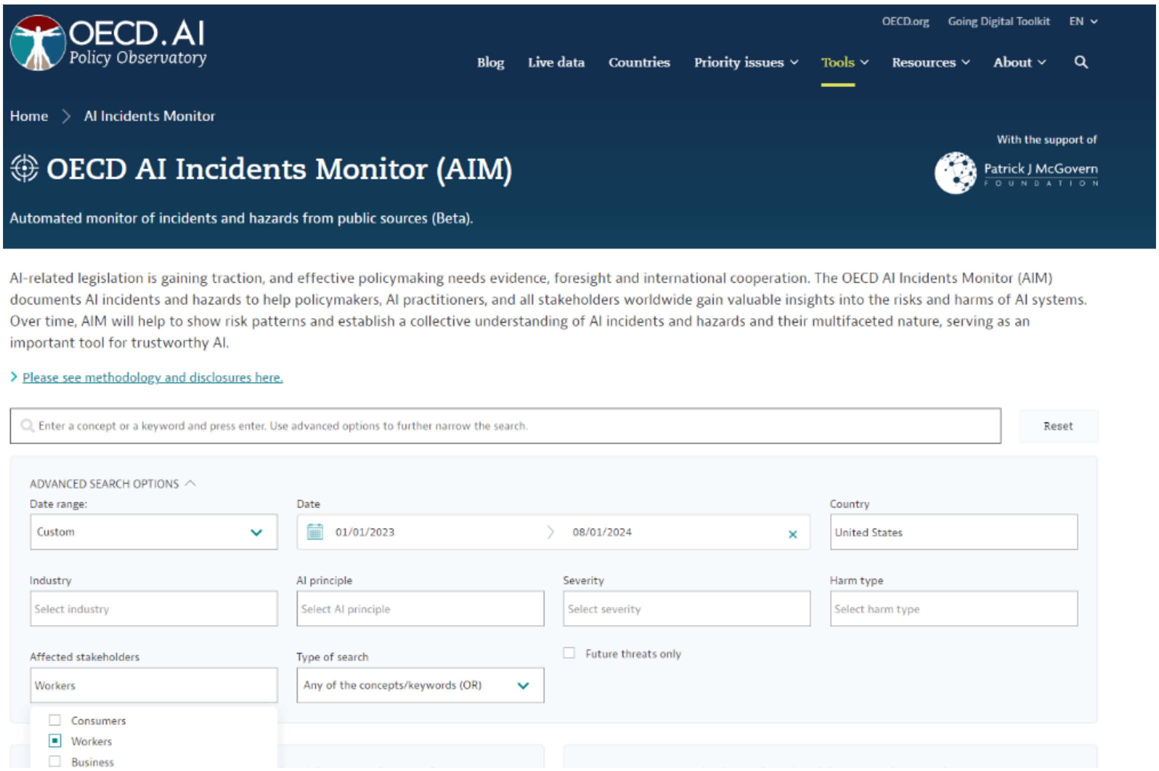
Aims

To determine how well AI incident trackers inform policymaking, we study how well the knowledge organization of different incident trackers reveals labour-related risks for AI in the workplace. We search for and analyse labour-related incidents in two AI incident trackers, the AI incidents monitor and the AI incident database, with a focus on how AI is impacting and is expected to impact workers within the United States.

Data collection and analysis

AI incident tracker 1: OECD AI incidents monitor (AIM)

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organization of 38 mostly high-income, industrialised countries. The OECD AI incidents monitor, (AIM) launched publicly in 2023. OECD AIM was developed as part of their efforts on AI governance to '*establish a knowledge foundation… and… terminology*' for an interactive evidence base that could help policymakers to define the scope of AI (OECD, 2023). The incident tracker's backend process is supported by the event registry (http://eventregistry.org) digital service, a third-party commercial entity. Event registry monitors news reports worldwide, drawing on an expert-created category system and machine learning algorithms to group news into incidents and automatically classify harms. For OECD AIM, each incident is automatically assigned a summary and headline from the primary news report (the news report from the source with the highest Alexa traffic rank).



OECD AI Incidents Monitor (AIM)

Automated monitor of incidents and hazards from public sources (Beta).

AI-related legislation is gaining traction, and effective policymaking needs evidence, foresight and international cooperation. The OECD AI Incidents Monitor (AIM) documents AI incidents and hazards to help policymakers, AI practitioners, and all stakeholders worldwide gain valuable insights into the risks and harms of AI systems. Over time, AIM will help to show risk patterns and establish a collective understanding of AI incidents and hazards and their multifaceted nature, serving as an important tool for trustworthy AI.

> Please see methodology and disclosures here.

Enter a concept or a keyword and press enter. Use advanced options to further narrow the search.

ADVANCED SEARCH OPTIONS

Date range: Custom

Date: 01/01/2023 > 08/01/2024

Country: United States

Industry: Select industry

AI principle: Select AI principle

Severity: Select severity

Harm type: Select harm type

Affected stakeholders: Workers

☐ Consumers
☑ Workers
☐ Business

Type of search: Any of the concepts/keywords (OR)

