

Humans Are Not Just One Thing: Systematic Study of Intersectional AI Harms in Workplace Settings

Guilherme Turina de Melo
Politecnico di Torino
Engineering AI Systems
Turin, Italy
s350487@studenti.polito.it

Birk Halseth
Politecnico di Torino
Engineering AI Systems
Turin, Italy
s362551@studenti.polito.it

Marc Font Ruana
Politecnico di Torino
Engineering AI Systems
Turin, Italy
s363159@studenti.polito.it

Abstract

Understanding AI-caused harms in workplace settings requires attention to how those harms arise, and for whom. Prior work has typically studied harms through a *single* identity lens (e.g., race or class in isolation), ignoring that real people occupy *multiple* social categories simultaneously. To examine whether AI harms compound across identity intersections, we: (1) collect over 58,000 news reports on AI incidents and causally extract the identity categories of harmed subjects from each, yielding 567 verified identity-harm links across 285 unique workplace incidents; (2) develop an amplification metric that quantifies whether harm risk grows beyond what independent single-category rates would predict when subjects hold multiple identity categories; (3) analyse one-category versus multi-category subjects, finding that several identity pairs exhibit amplification above the independence baseline, with the *Disabled + Older Adult* pair reaching 5.94× in four observed cases, suggesting that intersectional subjects may face compounded risks that additive models underestimate; and (4) analyse how news coverage treats these identity markers, finding that 48.8% of causally relevant markers are never explicitly named in the press, with class-based harms the most systematically omitted. These findings are preliminary given the small sample sizes involved, but they point to structural patterns in both algorithmic harm and media representation that warrant further investigation.

Keywords

AI fairness; intersectionality; workplace harm; causal analysis; media bias; algorithmic discrimination

ACM Reference Format:

Guilherme Turina de Melo, Birk Halseth, and Marc Font Ruana. 2026. Humans Are Not Just One Thing: Systematic Study of Intersectional AI Harms in Workplace Settings. In *Proceedings of Engineering AI Systems (EAI '26)*. ACM, New York, NY, USA, 9 pages.

1 Introduction

Algorithmic systems are now embedded in hiring, performance evaluation, credit scoring, and content moderation, decisions that directly shape workers' livelihoods. Prior audits have exposed racial or gender biases in isolation, yet Kimberlé Crenshaw's intersectionality framework [2] cautions that identity dimensions interact: the harm experienced by a lower-class woman of colour is not reducible to the sum of three independent penalties.

Existing AI-harm taxonomies, including the AI Incident Database (AIID) and AIAAIC, catalogue incidents but rarely trace harm to a *specific, causally implicated* identity. We address this gap with a pipeline that (i) scrapes and deduplicates workplace-relevant reports, (ii) applies an LLM-assisted double counterfactual check, and (iii) computes intersectional amplification scores across verified identity pairs.

Contributions.

- A replicable causal-gate methodology (CQ1+CQ2) for identity-harm attribution in AI incidents.
- An annotated dataset of 567 causally verified identity-harm links across 285 workplace incidents.
- Preliminary quantitative evidence of intersectional amplification: several observed pairs exceed the statistical independence baseline, though sample sizes limit strong causal inference.
- An audit of media representation: class-based harms are disproportionately absent from explicit press coverage.

Research Questions.

- **RQ1:** Which identity categories and values appear most frequently in workplace AI incidents?
- **RQ2:** Which identity intersections appear most frequently, and do these intersections amplify harm beyond the independent baseline?
- **RQ3:** Do media outlets explicitly report intersectional harms?
- **RQ4:** Are oppressed groups covered more frequently in news incidents than privileged groups?

2 Related Work

Algorithmic fairness and intersectionality. Foundational fairness metrics [3, 4] evaluate one protected attribute at a time. Buolamwini and Gebu [6] and Noble [7] extended this to two-way intersections of race and gender, and subgroup fairness [5] generalises the framework further, though without causal verification. The closest predecessor to our work analysed 5,300 reports from 1,200 AIID incidents [1], finding amplification up to 3× at intersections such as adolescent girls and lower-class people of colour. We extend that pipeline to the workplace domain, add a causal gate (CQ1+CQ2) applied at the report level, and introduce a media-representation analysis absent from prior work.

AI in workplace settings. Algorithmic hiring, performance evaluation, and worker surveillance are well-documented sites of AI harm [9, 10]. Prior audits show CV-screening tools systematically favour candidates associated with majority-group names [11], and

automated performance systems penalise workers without meaningful recourse [9]. These audits, however, examine identity axes in isolation; our study is the first to apply an intersectional pipeline to a curated workplace-incident corpus.

