Informing AI Risk Assessment with News Media: Analyzing National and Political Variation in the Coverage of AI Risks

Mowafak Allaham[1], **Kimon Kieslich** [2], **Nicholas Diakopoulos**[1]

[1]Northwestern University
[2]Institute for Information Law, University of Amsterdam
mowafakallaham2021@u.northwestern.edu, k.kieslich@uva.nl, nad@northwestern.edu

Abstract

Risk-based approaches to AI governance often center the technological artifact as the primary focus of risk assessments, overlooking systemic risks that emerge from the complex interaction between AI systems and society. One potential source to incorporate more societal context into these approaches is the news media, as it embeds and reflects complex interactions between AI systems, human stakeholders, and the larger society. News media is influential in terms of which AI risks are emphasized and discussed in the public sphere, and thus which risks are deemed important. Yet, variations in the news media between countries and across different value systems (e.g. political orientations) may differentially shape the prioritization of risks through the media's agenda setting and framing processes. To better understand these variations, this work presents a comparative analysis of a cross-national sample of news media spanning 6 countries (the U.S., the U.K., India, Australia, Israel, and South Africa). Our findings show that AI risks are prioritized differently across nations and shed light on how left vs. right leaning U.S. based outlets not only differ in the prioritization of AI risks in their coverage, but also use politicized language in the reporting of these risks. These findings can inform risk assessors and policy-makers about the nuances they should account for when considering news media as a supplementary source for risk-based governance approaches.

Extended version — https://arxiv.org/abs/2507.23718

1 Introduction

AI is increasingly integrated into various systems and applications that serve millions of users globally, despite its potential for significant risks to both individuals and society (Bommasani et al. 2021; Park et al. 2024; Burtell and Woodside 2023). In response, governments, companies, and researchers have contributed regulatory frameworks, (Madiega 2021; Biden 2023), risk assessments (Solaiman et al. 2023; Metcalf et al. 2021; Allaham, Kieslich, and Diakopoulos 2024; Nanayakkara, Hullman, and Diakopoulos 2021) and assessment methods, such as red-teaming and safety benchmarks (Ganguli et al. 2022; Mazeika et al. 2024; Zeng et al. 2024; Zhang et al. 2023), to help govern (Reuel et al. 2024), anticipate (Kieslich, Diakopoulos, and Helberger 2024; Hautala and Heino 2023; Avin, Gruetzemacher, and Fox 2020), and potentially mitigate such risks. Risk-based approaches (van der Heijden 2021) to managing potential harms from AI have come to dominate, yet such approaches often center the technological artifact as the primary focus of risk assessments, overlooking systemic risks of AI that see risk arising as a complex interaction between the AI system, human stakeholders (e.g. users but also impacted people), and the larger societal context which might include dimensions of culture, government, institutions, and more (Kieslich, Helberger, and Diakopoulos 2025; Weidinger et al. 2023; Uuk et al. 2024).

One route to incorporate more societal context into risk-based approaches is to leverage the news media. As observed in AI incident tracking initiatives such as the AI Incident Database and OECD AI Incidents Monitor (McGregor 2021; Filippucci et al. 2024; Diakopoulos 2025), news coverage, including the reporting of failures, biases, and broader impacts of algorithmic and AI systems on the public (Diakopoulos 2015a) can be leveraged as a source to help map AI risks in real-world contexts. In particular news media functions to identify and articulate risks and harms in the full complexity of society, emphasizing the socio-technical interactions around AI systems (Diakopoulos 2025). At the same time, news media and the journalistic processes underlying it, reflects its own set of normative and other biases about what is prioritized for coverage and how AI is covered (Chuan, Tsai, and Cho 2019; Nguyen and Hekman 2024). Moreover, news media plays an important role in shaping the narrative around AI risks that are relevant to the general public and the perception of AI by various stakeholders including the public and policy-makers (Gilardi et al. 2024). Thus, differences in the news media between societies (e.g. as indicated by country) and across different value systems (e.g. political orientations) may differentially shape the identification and perception of harms through agenda setting and framing processes (Ouchchy, Coin, and Dubljević 2020; Sun et al. 2020a; Brennen, Howard, and Nielsen 2018; McCombs and Shaw 1972; Scheufele 1999).