☐ Future threats only

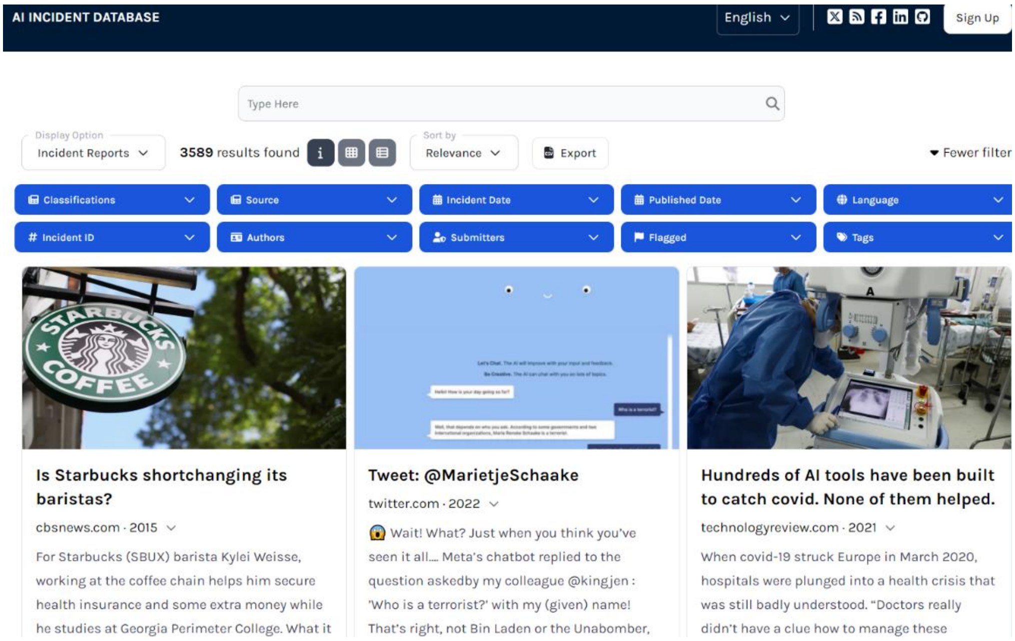
Figure 1. The OECD's AIM database query portal, showing some of our query settings

To collect data, we downloaded query results (incidents, summaries, headlines) from the OECD's AIM database (Figure 1) on August 1, 2024. We had to depend on the online portal because the most important data for our project, '*affected stakeholders*', was only available through this channel. (The batch data download option only provided the incident ID, title, summary, date of creation, concepts, and companies.) We always set '*affected stakeholders*' to workers, '*country*' to United States and '*date*' from January 1, 2023, through August 1, 2024.

Our queries retrieved 600 incidents with 6,744 associated news reports that the OECD AIM database categorised as having workers as '*affected stakeholders*'. We searched separately for each '*industry*' in the OECD AIM industry taxonomy and calculated the percentage of incidents with Workers as stakeholders. We searched with '*future threat only*' both unchecked and checked and calculated the percentage of future threats. We grouped industrial sectors by comparing the percentage of incidents that involved future threats, the percentage of incidents that involved workers as stakeholders, and to what extent these intersected (e.g., workers as stakeholders AND future threats; neither; or just one or the other).

AI incident tracker 2: AI incident database (AIID)

The AI Incident Database (AIID) (https://incidentdatabase.ai) was launched in 2020 with support from the Partnership on AI (https://partnershiponai.org/) (McGregor, 2021). Drawing from crowdsourced submissions, the AIID depends on a taxonomy and annotation guidance developed by the Georgetown University Center for Security and Emerging Technology (Hoffmann et al., 2023; Responsible AI Collaborative, 2022). Each news report is assigned to a single incident, but each incident may collect multiple related news reports.



Is Starbucks shortchanging its baristas?

cbsnews.com · 2015

For Starbucks (SBUX) barista Kylei Weisse, working at the coffee chain helps him secure health insurance and some extra money while he studies at Georgia Perimeter College. What it

Tweet: @MarietjeSchaake

twitter.com · 2022

😱 Wait! What? Just when you think you've seen it all.... Meta's chatbot replied to the question askedby my colleague @kingjen : 'Who is a terrorist?' with my (given) name! That's right, not Bin Laden or the Unabomber,

Hundreds of AI tools have been built to catch covid. None of them helped.

technologyreview.com · 2021

When covid-19 struck Europe in March 2020, hospitals were plunged into a health crisis that was still badly understood. "Doctors really didn't have a clue how to manage these

Figure 2. Screenshot of the AIID's '*discovery*' query tool

Since AIID did not provide a direct method for filtering incidents involving *workers*, we downloaded the whole dataset of 3,545 full-text news reports collected into 721 incidents as bulk data on August 1st, 2024. Initially, we explored the download by querying for each of the three keywords *labor*, *jobs*, and *worker*; we identified additional terminology from the incidents with these keywords in their associated news reports and ultimately constructed a dataset by searching for the following queries:

- *Compensation*
- *Firing*
- *Hiring*
- *Scheduling*
- *Unemployment*
- *Worker death*
- *Workplace*