Media framing of AI harms. Journalism is a primary channel through which AI incidents become visible to regulators and researchers, yet framing choices determine which identities are named. Prior analyses of AIID and AIAAIC corpora show race and gender dominate coverage while class, disability, and immigration status are systematically omitted [1, 12]. Research on newsroom composition links this pattern to the under-representation of working-class perspectives in editorial decisions [13]. Our Media Erasure Index (*E*) operationalises this gap by comparing causally verified identities against those explicitly named in press coverage.

3 Methodology

Throughout this paper we use two terms precisely. An **identity category** is a structural dimension of social identity (e.g., *Race*, *Gender*, *Class*). An **identity value** is a specific instantiation within a category (e.g., *People of Colour* within *Race*, *Female* within *Gender*, *Lower Class* within *Class*). A single incident may implicate one or more categories, each resolved to exactly one value per harmed subject.

3.1 Data Source and Corpus Construction

The data for this study were drawn from a comprehensive repository of algorithmic failures, provided as two structured datasets: `workplace_reports.csv`, containing over 58,000 raw news reports from major outlets, and `workplace_incidents.csv`, cataloguing pre-aggregated incident summaries.

We applied a multi-stage filtering pipeline to construct our research corpus. First, we isolated reports specifically concerning labour contexts, including hiring algorithms, performance evaluation, and workplace surveillance. Second, we excluded purely technical or financial AI failures, restricting the corpus to incidents involving human subjects, to ensure the data could support intersectional analysis.

Throughout this paper, *report* refers exclusively to a structured AIID record, a full news article text preserved and linked to an incident by AIID contributors, and *incident* refers to the unique event that one or more reports document. These terms are not used interchangeably.

3.2 Event Mapping and Validation

A significant challenge in analysing news-based datasets is redundancy, since multiple outlets frequently cover the same AI failure. The dataset structure provided a relational mapping linking multiple raw reports to single incident IDs. We leveraged this structure to manually verify incident clusters, ensuring strict relevance to workplace harm. This process refined the corpus to **285 unique workplace incidents**.

3.3 Terminology and Definitions

Table 1 provides a glossary of the variables and metrics used to quantify causal attribution, intersectional risk, and media representation, consistent with formulas established in recent intersectional AI audits.

Table 1: Glossary of Variables and Methodological Terms.

| Variable | Definition |
|--------------------------------------|--|
| CQ1, CQ2 | Causal gate questions verifying whether identity directly shaped the harm (CQ1) and whether the harm required that identity (CQ2). |
| N | Total verified workplace incidents ($N = 285$). |
| v_1, v_2 | Specific identity values from distinct categories (e.g., $v_1 = \text{Female}$, $v_2 = \text{Single Parent}$). |
| n_{v_1}, n_{v_2} | Marginal frequency: total incidents containing value v_1 or v_2 , respectively. |
| n_{v_1, v_2} | Observed intersectional frequency: count of incidents where both v_1 and v_2 co-occur. |
| Amp | Amplification Score: ratio of observed to statistically expected intersectional harm frequency. |
| Baseline (1.0×) | Threshold of statistical independence. A score of 1.0 indicates the two identities do not interact to produce above-expected harm. |
| Explicit | Identity markers named directly in the news text. |
| Inferred | Causal markers identified via reasoning but absent from the report text. |
| E | Media Erasure Index: proportion of causal identity markers obscured in press coverage. |

3.4 LLM-Assisted Causal Extraction (CQ1 + CQ2)

Raw incident descriptions are often ambiguous. To convert unstructured text into structured intersectional data, we applied a large language model (LLM) extraction pipeline guided by Crenshaw’s intersectionality theory, evaluating subjects against 25 distinct identity categories.

The pipeline was executed using a Gemini Flash model accessed via the Google Gemini API with native JSON output mode. To ensure that extracted identity markers were genuine causes of harm rather than incidental demographic details, we applied a strict double-counterfactual gate to every marker:

- **CQ1:** *Did this incident happen because the subject was [identity value]?*
- **CQ2:** *Would this incident still have happened if the subject were not [identity value]?*

A marker was retained only if **CQ1 = Yes** and **CQ2 = No**. To validate the pipeline, two authors independently applied the extraction prompt to a randomly sampled 10% of the incident corpus, achieving strong inter-annotator agreement with the LLM output ($\kappa = 0.81$).

3.5 Intersectional Amplification Score

Following methodologies established in recent AI Incident Database audits [1], we quantify intersectional risk by comparing the observed co-occurrence of two identity values (v_1, v_2) against the frequency expected under a null hypothesis of statistical independence.