Owing to the role of news media in shaping how risks are perceived by society, in this work we posit that risk assessors, policy-makers, and third-party auditors who evaluate and address the negative impacts of AI through risk-

based and regulatory means should be aware of and account for media variance, including national and political nuances, as they consider news media in risk-based regulatory approaches. To shed light on how these details may influence the reporting of AI risks, this work analyzes a sample of news media to examine the prevalence of AI risks, making cross-national comparisons of news reporting between countries from the Global North and Global South, and analyzing the role of political orientation of news outlets in shaping the coverage of AI risks in the U.S. context specifically.

Using the domain taxonomy of AI risks from the the MIT Risk Repository (Slattery et al. 2024) – a repository of AI risks synthesizing 56 taxonomies that categorizes AI risks by their cause and risk domain – we analyze the prevalence of AI risks reported in a sample of news articles published in English by news outlets spanning 6 countries from around the world: the United States of America, the United Kingdom, India, Australia, Israel, and South Africa. Furthermore, we examine how the prevalence and coverage of AI risks varies across U.S. news outlets with different political orientations, using domain-level ratings from Media Bias Fact Check, an independent website maintained by researchers and journalists that relies on human fact-checkers affiliated with the International Fact-Checking Network to evaluate media sources along different dimensions including political bias (Lin et al. 2023).

Through a comparative analysis of the prevalence of AI risks reported in our sample, we show how news media tend to emphasize the coverage of specific risks such as *Socioeconomic & Environmental* risks overall, but place less attention on other risks such as *Human-Computer Interaction* risks. We also illustrate the heterogeneous pattern of news coverage for AI risks showing notable and significant associations between countries and AI risks reported in articles published by news outlets in these countries. For instance, even among the most prevalent risk category in our sample, the proportion of articles covering *Socioeconomic & Environmental* risks is substantially less in Israeli and Indian outlets as compared to those in the United States, the United Kingdom, South Africa, and Australia. Furthermore, we illustrate how the prioritization and communication of AI risks are influenced and shaped by the political orientation of news outlets in our sample of articles from the U.S. Specifically, we show how *Malicious Actors & Misuse* risks are the most salient risks reported on by right-biased news outlets compared to *Socioeconomic & Environmental* risks by left-biased outlets. In addition, as part of this analysis, we share examples illustrating the use of politicized language in the reporting of AI risks by these outlets.

As news media continue to play a crucial role in highlighting risks and harms relevant to the general public–a key stakeholder in shaping the current and future development of AI regulations and public policy, especially in democratic countries– and shaping their perception of AI, this work explicitly focuses on examining the influence of national and political variations embedded in journalistic practices on the reporting of AI risks in news media. By accounting for such variations, we unravel insights about which of the AI risks identified by the MIT domain taxonomy of AI risks are dominating the public discourse (and which are not), per our cross-national comparison of the prevalence of AI risks in news media. In addition, by leveraging our sample of articles from the U.S., we illustrate the association between political orientations of news outlets and the risks covered by these outlets.

Our findings contribute to informing risk assessors, policymakers, and researchers about (1) dimensions that should be accounted for when considering news media as part of risk-assessment, incidents monitoring, and regulatory practices, and (2) the emerging politicized language around AI risks that may influence the public perception of AI and hinder progress on current and future development of AI regulations and policies aiming to make AI systems and technologies more inclusive and safe.

2 Related Work on Media Coverage of AI

The news media play an influential role in shaping the national and public discourse on AI by helping to set the standards and expectations for AI accountability (Diakopoulos 2025). In the traditional understanding of communication science, the news media function as agenda setters (McCombs and Shaw 1972). A key task of the news media is to inform a broad public about politically and socially relevant issues (Schäfer 2017) and thereby ensure a plurality of voices, i.e. the inclusion of various societal stakeholders. *How* the media portray these technologies is consequential, as media coverage has been shown to influence public opinion (Nisbet et al. 2002; Scheufele and Lewenstein 2005), especially for novel technologies like AI (Hilgard and Li 2017) – and public opinion plays a crucial role in technology adoption. On the one hand, citizens act as consumers of AI technology, and the media's portrayal of AI can influence whether or not people are willing to use the technology (Ouchchy, Coin, and Dubljević 2020). On the other hand, citizens can act as voters and thus influence regulatory aspects (Ouchchy, Coin, and Dubljević 2020; Kieslich, Lünich, and Došenović 2023; Kieslich 2024).