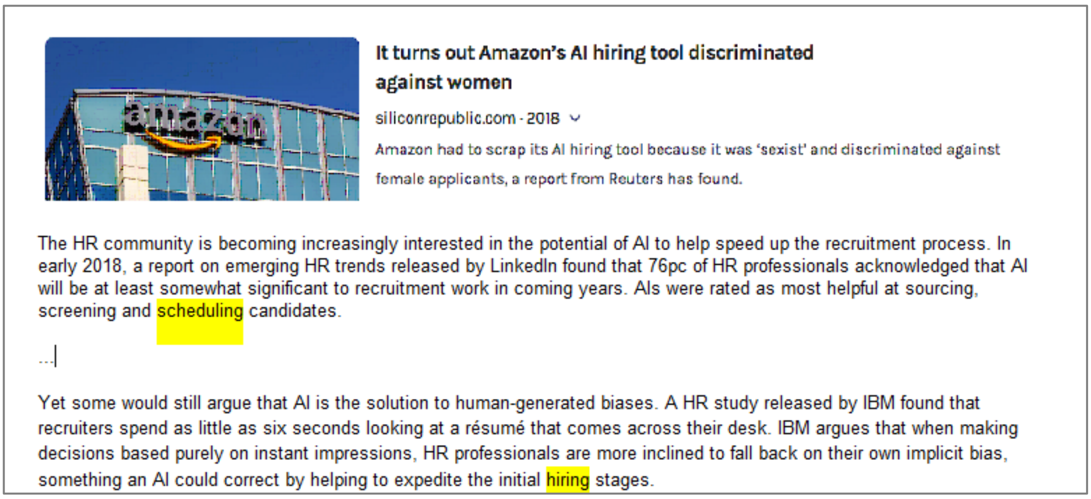
We recorded the number of results for each query and deduplicated, retrieving 304 keyword matches in 266 news reports grouped into 126 incidents, which were associated with a total of 1,183 news reports, both retrieved by and not retrieved by keywords. We calculated the number of news reports retrieved by each keyword and grouped by incident.

Ultimately, we categorised an incident as labour related only when a news report retrieved by a keyword indicated how AI has already contributed to harms or risks for workers' well-being in the workplace. Underlying our decision of whether an incident was eligible for inclusion in our labour-related dataset was the question: did automation or implementation of AI in the workplace either replace, disrupt, or augment the tasks and prospects of safe employment for workers? For instance, for celebrities and creative artists, we considered their capacity to extract value, so that we treated their bodies, personality, styles, or likeness as an intrinsic part of their labour.

We kept an inventory of thematic codes to classify recurring types of AI technologies, examples of workplace tasks, reported harms and risk mitigation topics included in the labour-related dataset. We excluded speculated harms that have not occurred, or that do not unambiguously bear directly on the workplace, such as gender or racial bias in search results for professions. Since we wanted to focus on AI's own impact on labour, we also excluded from the labour-related dataset:

1. Incidents related to the internal management of AI-focused companies (e.g., hiring, firing and leadership changes)
2. Incidents that originated outside employment, for instance related to:
    - Administration of social benefits, governance, justice, and law enforcement. Though relevant to labour, the administration of unemployment and social security benefits were considered to be outside the scope of the workplace.
    - Harmful representations pertaining to culture (e.g., generative AIs return violent imagery), content delivery and information retrieval (e.g., Image search for 'CEO' returns predominantly male-presenting results).
    - Anti-competition and monopolistic firm behaviour, except those specific to the job market
    - Deep fakes, except those generated as a condition of employment

Ultimately, we manually classified 57 incidents (with 184 keyword matches in 112 of their associated news reports – see Figure 3) as directly involving the effects of AI on labour. We compared these manually classified labour-related incidents with a keyword-based retrieval strategy: we collected the number of news reports retrieved by each keyword, grouped them by incident, and collected all news reports associated with the matching incidents.



It turns out Amazon's AI hiring tool discriminated against women

siliconrepublic.com · 2018

Amazon had to scrap its AI hiring tool because it was 'sexist' and discriminated against female applicants, a report from Reuters has found.

The HR community is becoming increasingly interested in the potential of AI to help speed up the recruitment process. In early 2018, a report on emerging HR trends released by LinkedIn found that 76pc of HR professionals acknowledged that AI will be at least somewhat significant to recruitment work in coming years. AIs were rated as most helpful at sourcing, screening and scheduling candidates.

...|

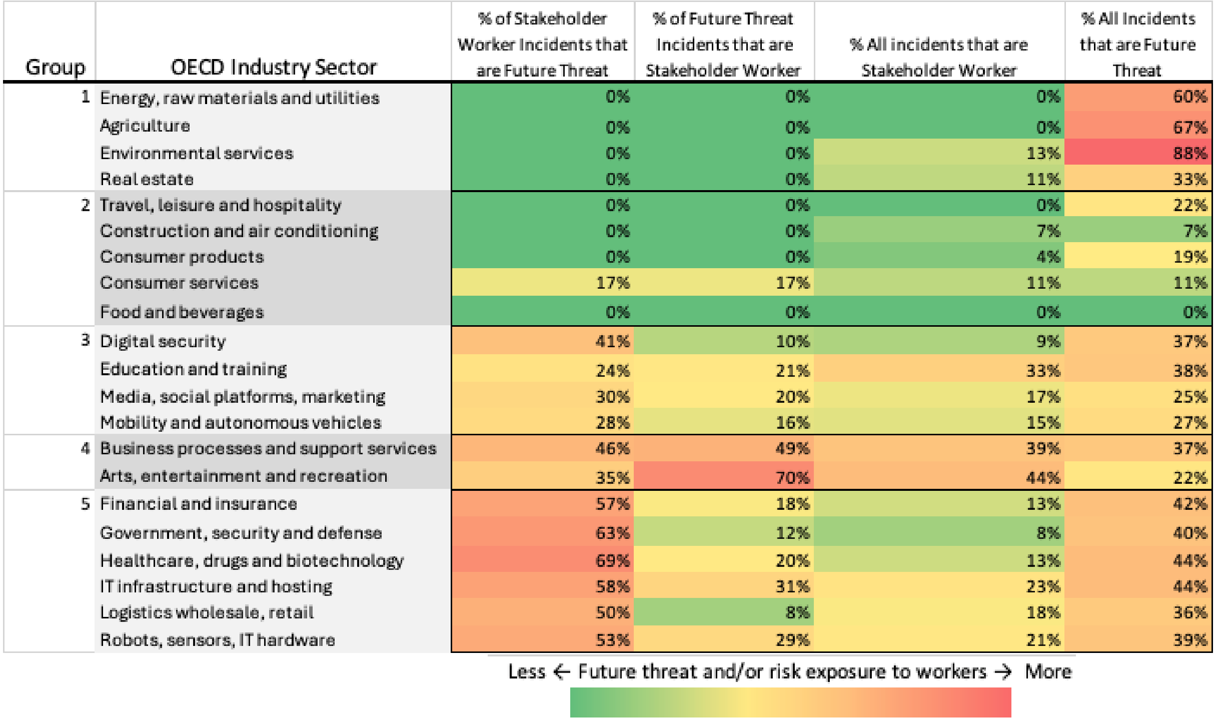
Yet some would still argue that AI is the solution to human-generated biases. A HR study released by IBM found that recruiters spend as little as six seconds looking at a résumé that comes across their desk. IBM argues that when making decisions based purely on instant impressions, HR professionals are more inclined to fall back on their own implicit bias, something an AI could correct by helping to expedite the initial hiring stages.