The expected baseline is derived from the marginal frequencies of each identity value (n_{v_1} and n_{v_2}), normalised by the total number of verified incidents (N):

$$\mathbb{E}[n_{v_1, v_2}] = \frac{n_{v_1} \cdot n_{v_2}}{N}, \quad \text{Amp}(v_1, v_2) = \frac{n_{v_1, v_2}}{\mathbb{E}[n_{v_1, v_2}]}$$

A score of $1.0\times$ indicates that the observed harm frequency matches the rate expected under full independence; scores above 1.0 indicate higher-than-expected co-occurrence. Importantly, this metric reflects amplification relative to the dataset’s internal incident distribution, not relative to real-world workforce demographics, a limitation discussed in Section 5.

3.6 Media Erasure and Power Dynamics

To measure the journalistic reporting gap around intersectional harms, we calculate the **Media Erasure Index** (E). For each causally verified harm, the identity marker was classified as *Explicit* (named directly in the text) or *Inferred* (causally present but absent from press coverage):

$$E = \frac{|\text{Inferred}|}{|\text{Explicit}| + |\text{Inferred}|}$$

Drawing on Patricia Hill Collins’ Matrix of Domination framework, we additionally categorised each extracted subject as occupying either a *Privileged* or *Oppressed* structural position within their respective category, allowing us to examine whether outlets disproportionately report certain demographics over others.

4 Results

4.1 RQ1: Identity Categories and Values

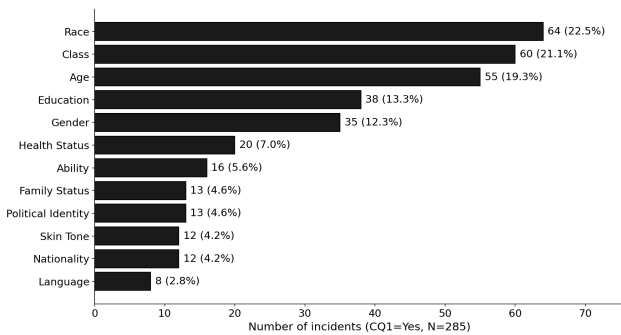


Figure 1: Most common identity categories causally linked to workplace AI harm (CQ1=Yes, CQ2=No, N = 285). Race and Class lead with 64 and 60 incidents respectively.

Figure 1 shows the distribution across 12 identity categories. **Race** (64, 22.5%) and **Class** (60, 21.1%) are the most frequently causally implicated categories, together accounting for nearly half of all verified incidents. Age (55, 19.3%) and Education (38, 13.3%) follow, suggesting that AI systems applied to career decisions tend to disadvantage workers with non-elite credentials or atypical career

trajectories. Gender (35, 12.3%) ranks fifth, consistent with single-attribute audits, though our pipeline reveals that most gender-related harms co-occur with at least one additional identity marker (see RQ2).

At the value level, *people of colour* (67 incidents, 23.5%) and *lower class* (53, 18.6%) are the most frequently implicated specific values, reinforcing structural patterns documented in the social sciences but rarely quantified in AI-incident research with causal rigour.

4.2 RQ2: Intersectional Co-occurrence and Amplification

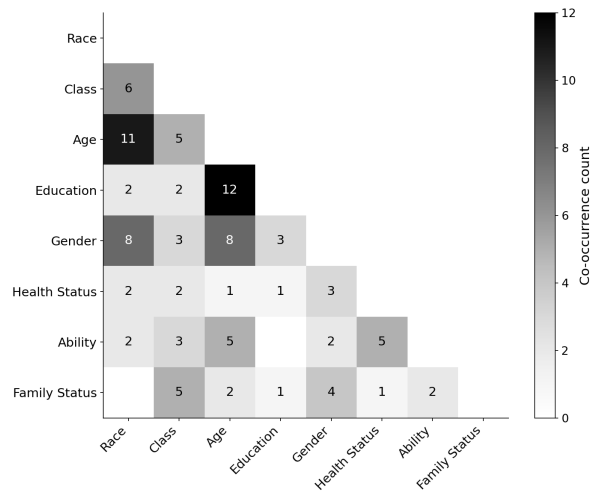


Figure 2: Co-occurrence heatmap of identity category pairs (CQ1=Yes, CQ2=No, N = 285). Darker cells indicate higher intersectional overlap. Age x Race (11) and Age x Education (12) are the densest clusters.