One of the main principles of journalistic news quality is to represent a plurality of voices that are relevant to the discourse, and the inclusion of these voices (e.g., activists, academics, civilians, NGOs) can enrich the discourse on AI (Brennen 2018). In particular, when it comes to reporting on AI risks, the efforts of investigative journalists have helped shed light on pressing issues such as the child benefit scandal in the Netherlands (Constantaras et al. 2023) or the COMPAS recidivism algorithm in the US (Angwin et al. 2022). Indeed, scholars first articulated the idea of "algorithmic accountability" as stemming from investigations published in the news media (Diakopoulos 2015b), and more recently have argued for the inclusion of journalists in third-party audits of AI systems, as they "were responsible for uncovering deeply-rooted socio-technical harms in algorithmic systems related mainly to representational harms due to discriminatory design choices." (Hartmann et al. 2024). This is supported by the fact that many of the sources of the AI Incident Database (McGregor 2021) are newspaper articles. As a result, news coverage of AI risks plays a key role in exposing

the risks of AI systems. Unlike self-reporting by companies (including their research teams), journalists are structurally independent and can uncover novel impacts that may conflict with corporate goals. Their inclusion "ensure[s] social accountability through domain knowledge and special access to affected communities" (Hartmann et al. 2024).

Another important factor when considering the impact of discourse on public perception is its politicization. Scholarly research in this area states that politicization of an issue requires three conditions (De Wilde 2011; Schattschneider 1957): (1) polarization of the issue, i.e., whether and how prevalent different (political) positions are on the issue. This could also be achieved by different framing or agenda setting of topics related to the issue (e.g. different prevalence of AI risks, or different positions on individual AI issues). (2) The intensity of media coverage. This refers to the visibility of an issue. The more it is covered, the more relevance is attributed to the issue. And (3) The resonance of the issue, i.e. how relevant the issue is in the eyes of the public. Media coverage plays a key role in this regard, as it provides an important arena in which AI is discussed. Several studies of media coverage have found a sharp increase in news coverage of AI in recent years (Fast and Horvitz 2017; Vergeer 2020; Ouchchy, Coin, and Dubljević 2020; Ittefaq et al. 2025; Chuan, Tsai, and Cho 2019), which satisfies the condition of intensity of coverage. Several scholars have also analyzed the influence of the political leaning of the news outlet on the framing of AI – with mixed results (Brennen 2018; Roe and Perkins 2023; Vergeer 2020). For example, for the UK case, Brennen (2018) state that: "Right-leaning outlets highlight issues of economics and geopolitics, including automation, national security, and investment", whereas "Left-leaning outlets highlight issues of ethics of AI, including discrimination, algorithmic bias, and privacy." However, when Roe and Perking looked at headlines about ChatGPT and AI in general a few years later, they didn't find any evidence of strong polarization (Roe and Perkins 2023). For the Dutch case, Vergeer (2020) reported that business newspapers were generally more favorable to AI than national newspapers. However, the politicization of the AI *risk debate* in particular has not been explored. Analyzing the politicization of AI risks in terms of political positions is important because it reflects political strategies in terms of regulation or policy enforcement. Furthermore, it shows how citizens who consume politically biased news are informed and perceive the AI risk discourse.

Recognizing the importance of the news media in relation to AI, a significant number of scholars have focused on analyzing how the news discusses AI (Brennen 2018; Chuan, Tsai, and Cho 2019; Fast and Horvitz 2017; Ouchchy, Coin, and Dubljević 2020; Kieslich, Došenović, and Marcinkowski 2022; Sun et al. 2020b; Vergeer 2020; Zeng et al. 2024; Brennen, Howard, and Nielsen 2022; Meißner 2024; Roe and Perkins 2023; Ittefaq et al. 2025; Nguyen and Hekman 2022, 2024; Bunz and Braghieri 2022). However, most studies of media coverage focus on countries in the Global North, such as the US (Chuan, Tsai, and Cho 2019; Fast and Horvitz 2017), the UK (Brennen 2018; Brennen, Howard, and Nielsen 2022; Roe and Perkins 2023), Germany (Meißner 2024; Kieslich, Došenović, and Marcinkowski 2022), the Netherlands (Vergeer 2020), or take a comparative approach between the US and the UK (Nguyen and Hekman 2024; Bunz and Braghieri 2022) or the US and China (Nguyen and Hekman 2022). Only a few studies focus on non-Western countries, with the exception of China (Zeng et al. 2024) and a comparative study of 12 countries, including countries of the Global North and the Global South (Ittefaq et al. 2025). Thematically, studies on media coverage mostly focus on mapping the general discourse on AI, for example by analyzing the thematic structure or sentiment of media discourse (e.g. Ittefaq et al. (2025); Chuan, Tsai, and Cho (2019); Brennen (2018)), while rarely focusing explicitly on risks or negative impacts (with the exception of (Ouchchy, Coin, and Dubljević 2020; Nguyen and Hekman 2024; Chuan, Tsai, and Cho 2019)).