Figure 3. Sometimes multiple keyword searches retrieve a single news report: both '*scheduling*' and '*hiring*' are found in the news report (Short, 2018). The news report Short (2018) is associated with AIID's Incident 37 'female applicants down-ranked by Amazon recruiting tool.'

Results

AI incident tracker 1: OECD AIM

We compared the OECD AIM database as a whole to the subset of workers-related incidents (600 incidents with 6,744 associated news reports that the OECD AIM database categorised as having *workers* as '*affected stakeholders*'). Future threats are more pronounced for Workers: we found that 35% of *workers*-related incidents are marked as '*future threat*' compared to only 25% of incidents in the OECD AIM database as a whole. Figure 4 shows the percentage of all incidents involving workers broken down by industry sectors. The share of future threats with *workers* listed as affected as stakeholders (Figure 4 column 3) is particularly high for two sectors: *arts, entertainment, and recreation* (70%) and *business processes and support services* (49%) [Group 4]. Some industries [Group 1] have a larger percentage of future threat incidents but face little AI-related threat specifically to *workers*, now or in the future, such as *agriculture*; *energy, raw materials, and utilities*; *real estate*; and *environmental services*. Other sectors [Group 2] not only have incidents with minimal or zero AI impact for workers (Figure 4 column 2), but relatively few future threats generally (Figure 4 column 4): *food and beverages*; *construction and air conditioning*; *consumer products; consumers services*; *and travel, leisure, and hospitality*. Other industry sectors [Group 5] contain a moderate percentage of future threats, but their incidents do not proportionally affect *workers* more than other stakeholders (Figure 4 column 2); however, for this group, an incident is more likely to be a future threat when involving Workers as stakeholders in these industries, as seen in Table 1 column 1 for Group 5: *healthcare, drugs and biotechnology* (69%); *government, security and defence* (63%); *financial and insurance* (57%); *IT infrastructure and hosting* (58%); *logistics wholesale, retail* (50%); *robots, sensors, IT hardware* (53%). For some other industries [Group 3], *workers* have been identified as stakeholders in incidents, but incidents are less likely to be future threats, from Table 1 column 1: *digital security* (41%); *education and training* (24%); *media, social platforms, marking* (30%); *mobility* and autonomous vehicles (28%). These sectors face a moderate percentage of future threats that mostly concern other concerns besides *workers*, as seen in Figure 4 column 2 for Group 3.



Group	OECD Industry Sector	% of Stakeholder Worker Incidents that are Future Threat	% of Future Threat Incidents that are Stakeholder Worker	% All incidents that are Stakeholder Worker	% All Incidents that are Future Threat
1	Energy, raw materials and utilities	0%	0%	0%	60%
	Agriculture	0%	0%	0%	67%
	Environmental services	0%	0%	13%	88%
	Real estate	0%	0%	11%	33%
2	Travel, leisure and hospitality	0%	0%	0%	22%
	Construction and air conditioning	0%	0%	7%	7%
	Consumer products	0%	0%	4%	19%
	Consumer services	17%	17%	11%	11%
	Food and beverages	0%	0%	0%	0%
3	Digital security	41%	10%	9%	37%
	Education and training	24%	21%	33%	38%
	Media, social platforms, marketing	30%	20%	17%	25%
	Mobility and autonomous vehicles	28%	16%	15%	27%
4	Business processes and support services	46%	49%	39%	37%
	Arts, entertainment and recreation	35%	70%	44%	22%
5	Financial and insurance	57%	18%	13%	42%
	Government, security and defense	63%	12%	8%	40%
	Healthcare, drugs and biotechnology	69%	20%	13%	44%
	IT infrastructure and hosting	58%	31%	23%	44%
	Logistics wholesale, retail	50%	8%	18%	36%
	Robots, sensors, IT hardware	53%	29%	21%	39%

Less ← Future threat and/or risk exposure to workers → More

Figure 4. The proportion of OECD AIM incidents with workers as a stakeholder, broken down by each OECD-classified Industry, indicating the proportion that the database classified as '*future threat*'.

AI incident tracker 2: AIID

The 57 labour-related incidents were not the only incidents retrieved with the query terms we chose; Table 1 shows that different percentages of labour-related incidents were retrieved for different query terms, ranging widely, from 6/6 (100%) for scheduling to 11/33 (33%) for compensation. While we classified incidents as labour-related or not labour-related at the incident level, different articles related to the same incident presented the incident differently.