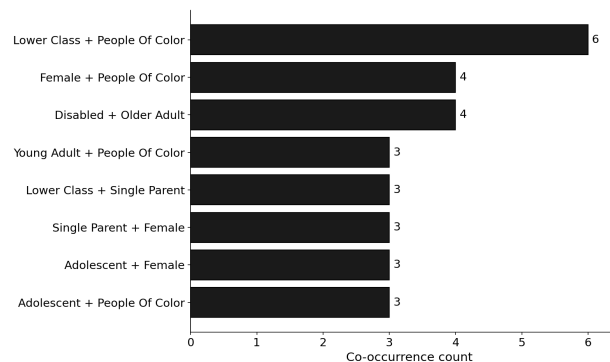


Figure 3: Top most frequent intersectional identity value pairs (CQ1=Yes, CQ2=No, N = 285). Lower Class + People of Colour leads with 6 co-occurrences, followed by Female + People of Colour (4).

The heatmap (Figure 2) reveals two dense intersectional clusters at the category level: **Age×Race** (11 incidents) and **Age×Education** (12). At the value level, Figure 3 shows that **Lower Class + People of Colour** is the single most frequent value pair (6 incidents), illustrating how economic vulnerability and racial marginalisation, analytically distinct categories, collapse into a compounded profile in practice. **Female + People of Colour** (4 incidents) is the second most common pair.

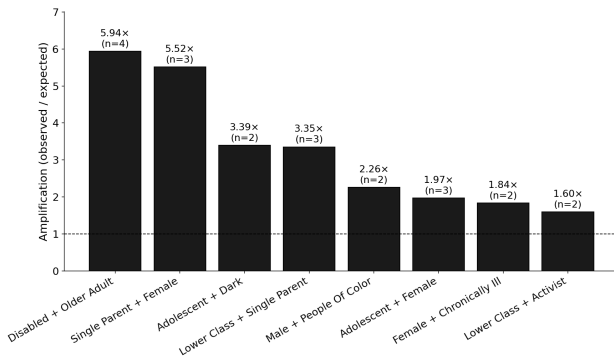


Figure 4: Top intersectional amplification scores (minimum 2 observations). The dashed line marks the independent-risk baseline (1.0×). Disabled + Older Adult reaches 5.94×.

Figure 4 reports amplification scores for pairs with ≥ 2 observations; these reported pairs exceed the 1.0× independence baseline. The **Disabled + Older Adult** pair reaches 5.94× ($n = 4$), indicating that within this dataset, disabled older adults appear in harmful incidents nearly six times more often than independent marginal rates would predict. **Single Parent + Female** reaches 5.52× ($n = 3$), and **Adolescent + Dark Skin Tone** reaches 3.36× ($n = 2$).

4.3 RQ3: Media Reporting of Intersectional Harms

Of the 411 causally verified identity markers, Explicit Oppressed markers constitute the largest share (50.4%, $n = 207$). However, 41.4% ($n = 170$) of oppressed-group markers are Inferred, causally present but never named in press coverage, yielding a Media Erasure Index of $E = 0.472$. Privileged-group identities account for only 8.2% of the corpus combined ($n = 34$), reflecting that workplace AI harms in this dataset overwhelmingly affect structurally disadvantaged groups.

This erasure is most pronounced for class. *Lower Class* is the single most frequently inferred marker ($n = 104$), yet appears in only 10 explicit reports, a gap of 94 cases. This contrasts sharply with race and gender coverage (*People of Colour*: 96 explicit, 41 inferred; *Female*: 48 explicit, 9 inferred), suggesting that economic vulnerability is substantially less visible in public discourse than racial or gender-based framings. Notably, *Education = Elite Graduate* ($n = 35$) and *Skin Tone = Dark* ($n = 23$) also rank among the most frequently inferred but absent markers. Full marker-level breakdowns are reported in Tables 6 and 7, with outlet-level coverage in Table 8.

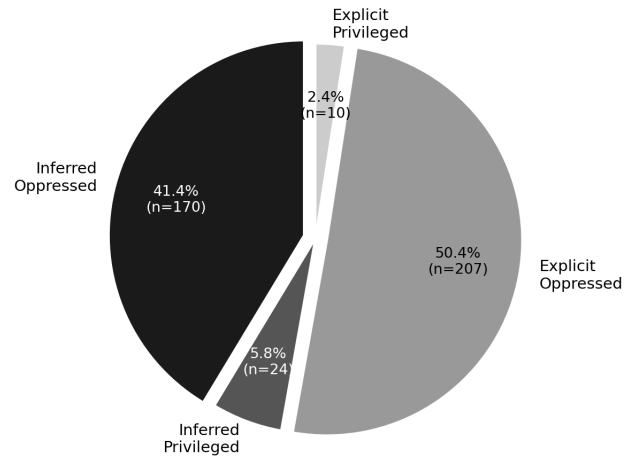


Figure 5: Distribution of causally verified identity markers by reporting status and power position ($N = 411$ markers). Nearly half of all oppressed-group markers are *Inferred*, present as causal factors but never named in the press, while *Explicit Privileged* markers account for only 2.4% of the total.