Overall, we find that the extant literature on news analysis of AI doesn't tend to engage extensively with AI *risks* and that there is an opportunity to more fully leverage news sources in AI risk assessment practices by incorporating this focus on risk and its intersection with national and political variations. This paper addresses these opportunities by (1) explicitly focusing on AI risks in the study of news content, (2) using news media as a source to inform risk assessments by analyzing the prevalence of AI risks covered, (3) taking a cross-national perspective in analyzing a sample of news articles published by news outlets from a few countries in the Global North and Global South, thus contributing to the inclusion of more diverse perspectives and analysis of AI risks, and (4) analyzing the effect of political positioning of news coverage on the prevalence of AI related risks.

3 Data

To establish a dataset of news articles related to AI, we base our selection of articles on the following three criteria: (1) data availability from different countries, (2) the goal of maintaining geographic diversity in our sample, and (3) a preference for outlets publishing articles in English to facilitate evaluation and analysis of articles by the English-speaking authors. Based on these considerations, we selected articles from the following countries for analysis: the U.S., the U.K., India, Australia, Israel, and South Africa. Although domains in our sample include some of the most read news outlets, as well as others, in each country (per Newman et al. (2024)), we recognize that our selection criteria is likely to exclude other countries and their news outlets that could be of interest for this research as further elaborated on in the Limitations section (see Section 7).

Using GDELT (Leetaru and Schrodt 2013) (Global Data on Events, Location and Tone Project), we collected online news articles published in English by outlets in these countries between January 2022 and October 2024, giving enough time to capture the coverage of several emerging AI technologies at the time (e.g., ChatGPT) and their implications on society. We chose to source our sample of articles from GDELT because it captures and provides an extensive coverage of what is reported on in news media across different news outlets, languages, and offers an accessible way of querying such coverage via an API (Leetaru and

Schrodt 2013; Ward et al. 2013). In addition, it presents an alternative to scraping news portals and aggregators, such as Google News, that display or rank news content that is recently published by popular news outlets, politically biased (i.e., slight leftward bias), or limited in exposure to cross-national perspectives (Nechushtai, Zamith, and Lewis 2024; Ulken 2005; Hernandes and Corsi 2024).

To retrieve AI-related articles from GDELT, we first developed a set of data-driven keywords of relevance. To do this we first retrieved articles from The New York Times (NYT) using two seed search words ("A.I." and "Artificial Intelligence") and extracted a list of the most prevalent n-grams from each retrieved article. By manually selecting n-grams with the highest frequency that are relevant to AI, we identified a total of 31 keywords spanning numerous topics related to AI. To further expand the span of coverage of the identified keywords for AI systems and technologies, we also scraped the full text of 2,724 articles associated with 529 incidents between January 2017 and June 2023 that are relevant to AI from the AI Incident Database (McGregor 2021), which curates news items and other reports indicating AI failures around the world. Using the same n-gram extraction method mentioned earlier, we identified nine new keywords that didn't overlap with the 31 already found. This brought the total number of the curated keywords up to 40. A full list of the keywords is provided in the Appendix (A.1). Next, we used GDELT's v2 API endpoint to query for each of the AI-relevant keywords and retrieve the URLs and metadata of the daily published news articles mentioning that keyword.

In total, we retrieved 921,057 URLs for online articles published by 244 unique news domains spanning the U.S., the U.K., India, Australia, Israel, and South Africa. To ensure that our sample excludes content from domains that aggregate news or distribute press-releases (e.g., prnewswire.com), we filtered the data retrieved from GDELT based on a curated list of 1,064 news domains spanning 177 countries compiled in the Global English Language Sources list by Media Cloud (MC) (Roberts et al. 2021; Media Cloud 2025), which also includes domains from the countries included in our research. After applying the filter, our sample included 178,172 URLs across all six countries. Using a custom web scraper that leverages the `newspaper` library (Ou-Yang 2025), we successfully scraped 31,252 articles (17.5% of 178,172 articles) from 115 news domains spanning all six countries. However, we could not scrape the remaining articles due to either missing content (i.e., 404 errors) or content being blocked behind pay- or sign up walls. The resulting sample thus reflects media that is freely and perhaps more broadly accessible online and via social media than might be the case if we were able to include more gated media sources (see Section 7 for further discussion of this).