To examine the strength of the keyword-based signal that the incident should be interpreted as labour-related, we examined how many news reports associated with labour-related incidents were returned by a given keyword, also shown in Table 1.

		compensation	**firing**	**hiring**	**scheduling**	**unemployment**	**worker death**	**workplace**	**total matches**	**deduplicated**
Number of news reports retrieved by the keyword	From Labour-related Incidents	16	20	55	15	4	30	46	186	**157**
	From All Incidents	51	28	84	15	19	40	68	304	**266**
Number of matching incidents (news reports retrieved by the keyword, grouped by incident)	Labour-related Incidents	11	11	21	6	4	8	22	86	**57**
	All Incidents	33	19	44	6	8	17	37	164	**127**
Total news reports associated with matching incidents	From Labour-related Incidents	86	164	189	52	40	84	353	968	**533**
	From All Incidents	320	246	425	52	108	137	519	1807	**1183**

Table 1. The number of incidents and labour-related incidents returned for each query term in AIID. We deduplicated the total since the same news report may be retrieved by multiple keyword searches.

We iteratively classified AIID's labour-related AI incidents as shown in Table 2. This resulted in a total of 33 labels which we grouped into 4 categories: technology, workplace task, risks, and policy domain.

Technology	Workplace Task	Risks	Policy Domain
- Robotics - Algorithmic design - Autonomous driving - Predictive analytics - Generative AI - Computer vision - Networks - Manual data classification	- Warehouse operations - Recruitment, Personnel and hiring decisions - Assembly - Automating contracts and business rules - Security monitoring and surveillance - Termination decisions - Data classification - Job performance assessment - Journalism and reporting - Health and safety protocol - Creative content production - Content filtering - Jurisprudence - Chatbot - Conduct and behaviour	- Occupational hazard overwork or fatigue - Physical harm - No human override - Gender inequity - Racial inequity - Unreliable information - Psychological harm - Technological under comprehension - Human attribution issues - Financial harm or reputational harm - Political harm	- Labour regulations and workplace protections - Worker privacy, likeness, and labour ownership - Scaling automation versus labour cost benefit - Employment, contracting and termination - Unreasonable job expectations, assessment, and disciplining - Worker privacy, likeness, and labour ownership - Worker integrity and whistleblowing protections - Content moderation/Data labour

Table 2. Our own categorization of AIID's labour-related AI incidents into 33 labels in 4 categories

Discussion

The AI incident trackers do not facilitate ready retrieval of labour-related incidents. Being unable to readily see the incidents related to labour makes it impossible to understand what problems are associated with the AI-related risks and harms to labour.

The two AI incident trackers we examined used limited existing labour-related terminology. In OECD AIM, we found only one relevant term—the stakeholder *workers*—which is applied not only to AI end-users whose labour is replaced or augmented by AI systems, but also to AI producers (e.g., workers under pressure to engineer AI systems).

In AIID we identified three relevant term sets—'*data input*', 'AI *task*' and '*end user amateur/expert*'—from the CSET taxonomy manual for incident classification. However, '*data input*' did not consider the human labour and roles involved in producing input data (such as training data that the AI system was trained on). 'AI *task*' considers tasks such as '*resume reading*' or '*chatbot*', which can be viewed as labour replacement (Hoffmann et al., 2023). The inherent problem of staffing a support hotline for eating disorders with chatbots instead of a human call centre workforce, for example, is that this violates our expectations that therapeutic applications may require emotional labour (Posada, 2020), but 'AI *task*' elides emotional labour.

While the OECD AIM has a taxonomy for industry sectors, its reliance on news reports risks overrepresenting some sectors such as entertainment. *Workers* stakeholder incidents tend to concentrate in the *arts, entertainment, and recreation* industry sector, covering workers' copyright infringement, data protections and personal likeness. And the challenges of generative AI in the workplace have been raised by recent Hollywood strikes calling for controls over licensing creative content. More attention should be paid to understanding AI's effects on different sectors, especially given the likelihood that some sectors are overrepresented. It was unclear when reconciling how classification terms like worker and '*future threat*' should be interpreted when combined. Should policymakers rely on OECD AIM's classification technique when it excludes workers as stakeholders and simultaneously designates incidents in industrial sectors pertaining to earthborn land uses (like real estate, energy, agriculture, etc.) as '*future threat*?' Representing incidents in this manner presents its own risk and high stakes: Can policymakers properly gauge why heightened future risks in a sector should not portend in harm for workers?

The heavy reliance on news coverage, common to current AI trackers (Turri & Dzombak, 2023), leads to some challenges. News reports can frame different stories with the same evidence, by choosing what to ignore, background or highlight. When Amazon's implementation of a predictive hiring algorithm resulted in gender discrimination (Dastin, 2018), some news reports attributed the harm to sexist input data (Kraus, 2018) or mischaracterised output from reinforcement learning algorithms as outperforming human judgment (Short, 2018) and free from '*user abuse*' (Doctorow, 2018; Papadopaulos, 2018). Current AI incident report approaches group incidents geographically and temporally, which reduces the structured data available about where and when incidents took place. Consequently, in AIID, determining which incidents are labour related requires a substantial amount of judgement. When a single news report frames the news event as a labour issue, it makes the case that the incident is labour-related. Yet each incident collects multiple news reports, and labour was often in the background or used as supplemental frame. Finding incidents for labour-related policy topics was different from finding mentions of keywords. Some keywords have high precision, retrieving only labour-related incidents, but low recall, giving a limited picture of the risks and harms. For instance, scheduling only matched incidents about controversial algorithmically driven worker shift management software. Keywords with stronger recall, like '*compensation*' and '*hiring*', tended to be less precisely labour-focused, with matching incidents often referring to internal AI enterprise practices and consumer settlements.