4.4 RQ4: Coverage of Oppressed vs. Privileged Groups

Across all analysed outlets, 86.4% of identity markers are associated with oppressed groups, compared to 13.6% linked to privileged groups. Vice.com (100% oppressed-group coverage, 32 reports) and ProPublica (100%, 11 reports) show the strongest concentration of marginalised-subject reporting among outlets with sufficient volume (Table 8).

However, this visibility does not imply intersectionally complete representation. When read alongside the Media Erasure Index, the data suggest that outlets frequently acknowledge harms affecting oppressed populations while omitting the specific interacting identities that shape the mechanism of harm. Public narratives thus tend to simplify structurally intersectional incidents into narrower, single-axis explanations.

5 Discussion

Causal rigour matters. By filtering on CQ1+CQ2, we exclude co-occurrence artefacts and retain only incidents where an identity marker is a plausible cause rather than incidental context. This reduces dataset size but increases analytic validity: a racial bias in loan approval is meaningfully different from a loan-denial report that *happens* to involve a person of colour.

Class invisibility. The finding that Lower Class is the second most prevalent causal factor yet the least explicitly reported is substantively important. Economic class lacks the legal protection

status of race or gender in most jurisdictions, and our data indicate it is implicated in as many harmful AI incidents as race. Failing to account for economic vulnerability in AI audits effectively allows automated systems to reproduce class stratification without public oversight or regulatory scrutiny.

Intersectional amplification as a pattern, not a proof. The top-10 pairs all exceed the 1.0× independence baseline, with *Disabled + Older Adult* reaching 5.94× and *Single Parent + Female* reaching 5.52×. However, the bottom-10 pairs tell a different story: all fall below baseline, with *Lower Class + Female* (0.3×), *Lower Class + People of Colour* (0.5×), and *Female + People of Colour* (0.6×) showing substantially *suppressed* co-occurrence relative to what independence would predict. The dataset therefore does not exhibit uniform amplification; rather, it shows a divergent pattern in which some intersections are associated with higher-than-expected harm frequency while others, including some of the most socially salient combinations, appear less often than marginal rates would suggest. This divergence is itself difficult to interpret without population-level demographic denominators: identity categories are correlated in the real-world workforce (disability and older age co-occur more than chance; lower class and single parenthood co-occur more than chance), meaning observed co-occurrence rates reflect demographic base rates as much as any algorithmic targeting. The results should therefore be read as descriptive of this incident dataset, not as population-level causal claims in either direction.

Limitations and future work. Our corpus is predominantly English-language and concentrated in US and UK media. The total of $N = 285$ incidents is small for intersectional analysis, resulting in sparse counts for most identity value pairs: the majority of two-way intersections are observed in fewer than three incidents, making three-way or higher-order intersections entirely infeasible. Amplification scores for rare pairs ($n < 3$) are reported in the Appendix but should be treated as illustrative rather than statistically reliable.

LLM-assisted coding introduces two additional risks: model hallucination and the potential reproduction of biases encoded in training data. We partially mitigated these through human validation ($\kappa = 0.81$), though human annotators were trained using the same extraction prompt as the model, which may overstate true inter-annotator independence.

Finally, we initially aimed to examine whether high-profile incidents exhibit greater intersectional complexity than less-covered cases. Data sparsity within our workplace corpus prevented statistically meaningful analysis of this question, which remains a productive direction for future research with larger datasets.

6 Conclusion

This paper presents a causal intersectional framework for analysing AI harms in workplace settings, combining counterfactual identity verification with intersectional amplification analysis. The results suggest that workplace AI harms are rarely confined to a single demographic axis; they tend to emerge through overlapping social structures that conventional single-attribute fairness evaluations do not capture.

Beyond amplification, the study identifies a second structural problem: the identities most causally central to algorithmic harm

are frequently the least visible in press coverage. Class-based vulnerabilities in particular remain obscured despite their consistent causal presence across incidents.

Taken together, these findings suggest that current AI governance frameworks are constrained by a predominantly single-axis conception of fairness. Addressing workplace AI harms therefore requires not only more intersectional technical auditing, but also reporting standards and regulatory frameworks capable of recognising compounded forms of vulnerability, and naming them explicitly.