4 Methods

To prepare our sample for analysis, we (1) filter for negative impacts and summarize them (Section 4.1), and then (2) classify these negative impacts according to the MIT domain taxonomy of AI risks using an LLM (Section 4.2). We describe each step in detail in the following sections.

4.1 Filtering & Summarizing negative impacts of AI from news media using an LLM

We follow a zero-shot prompting approach similar to the one reported in previous research with a similar objective of detecting the negative consequences of AI in news media (Pang et al. 2024). To develop our prompt, we referred to prior research on social impact assessments (Becker 2001; Nanayakkara, Hullman, and Diakopoulos 2021) to synthesize the following conceptual definition of an impact of an AI technology that we used to steer the LLM towards identifying articles reporting impacts of AI: *An impact refers to an effect, consequence, or outcome of an AI system (i.e., model or application) that positively or negatively affects individuals, organizations, communities, or society*. We did not limit the conceptual definition to negative impacts per se, so as to provide opportunity for future work that may want to focus on positive impacts of AI technologies (Kieslich, Helberger, and Diakopoulos 2024).

Due to the large number of articles in our corpus, we used GPT-4o [1](OpenAI 2024) to assist in filtering articles reporting on impacts of AI. First, we randomly sampled 300 articles from our corpus for authors to annotate whether it contained at least one impact of AI, based on our definition and found that most articles (77%) had an impact. Next, using this annotated sample and prompt A.7, we used GPT-4o to classify each article as either containing an impact or not. The model performed well on this task achieving an F1-macro score of 0.82. We then applied this prompt using the OpenAI batch API to the rest of the dataset. Out of the 31,252 articles in our sample, as reported in Section 3, GPT-4o classified 20,935 (66.98%) as containing impacts based on the aforementioned definition of an impact.

Summarizing negative impacts of AI in articles. After identifying 20,935 articles as having impacts of AI, we instructed GPT-4o with prompt A.8, including the entire article text as context to the model, to summarize all the negative impacts reported in each article. The result is a list of negative impacts each described by a sentence summarizing the impact. We chose to summarize the negative impacts in articles, rather than extract them verbatim, because we were concerned that quoting only specific sentences might miss additional information that may help contextualize the impacts, which often requires an understanding of the context from the full article. To evaluate the model performance in summarizing negative impacts, we randomly selected 50 articles and the corresponding lists of summarized impacts from these articles. Based on our manual assessment we find that our method captures the granular and specific context of impacts as reported in 48 of the 50 articles. For the remaining two articles, one had negative impacts that the LLM did not identify and the other contained a negative impact, and was annotated by the LLM as such, but it did not fit our definition of impact since the impact wasn't directly linked to an AI system.

[1]Model version: `gpt-4o-2024-08-06`

Applying this process resulted in 36,793 negative impacts of AI that were identified from 12,385 articles sourced from 105 domains spanning 6 countries: the U.S., the U.K., India, Australia, Israel, and South Africa. For descriptive details on the proportion of articles per country and news domain, see Tables A.2 and A.4, respectively.

4.2 Using the MIT risk taxonomy to classify impacts from news media

To classify the negative impacts of AI reported in news media, we manually annotated each reported impact from a random sample of 300 negative impacts from our corpus into one of the seven risk categories defined by the "domain taxonomy" of the MIT AI Risk Repository: *Discrimination & toxicity*, *Privacy & security*, *Misinformation*, *Malicious actors & misuse*, *Human-computer interaction*, *Socioeconomic & environmental harms*, *AI system safety failure & limitations*. A full list of the sub-categories defining these seven domain risk categories can be found in Appendix A.3. Although other expert-driven taxonomies of AI risks exist (Solaiman et al. 2023; Shelby et al. 2023; Weidinger et al.), prior research found that these taxonomies may suffer from inadvertent expert or selection biases and may not be as representative of international perspectives of AI risks and harms (Allaham, Kieslich, and Diakopoulos 2024; Bonaccorsi, Apreda, and Fantoni 2020; Crawford 2016; Hagerty and Rubinov 2019; Jobin, Ienca, and Vayena 2019). Accordingly, rather than relying on a single expert-driven taxonomy, we chose to focus on the MIT Risk Repository because it presents an aggregated perspective into the risks and harms of AI across 56 taxonomies that are sourced from academia, government, and industry and are authored by teams of researchers from several countries around the world (Slattery et al. 2024). The resulting annotated sample using the domain taxonomy of AI risks serves as a baseline to evaluate the capability and performance of the LLM on this classification task.