Power relationships between the working class and management (and by extension, capital) are salient in the labour-related incidents we flagged: who introduces AI into the workplace? Who is responsible for AI when it backfires? Automated staffing decisions using AI-supported predictive analytics led to multiple problems (e.g., Southwest Airlines' flight cancellations (Sider, 2022), Tesla's factory delays (Duhigg, 2018)). Lack of human override and management's technical under-comprehension exacerbated these problems. Values need to be taken into consideration to determine which labour conflicts themselves qualify as harms.

Future work

Future work should examine how the distribution of news reports used in AI incident trackers vary over time, across industries and in relation to AI principles and news media sources and audiences. The ratio of news reports to incidents varies, with some incidents (such as the Hollywood strikes) heavily reported, without regard to the actual exposure to AI-related harms in a given industrial sector. Comparing how news reports frame the same AI incident (e.g., with framing analysis) would be valuable. Likewise, researchers should seek to understand how different groupings into incidents can contribute to policymakers' problem definitions, perhaps by examining the variation in how news reports are grouped in different AI incident trackers.

Future research could systematically identify which query terms to use, including testing stemming for words such as hiring, firing, and scheduling as well as considering additional terminology such as '*crowd labor*'. The use of terminology in AI governance could also be fruitfully examined, e.g., '*trustworthy AI*', '*ethical AI*', '*AI for good*', '*beneficial AI*', and '*responsible AI*' (Stix, 2022).

Current definitions for harm or hazard (Placani, 2017; Rowe, 2021) need to be revisited when evaluating the relationship between AI and labour to enable incident risk prevention (Meyer, 2023). Incident reporting, which has roots in risk management and safety (Johnson, 2003), may not be sufficient for understanding AI risks to workers. Alternative conceptual frameworks for risk management from epidemiology, audit culture and social work may be more suitable for centring workers.

Future work should investigate what harm, incident, potential harm, or hazard mean in the context of labour. For instance, are '*disruptions*' ever supposed to be weathered as part of events contributing to the normal functioning of the economy? Tradeoffs must consider multiple perspectives (e.g., worker, manager, capital) in the political economy of labour.

Sociotechnical perspectives will be needed. Above all, it is important to consider *'how algorithms may reshape organizational control'*, as Kellog et al. (2020) review. Sociotechnical analysis of fairness in machine learning (Selbst et al., 2019) can inspire new approaches for attending to the power dynamics of AI and labour, drawing on fields such as organizational behaviour and management science, human resources management, labour law and the sociology of labour.

Future classifications of AI and labour should document when a system extracts labour value unjustly. Key examples would be when a '*worker*' can claim direct credit or likeness to the source data (such as Scarlett Johanssen's voice appropriated by OpenAI (Pisani & Albert, 2024)) or when input data is created with the express purpose of serving an AI system (Satariano & Mozur, 2023) (see our content moderation/data labour category in Table 3). Taxonomies for organizing how AI interacts with prior types of '*people problems*' inside the workplace (Moore, 2019) may be helpful since both AI and AI incident reporting ultimately depend on how people are organised to share information within a firm. Standardised terminologies and definitions for workplace safety can be informed by International Labour Organization worker protections requiring information disclosure in the event of mass dismissal and redundancy (De Stefano, 2019). Incidents are made visible based on the power structures upheld by a given taxonomy: naming and power is worth specific consideration.

Conclusions

The two incident trackers we investigated do not adequately capture the nuanced impacts of AI on labour. Particular attention needs to be paid to power imbalances that may increase risks for harm in the workplace. Better definitions are needed to capture labour-related AI incidents, to help policymakers gather evidence to anticipate and mitigate the risks and harms threatening workers across diverse industries. AI incident trackers' reliance on news reports and limited vocabulary for identifying worker-related harms and risks leads to gaps in understanding AI for policy formulation and problem definition.

Are the labour-related risks of AI intelligible in the incident trackers with a level of detail and reliability that could support comprehensive policy and problem definition for workers' protection from the harms of AI? Our answer is no.

Acknowledgements

This research is funded by the United States Institute of Museum and Library Services RE-250162-OLS-21. Thanks to Zachary Kilhoffer for discussions of AI trackers and to Corinne McCumber, Heng Zheng, Yuanxi Fu and Zhixuan 'Kyrie' Zhou for providing feedback on a draft. Jodi Schneider was supported in part as the 2024–2025 Perrin Moorhead Grayson and Bruns Grayson Fellow, Harvard Radcliffe institute for advanced study. CRediT roles: TL: conceptualization, data curation, formal analysis, methodology, visualization, writing – original draft, writing – review and editing; Jodi Schneider: conceptualization, funding acquisition, supervision, writing – review and editing.