References

- [1] Anonymous Author(s). 2026. Why AI Harms Can't Be Fixed One Identity at a Time: What 5300 Incident Reports Reveal About Intersectionality. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT '26)*. ACM, New York, NY, USA.
- [2] Kimberlé Crenshaw. 1989. Demarginalizing the intersection of race and sex. *University of Chicago Legal Forum* 1989, 1 (1989), 139–167.
- [3] Solon Barocas and Andrew D. Selbst. 2016. Big data's disparate impact. *California Law Review* 104 (2016), 671–732.
- [4] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *Proceedings of ITCS 2012*. ACM, 214–226.
- [5] Michael Kearns, Seth Neel, Aaron Roth, and Zhiwei Steven Wu. 2018. Preventing fairness gerrymandering: Auditing and learning for subgroup fairness. In *Proceedings of ICML 2018*.
- [6] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of FAT* 2018*. 77–91.
- [7] Safiya Umoja Noble. 2018. *Algorithms of Oppression*. NYU Press.
- [8] Sean McGregor. 2021. Preventing repeated real world AI failures by cataloguing incidents: The AI Incident Database. In *Proceedings of AAAI 2021*.
- [9] Virginia Eubanks. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.
- [10] Cathy O'Neil. 2016. *Weapons of Math Destruction*. Crown Publishers.
- [11] Kyra Wilson and Aylin Caliskan. 2025. Gender, race, and intersectional bias in resume screening via language model retrieval. In *Proceedings of AAAI 2025*. AAAI Press, 1578–1590.
- [12] Rifat Ara Shams, Didar Zowghi, and Muneera Bano. 2025. AI for all: Identifying AI incidents related to diversity and inclusion. *Journal of Artificial Intelligence Research* 83 (2025). 10.1613/jair.1.17806
- [13] Pierre Bourdieu. 1998. *On Television*. New Press.

A LLM Causal Extraction Prompt

The following system prompt was used to extract identity markers and assess intersectional AI harms from raw incident reports. The prompt explicitly embeds Kimberlé Crenshaw's intersectionality theory and enforces the double-counterfactual logic (CQ1/CQ2) to ensure strict causal attribution.

You are an expert AI Incident Analyst specializing in the application of Kimberlé Crenshaw's intersectionality theory to AI incident reports. You analyze harms using causal and structural reasoning, tracing how system design choices or detection failures contributed to the outcome.

INCIDENT ID: {incident_id}
 INCIDENT TITLE: {incident_title}
 INCIDENT DESCRIPTION: {incident_description}

REPORTS:
 {reports_text}

TASK 1 – Identify harmed subjects
 - Extract every living entity harmed by the AI system.
 - Exclude organizations and inanimate objects.
 - Assign identity markers separately for each subject.
 - Merge duplicated subjects across reports.

TASK 2 – Extract identity markers
 For each subject, identify markers from the following categories:
 {IDENTITY_CATEGORIES}

Extraction rules:
 - Explicit: directly stated in the report.
 - Inferred: clearly implied by a specific detail in the report.
 - Uncertain markers must not be included.

For each marker, apply the following counterfactual questions:

CQ1:
 "Did this incident happen because the AI Subject was [identity]?"
 - Yes only if the system behavior changed because of this identity.
 - No if the identity is only contextual information.

CQ2:
 "Would this incident still have happened if the AI Subject was not [identity]?"
 - No if a person with a different identity would not have experienced the harm.
 - Yes if the harm would have occurred regardless of identity.

Guidelines:
 - Do not be overly conservative.
 - Include identity markers whenever they clearly shaped the AI system's behavior toward the subject.

For marker_type use only:
 - "Explicit"
 - "Inferred"

For power_position use only:
 - "Privileged"
 - "Oppressed"

TASK 3 – Assess deployer
 Determine whether a company or organization deployed the AI system.

Return only valid JSON using the following structure:

```
{
  "incident_id": "{incident_id}",
  "incident_title": "{incident_title}",
  "description": "...",
  "deployer": {
    "is_company": "Yes or No",
    "name": "company name or Unknown"
  },
  "sources": [],
  "subjects": [
    {
      "name": "...",
      "type": "Individual / Group of persons / Society",
      "identity_markers": {
        "category": {
          "marker": "...",
          "marker_type": "Explicit or Inferred",
          "power_position": "Privileged or Oppressed",
          "source": "...",
          "DirectScore": "Yes or No",
          "AlternateScore": "Yes or No",
          "reasoning": "...",
          "MarkerHarm": "..."
        }
      }
    }
  ]
}
```

Critical rules:
 - One value per category per subject.
 - Do not append "(O)" or "(P)" to marker values.
 - Inferred markers require explicit evidence from the report.
 - DirectScore=Yes only if identity directly shaped the harm.
 - AlternateScore=No only if a different identity would have changed the outcome.
 - Merge duplicated subjects across reports.
 - Count all reports as a single incident.
 - Return only JSON, without explanations or markdown.