Next, using a zero-shot prompting approach, we sent requests via the OpenAI API instructing GPT-4o to classify each impact summary from the 300 randomly sampled impacts into only one of the defined domain risk categories of the MIT Risk Repository, as outlined in Prompt A.9. In addition, we included in the prompt an instruction for the LLM to assign an "other" label for impacts that do not fit any of the categories defined in the prompt, mirroring the best practices of thematic analysis (Braun and Clarke 2012). We prompted the model using the sub-domain categories of AI risks to align with the conceptual framework of the domain taxonomy (see A.3), which organizes risks into specific sub-domains, as outlined in the original paper of the MIT Risk Repository (Slattery et al. 2024). For instance, *Human-computer interaction* risks are defined in the domain taxonomy in terms of *Overreliance and unsafe use* and *Loss of human agency and autonomy* risks, which were used in prompt A.9. Accordingly, if GPT-4o deemed an excerpt of text to fit the definition of any of these two sub-categories, the excerpt is considered a risk relevant to Human-computer interaction. Therefore, for our analysis aimed at describing the prevalence of the *categories* of AI risks identified by the MIT Risk Repository and reported in the news media across different countries, we aggregate and present results at the domain, rather than the sub-domain, level of AI risk categories as they are anchored to capture the incremental and expanding nature of the sub-categories of risks that emerge over time with the (mis)use of AI across domains.

Based on the human annotated sample of negative impacts, the performance evaluation of the LLM for classifying impacts from news media into the seven categories of risks from the domain taxonomy of AI risks demonstrates a strong performance by GPT-4o in classifying impacts into their corresponding risk category with a macro-averaged F1-score of 0.90 [2]. To scale up the classification for the remaining summaries of negative impacts in our corpus, we applied the same zero-shot approach described previously. The prevalence of the categories of risks across countries is analyzed and described in the results section.

5 Results

In section 5.1, we present our findings from analyzing the prevalence of AI risks in a cross-national sample of news media at the country-level, and then further analyze the association between these risks and the countries of the news outlets covering them. Next, in section 5.2, we analyze the potential influence of the political orientation of news outlets on the reporting of AI risks, focusing our analysis only on a sample of articles from the U.S.

5.1 Prevalence of AI Risks in a cross-national sample of news media

To measure the prevalence of AI risk categories identified by the MIT domain risk taxonomy in our cross-national sample of news media, we calculate, for each country, the proportion of articles reporting on each risk category relative to the total number of articles from that country (see Table A.6 for details). We find that the prevalence of AI risks in news media tend to vary by risk category and across countries, as illustrated in Figure 1.

In order to examine whether the prevalence of AI risks is associated with the country reporting on these risk, for each risk category, we conducted a Chi-square test of independence on the number of articles covering that risk in each country. Results indicate a statistically significant association between countries and the coverage of AI risks in news articles, with an exception of *Privacy & Security* which was not found to be statistically significant, indicating that the coverage of AI risks varies by country (details of these tests are included below). We elaborate on our findings for each risk category in the subsequent sections.

The proportion of articles covering issues related to increased inequality, economic and cultural impacts, environmental harms, and more as described by the *Socioeconomic*

[2]The per-category F1-scores range from 0.76 (for Human-Computer Interaction) to 1.00 (for AI system safety, failures, & limitations), with most categories achieving F1-scores above 0.85, indicating high overall classification performance across categories.