About the author

Theodore Dreyfus Ledford is a PhD student in Information Sciences at the University of Illinois at Urbana-Champaign. tledfo2@illinois.edu

References

Abercrombie, G., Benbouzid, D., Giudici, P., Golpayegani, D., Hernandez, J., Noro, P., Pandit, H., Paraschou, E., Pownall, C., Prajapati, J., Sayre, M. A., Sengupta, U., Suriyawongkul, A., Thelot, R., Vei, S., & Waltersdorfer, L. (2024). *A collaborative, human-centred taxonomy of AI, algorithmic, and automation harms* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2407.01294

Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2), 3–30. https://doi.org/10.1257/jep.33.2.3

Arda, S. (2024). *Taxonomy to regulation: A (geo)political taxonomy for AI risks and regulatory measures in the EU AI act* [Preprint]. arXiv. http://arxiv.org/abs/2404.11476

Arntz, M., Gregory, T., & Zierahn, U. (2016). The risk of automation for jobs in OECD countries: A comparative analysis (OECD Social, Employment and Migration Working Papers, 189). OECD. https://doi.org/10.1787/5jlz9h56dvq7-en

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. Journal of Economic Perspectives, 29(3), 3–30. https://doi.org/10.1257/jep.29.3.3

Autor, D.H. (2022). The labor market impacts of technological change: From unbridled enthusiasm to qualified optimism to vast uncertainty (Working Paper w30074). National Bureau of Economic Research. https://doi.org/10.3386/w30074

Baumgartner, F. R., & Jones, B. D. (2015). The politics of information: Problem definition and the course of public policy in America. University of Chicago Press.

Booth, H., Rike, D., & Witte, G. A. (2013 December). The National Vulnerability Database (NVD): Overview (Information Technology Laboratory Bulletin Series). National Institute of Standards. https://www.nist.gov/publications/national-vulnerability-database-nvd-overview

Cameron, J. (Director). (1984, October 26). The Terminator [Film]. Cinema '84; Euro Film Funding; Hemdale.

Cattell, S., Ghosh, A., & Kaffee, L.-A. (2024). Coordinated flaw disclosure for AI: Beyond security vulnerabilities [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2402.07039

Critch, A., & Russell, S. (2023). TASRA: A Taxonomy and Analysis of Societal-Scale Risks from AI [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2306.06924

Dastin, J. (2018, October 10). Insight—Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G/

De Stefano, Valerio. (2019). 'Negotiating the algorithm': Automation, artificial intelligence, and labor protection. Comparative Labor Law & Policy Journal, 41(1), 15-46.

Doctorow, C. (2018, October 11). Amazon trained a sexism-fighting, resume-screening AI with sexist hiring data, so the bot became sexist. Boing Boing. https://boingboing.net/2018/10/11/garbage-conclusions-out.html

Duhigg, C. (2018, December 13). Dr. Elon & Mr. Musk: Life inside Tesla's production hell. Wired. https://www.wired.com/story/elon-musk-tesla-life-inside-gigafactory/

Head, B. W. (2010). Reconsidering evidence-based policy: Key issues and challenges. Policy and Society, 29(2), 77–94. https://doi.org/10.1016/j.polsoc.2010.03.001

Hoffmann, M., & Frase, H. (2023). Adding structure to AI harm. Center for Security and Emerging Technology. https://doi.org/10.51593/20230022

Hoffmann, M., Narayanan, M., Mitra, A., Liao, Y.-J., & Frase, H. (2023). CSET AI Harm Taxonomy for AIID and Annotation Guide. https://github.com/georgetown-cset/CSET-AIID-harm-taxonomy

Hutiri, W., Papakyriakopoulos, O., & Xiang, A. (2024). Not my voice! A taxonomy of ethical and safety harms of speech generators. Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency, 359–376. https://doi.org/10.1145/3630106.3658911

Johnson, C. (2003). Failure in safety critical systems: A handbook of accident and incident reporting. University of Glasgow Press.

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. Academy of Management Annals, 14(1), 366–410. https://doi.org/10.5465/annals.2018.0174

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (Eds.). (2000). Creating safety systems in health care organizations. In To err is human: Building a safer health system (pp. 155–204). National Academies Press. https://doi.org/10.17226/9728

Kraus, R. (2018, October 10). Amazon's sexist recruiting algorithm reflects a larger gender bias. Mashable. https://mashable.com/article/amazon-sexist-recruiting-algorithm-gender-bias-ai

Laker, B. (2023, October 30). The future of work: Navigating the complex landscape of flexibility. Forbes. https://www.forbes.com/sites/benjaminlaker/2023/10/30/the-future-of-work-navigating-the-complex-landscape-of-flexibility/

Lane, M., & Saint-Martin, A. (2021). The impact of Artificial Intelligence on the labour market: What do we know so far? (OECD Social, Employment and Migration Working Papers 256). https://doi.org/10.1787/7c895724-en

MacKillop, E., & Downe, J. (2023). What counts as evidence for policy? An analysis of policy actors' perceptions. Public Administration Review, 83(5), 1037–1050. https://doi.org/10.1111/puar.13567

Macrae, C. (2016). The problem with incident reporting. BMJ Quality & Safety, 25(2), 71–75. https://doi.org/10.1136/bmjqs-2015-004732

McGregor, S. (2021). Preventing repeated real world AI failures by cataloguing incidents: The AI incident database. Proceedings of the AAAI Conference on Artificial Intelligence, 35(17), 15458–15463. https://doi.org/10.1609/aaai.v35i17.17817

Meyer, C. O. (2024). Can one 'prove' that a harmful event was preventable? Conceptualizing and addressing epistemological puzzles in post incident reviews and investigations. Risk, Hazards & Crisis in Public Policy, 15(3), 374–392. https://doi.org/10.1002/rhc3.12281

Moore, P. V. (2019). The mirror for (artificial) intelligence: In whose reflection? Automation, artificial intelligence, & labor law. Comparative Labor Law & Policy Journal, 41(1), 47–68.