B RQ1, Full Category Prevalence Data

Tables 2 and 3 give the complete frequency distribution across all 21 identity categories and the top-20 specific identity values (CQ1=Yes, CQ2=No, $N = 285$).

Table 2: All 21 identity categories causally linked to workplace AI harm.

| Category | N | % |
|-------------------|----|------|
| Race | 64 | 22.5 |
| Class | 60 | 21.1 |
| Age | 55 | 19.3 |
| Education | 38 | 13.3 |
| Gender | 35 | 12.3 |
| Health Status | 20 | 7.0 |
| Ability | 16 | 5.6 |
| Family Status | 13 | 4.6 |
| Political Id. | 13 | 4.6 |
| Skin Tone | 12 | 4.2 |
| Nationality | 12 | 4.2 |
| Language | 8 | 2.8 |
| Heritage | 6 | 2.1 |
| Immig. Status | 6 | 2.1 |
| Religion | 5 | 1.8 |
| Neurodiversity | 3 | 1.1 |
| Indigeneity | 2 | 0.7 |
| Gender Identity | 2 | 0.7 |
| Body Size | 1 | 0.4 |
| Gender Expression | 1 | 0.4 |
| Appearance | 1 | 0.4 |

Table 3: Top 20 specific identity values causally linked to workplace AI harm.

| Category = Value | N | % |
|-------------------------|----|------|
| Race = People of colour | 67 | 23.5 |
| Class = Lower class | 53 | 18.6 |
| Gender = Female | 33 | 11.6 |
| Edu. = Elite grad | 17 | 6.0 |
| Ability = Disabled | 16 | 5.6 |
| Edu. = Student | 15 | 5.3 |
| Age = Adolescent | 14 | 4.9 |
| Skin Tone = Dark | 13 | 4.6 |
| Age = Young adult | 12 | 4.2 |
| Age = Child | 12 | 4.2 |
| Age = Older adult | 12 | 4.2 |
| Health = Chron. ill | 10 | 3.5 |
| Class = Middle class | 9 | 3.2 |
| Pol. Id. = Activist | 7 | 2.5 |
| Family = Caregiver | 6 | 2.1 |
| Health = Mental hlth. | 6 | 2.1 |
| Age = Adult | 5 | 1.8 |
| Edu. = Self-taught | 5 | 1.8 |
| Lang. = ESL | 5 | 1.8 |
| Health = Phys. disabled | 5 | 1.8 |

C RQ2, Full Intersectional Data

Table 4 lists the top-15 category-level co-occurrence pairs. Table 5 reports the highest and lowest value-pair amplification scores. Pairs marked with [†] have $n < 3$ and are statistically unstable.

Table 4: Top 15 category pairs by co-occurrence (RQ2).

| Pair | N | % |
|-----------------------|----|-----|
| Age + Education | 12 | 4.2 |
| Race + Skin Tone | 12 | 4.2 |
| Age + Race | 11 | 3.9 |
| Age + Gender | 8 | 2.8 |
| Gender + Race | 8 | 2.8 |
| Class + Race | 6 | 2.1 |
| Class + Family Status | 5 | 1.8 |
| Ability + Health | 5 | 1.8 |
| Ability + Age | 5 | 1.8 |
| Age + Class | 5 | 1.8 |
| Family + Gender | 4 | 1.4 |
| Education + Gender | 3 | 1.1 |
| Heritage + Race | 3 | 1.1 |
| Gender + Health | 3 | 1.1 |
| Ability + Class | 3 | 1.1 |

Table 5: Value-pair amplification scores (Top 10 and Bottom 10, $n \geq 2$).

| Value Pair | O | E | Amp |
|--------------------------------|---|-------|----------------------|
| Top 10 | | | |
| Adult + Elite graduate | 2 | 0.30 | $6.7 \times^\dagger$ |
| Disabled + Older adult | 4 | 0.67 | $5.9 \times$ |
| Single parent + Female | 3 | 0.58 | $5.2 \times$ |
| Lower class + Single parent | 3 | 0.93 | $3.2 \times$ |
| Adolescent + Dark skin tone | 2 | 0.64 | $3.1 \times^\dagger$ |
| Young adult + Elite graduate | 2 | 0.72 | $2.8 \times^\dagger$ |
| Male + People of colour | 2 | 0.94 | $2.1 \times^\dagger$ |
| Adolescent + Female | 3 | 1.62 | $1.9 \times$ |
| Female + Chronically ill | 2 | 1.16 | $1.7 \times^\dagger$ |
| Lower class + Activist | 2 | 1.30 | $1.5 \times^\dagger$ |
| Bottom 10 | | | |
| Young adult + People of colour | 3 | 2.82 | $1.1 \times$ |
| Adolescent + People of colour | 3 | 3.29 | $0.9 \times$ |
| Young adult + Lower class | 2 | 2.23 | $0.9 \times^\dagger$ |
| Older adult + Lower class | 2 | 2.23 | $0.9 \times^\dagger$ |
| Older adult + People of colour | 2 | 2.82 | $0.7 \times^\dagger$ |
| Disabled + Lower class | 2 | 2.98 | $0.7 \times^\dagger$ |
| Female + People of colour | 5 | 7.76 | $0.6 \times$ |
| Student + People of colour | 2 | 3.53 | $0.6 \times^\dagger$ |
| Lower class + People of colour | 6 | 12.46 | $0.5 \times$ |
| Lower class + Female | 2 | 6.14 | $0.3 \times^\dagger$ |