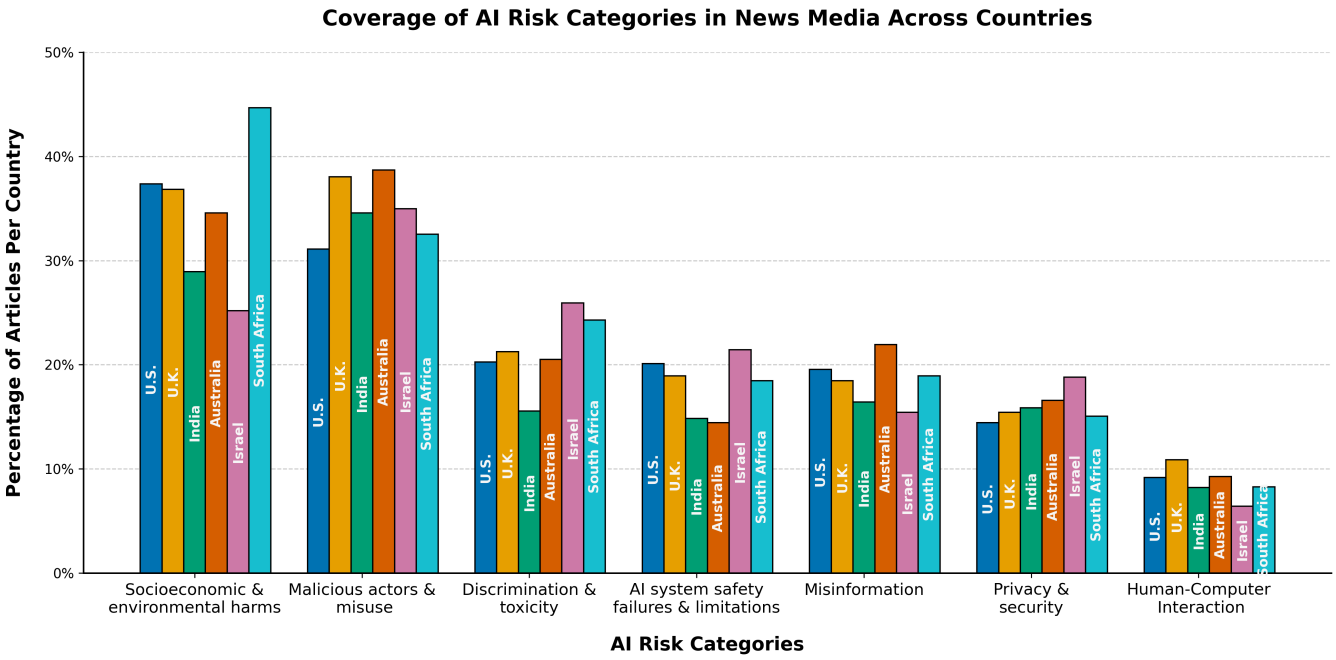

Figure 1: Prevalence of AI risk categories in news coverage across six countries in our sample, based on the proportion of articles reporting each AI risk category in the United States, the United Kingdom, India, Australia, Israel, and South Africa.

& Environmental Risks varied notably, with the highest proportions observed in South Africa (44.6%), United States (37.3%), United Kingdom (36.8%), and Australia (34.5%), while lower proportions were found in India (28.9%) and Israel (25.1%). A Chi-Square test for this risk category was conducted on the number of articles covering this risk yielding a statistically significant result ($\chi^2(5) = 54.76$, $p < 0.001$), indicating that the distribution of articles discussing socio-economic and environmental risks is not independent of country. Rather, the findings suggest that media coverage of socioeconomic and environmental risks vary in the extent to which each country emphasize this risk category, reflecting differences in public discourse or local reporting on socio-economic and environmental harms around AI.

Next, for AI risks relevant to disinformation, cyberattacks, and targeted manipulation that are encompassed by *Malicious Actors & Misuse* risk category, we use Chi-square test of independence to assess the cross-country variation in media coverage of this risk category relative to expected number of articles covering this category per country. The test reveals significant differences in the proportion of articles addressing this risk across countries ($\chi^2(5) = 46.9$, $p < 0.001$). Particularly, the United States had the least coverage of this risk category (31.1%) compared to any other country. In contrast, Australia (38.6%) had the highest proportion of articles focusing on this risk category, very similar to the proportion of articles exhibited in the U.K. (38.0%). All remaining countries, Israel (34.9%), India (34.5%), and South Africa (32.5%), had a more comparable proportion of coverage of this risk to each other. These findings highlight geographic variation in the framing of AI risks, with some countries prioritizing the coverage of potential implications and concerns of the malicious use and misuse of AI more prominently in public discourse than others.

Discrimination & Toxicity Risks - Both Israel (25.9%) and South Africa (24.2%) have the highest proportion of articles covering this risk category, which captures issues pertaining to discrimination, misrepresentation, and exposure to toxic content. This could be attributed to salient historical and political contexts surrounding the debates related to the post-apartheid state in South Africa and the implications of the ongoing conflicts in the Middle East on identity and social cohesion which could contribute to heightened coverage for risks related to AI-generated misrepresentation or AI-powered content moderation algorithms. In comparison, Australia (20.5%), U.K (21.2%), and the U.S. (20.2%) have a comparable emphasis on this category in their news coverage. However, India (15.5%) had the least coverage of this risk category in comparison to other countries in our sample. Similar to the previous risk categories, the Chi-Square test of independence was found to be significant ($\chi^2(5) = 25.96$, $p < 0.001$) showing a consistent pattern, as observed from the other risk categories so far, of varying prioritization in the coverage of risk categories by the news media across different countries.