NASA Aviation Safety Reporting System. (n.d.). ASRS: the case for confidential incident reporting systems (ASRS Research Papers 60). NASA Aviation Safety Reporting System. https://asrs.arc.nasa.gov/docs/rs/60_Case_for_Confidential_Incident_Reporting.pdf

OECD. (2022a). OECD framework for the classification of AI systems (OECD Digital Economy Papers 323). OECD. https://doi.org/10.1787/cb6d9eca-en

OECD. (2022b). Harnessing the power of AI and emerging technologies: Background paper for the CDEP Ministerial meeting (OECD Digital Economy Papers 340). OECD. https://doi.org/10.1787/f94df8ec-en

OECD. (2023). Stocktaking for the development of an AI incident definition (OECD Artificial Intelligence Papers 4). OECD. https://doi.org/10.1787/c323ac71-en

OECD. (2024). Defining AI incidents and related terms (OECD Artificial Intelligence Papers 16). OECD. https://doi.org/10.1787/d1a8d965-en

Papadopoulos, L. (2018, Oct 12). Amazon shuts down secret AI recruiting tool that taught itself to be sexist. Interesting Engineering. https://interestingengineering.com/innovation/amazon-shuts-down-secret-ai-recruiting-tool-that-taught-itself-to-be-sexist

Pawson, R., Wong, G., & Owen, L. (2011). Known knowns, known unknowns, unknown unknowns: The predicament of evidence-based policy. American Journal of Evaluation, 32(4), 518–546. https://doi.org/10.1177/1098214011403831

Pisani, J., & Albert, V. (2024, May 20). Scarlett Johansson rebukes OpenAI over 'Eerily Similar' ChatGPT voice. Wall Street Journal. https://www.wsj.com/tech/ai/openai-chatgpt-sky-voice-scarlett-johansson-43d13bbf

Placani, A. (2017). When the risk of harm harms. Law and Philosophy, 36, 77–100. https://doi.org/10.1007/s10982-016-9277-x

Posada, J. (2020). The future of work is here: Toward a comprehensive approach to Artificial Intelligence and labour [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2007.05843

Responsible AI Collaborative. (2022). Founding Report. Responsible AI Collaborative. https://asset.cloudinary.com/pai/cf01cce1af65f5fbb3d71fa092d001db

Reynard, W. D. (1986). The development of the NASA aviation safety reporting system (NASA reference publication 114). National Aeronautics and Space Administration, Scientific and Technical Information Branch. Available as ASRS Research Papers 34 https://asrs.arc.nasa.gov/docs/rs/34_Development_of_NASA_ASRS.pdf

Rodrigues, R., Resseguier, A., & Santiago, N. (2023). When Artificial Intelligence fails: The emerging role of incident databases. Public Governance, Administration and Finances Law Review, 8(2), 17–28. https://doi.org/10.53116/pgaflr.7030

Rowe, T. (2021). Can a risk of harm itself be a harm? Analysis, 81(4), 694–701. https://doi.org/10.1093/analys/anab033

Satariano, A., & Mozur, P. (2023, February 7). The people onscreen are fake. The disinformation is real. The New York Times. https://www.nytimes.com/2023/02/07/technology/artificial-intelligence-training-deepfake.html

Schiff, D. S. (2024). Framing contestation and public influence on policymakers: Evidence from US artificial intelligence policy discourse. Policy and Society, 43(3), 255–288. https://doi.org/10.1093/polsoc/puae007

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. Proceedings of the Conference on Fairness, Accountability, and Transparency, 59–68. https://doi.org/10.1145/3287560.3287598

Shelby, R., Rismani, S., Henne, K., Moon, AJ., Rostamzadeh, N., Nicholas, P., Yilla-Akbari, N., Gallegos, J., Smart, A., Garcia, E., & Virk, G. (2023). Sociotechnical harms of algorithmic systems: Scoping a taxonomy for harm reduction. Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society, 723–741. https://doi.org/10.1145/3600211.3604673

Short, E. (2018, October 11). It turns out Amazon's AI hiring tool discriminated against women. Silicon Republic. https://www.siliconrepublic.com/careers/amazon-ai-hiring-tool-women-discrimination

Shrishak, K. (2023). How to deal with an AI near-miss: Look to the skies. Bulletin of the Atomic Scientists, 79(3), 166–169. https://doi.org/10.1080/00963402.2023.2199580

Sider, A. (2022, December 28). How Southwest Airlines melted down. Wall Street Journal. https://www.wsj.com/articles/southwest-airlines-melting-down-flights-cancelled-11672257523

Stix, C. (2022). Artificial intelligence by any other name: A brief history of the conceptualization of "trustworthy artificial intelligence." Discover Artificial Intelligence, 2, 26. https://doi.org/10.1007/s44163-022-00041-5

Turri, V., & Dzombak, R. (2023). Why we need to know more: Exploring the state of AI incident documentation practices. Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society, 576–583. https://doi.org/10.1145/3600211.3604700

van der Kleij, R., Schraagen, J. M., Cadet, B., & Young, H. (2022). Developing decision support for cybersecurity threat and incident managers. Computers & Security, 113, 102535. https://doi.org/10.1016/j.cose.2021.102535

Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P.-S., Mellor, J., Glaese, A., Cheng, M., Balle, B., Kasirzadeh, A., Biles, C., Brown, S., Kenton, Z., Hawkins, W., Stepleton, T., Birhane, A., Hendricks, L. A., Rimell, L., Isaac, W., … Gabriel, I. (2022). Taxonomy of risks posed by language models. Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 214–229. https://doi.org/10.1145/3531146.3533088

Appendix

Data availability

Data underlying this research is available at https://doi.org/10.13012/B2IDB-1156758_V1