$^\dagger n < 3$, statistically unstable.

D RQ3, Full Media Representation Data

Of 411 total identity markers, 217 are Explicit and 194 are Inferred ($E = 52.8\%$; Erasure Score = 47.2%). Tables 6 and 7 list the top-15 markers in each group. Table 8 reports oppressed-group coverage per outlet (86.4% overall; outlets with fewer than 5 oppressed-group markers omitted).

Table 6: Top 15 Explicit markers (directly named in press).

| Marker | N |
|-----------------------------|----|
| Race = People of colour | 96 |
| Gender = Female | 48 |
| Education = Student | 33 |
| Age = Older adult | 19 |
| Age = Child | 18 |
| Ability = Disabled | 18 |
| Age = Adolescent | 17 |
| Age = Young adult | 16 |
| Geography = Urban | 11 |
| Health = Chronically ill | 11 |
| Class = Lower class | 10 |
| Pol. Id. = Activist | 9 |
| Health = Mental hlth. cond. | 8 |
| Age = Adult | 8 |
| Gender = Male | 8 |

Table 7: Top 15 Inferred markers (causally present, not reported).

| Marker | N |
|--------------------------|-----|
| Class = Lower class | 104 |
| Race = People of colour | 41 |
| Edu. = Elite grad | 35 |
| Skin Tone = Dark | 23 |
| Geography = Urban | 22 |
| Class = Middle class | 19 |
| Health = Chronically ill | 12 |
| Ability = Disabled | 12 |
| Edu. = Self-taught | 10 |
| Age = Adolescent | 10 |
| Gender = Female | 9 |
| Age = Young adult | 8 |
| Language = ESL | 7 |
| Age = Older adult | 7 |
| Age = Child | 6 |

Table 8: Oppressed-group coverage per outlet, RQ4 ($n \geq 5$ oppressed markers; 86.4% overall).

| Outlet | Oppressed | % |
|----------------------|-------------|-------------|
| theguardian.com | 57 | 86 |
| nytimes.com | 57 | 89 |
| washingtonpost.com | 45 | 85 |
| theverge.com | 41 | 89 |
| wired.com | 38 | 93 |
| vice.com | 32 | 100 |
| bbc.com | 25 | 86 |
| arstechnica.com | 24 | 96 |
| futurism.com | 22 | 81 |
| gizmodo.com | 18 | 86 |
| apnews.com | 17 | 81 |
| cbsnews.com | 16 | 73 |
| reuters.com | 15 | 83 |
| theregister.com | 14 | 88 |
| independent.co.uk | 13 | 93 |
| technologyreview.com | 13 | 93 |
| techcrunch.com | 13 | 87 |
| businessinsider.com | 12 | 63 |
| cnbc.com | 12 | 92 |
| nbcnnews.com | 12 | 71 |
| propublica.org | 11 | 100 |
| bloomberg.com | 11 | 92 |
| usatoday.com | 10 | 77 |
| telegraph.co.uk | 10 | 67 |
| techdirt.com | 10 | 91 |
| wsj.com | 9 | 90 |
| mirror.co.uk | 9 | 100 |
| cnn.com | 9 | 75 |
| aclu.org | 8 | 100 |
| forbes.com | 8 | 80 |
| qz.com | 8 | 89 |
| thehill.com | 8 | 67 |
| biometricupdate.com | 8 | 100 |
| thetimes.co.uk | 8 | 73 |
| npm.org | 7 | 54 |
| theconversation.com | 7 | 88 |
| boingboing.net | 7 | 70 |
| news.com.au | 7 | 100 |
| abc.net.au | 7 | 100 |
| statnews.com | 7 | 100 |
| insidehighered.com | 6 | 100 |
| algorithmwatch.org | 6 | 100 |
| restofworld.org | 6 | 100 |
| Overall | 1592 | 86.4 |