Despite the prominence of research related to misinformation (and its risks from AI-generated content) as well as the safety of AI systems in possessing dangerous capabilities (Mitchell et al. 2025; Kolt 2025), we observe the two categories of *AI System Safety, Failures, & Limitations* ($\chi^2(5)=28.88$, $p < 0.001$) and *Misinformation Risks*

($\chi^2(5)=12.69$, p=0.0264) not being among the leading risk categories covered by news media. Both Israel (21.4%) and U.S. (20.0%) had the highest proportion of articles in our sample emphasizing AI systems, safety, failures, & limitations in their coverage. As for misinformation risks, Australia (21.9%) had the most prominent coverage of this risk, with a comparable representation of coverage in U.S (19.5%), South Africa (18.9%), and U.K. (18.4%). However, India and Israel had the least proportion of articles covering this risk relative to other countries, accounting for only 16.4% and 15.4% of their respective coverage.

Lastly, *Privacy & Security Risks* and *Human-Computer Interaction Risks* were the least covered risks in our sample across all six countries. Although *Privacy & Security Risks* was the second to last most prevalent category in our sample, the variation in news coverage for this risk between countries was found not be significant based on a Chi- square test. As for the *Human-Computer Interaction Risks*, which reflect risks related to over-reliance, unsafe use, and loss of human agency, it received the least news coverage of AI risks across all countries (9.3%). On average, 8.3% of the news coverage in each country focused on reporting human-computer interaction risks, with the U.K. leading this coverage (10.8%). Despite these minor differences in the proportion of articles covering this risk in each country, still a Chi-square test resulted in a statistically significant differences in the distribution of articles across countries for this risk ($\chi^2(5)=11.17$, p=0.04).

5.2 Analyzing the coverage of AI risks across bias categories of U.S. news outlets

Although news sources in our sample spanning across six countries is useful for insights about patterns of media coverage of AI risks, prior research has shown that the political bias of news domains tend to influence the discourse around scientific topics such as climate change (Chinn, Hart, and Soroka 2020; Allaham et al. 2025) or emerging technologies (e.g., nuclear energy), including AI (Brennen 2018; Roe and Perkins 2023; Vergeer 2020). This has also been reflected in U.S. politics, with President Trump rescinding the Executive Order 14110 on Artificial Intelligence signed by former President Biden.

Based on the limitations of our sample, specifically with respect to the number of articles and representation of media coverage in some countries (see section 7), and difficulties in taking into account and attributing the political-orientation for news domains beyond the authors' expertise in U.S.-based media, we chose to only focus on articles in our sample from the U.S. to explore the influence of political orientation of news outlets on risk reporting of AI. To this end, we used domain-level bias ratings from Media Bias Fact Check (MBFC) (Check 2025) to identify the political orientations of the U.S.-based outlets included in our dataset.

Although MBFC proposes one way of rating bias in news domains, we recognize that it is not the only approach. However, we selected MBFC because it is an independent website maintained by researchers and journalists that relies on human fact-checkers affiliated with the International Fact-Checking Network to evaluate media sources along different dimensions such as factual reporting and bias (Lin et al. 2023). After annotating the 69 U.S. domains in our sample with MBFC ratings, all 7,893 articles from 69 domains are distributed across four political bias categories: Left (12.25%) Left-Center (53.4%), Least Biased (5.8%), Right-Center (22.6%), and Right (5.8%). A list of news domains, their associated bias categories, and the corresponding number of articles from each domain in our sample is provided in Table A.5 in the Appendix.

To statistically test the association between the coverage of various AI risks and domain-bias categories in U.S.-based outlets, we conducted a Chi-square test of independence on the number of articles reporting each risk category for each bias category. In addition, we report the proportion of articles covering each risk category by counting the number of articles related to each risk category, and divided this number by the total number of articles in each bias category.

As shown in Figure 2, we find that right-biased sources under-emphasize *Socioeconomic & Environmental Risks* based on the proportion of articles reporting on this risk (25.7%) in comparison to the proportion of articles published by least-biased (35.1%), left-center (37.7%), left-biased (36.7%), and event right-center (40.2%) news outlets. Indeed the contrast between right-biased and right-center outlets had the greatest gap. This observed variability in news coverage across the different categories of domain bias is statistically significant, as indicated by the Chi-square test of independence ($\chi^2(4) = 34.89$, $p < 0.001$). This highlights the association between news coverage of AI risks and domain bias, potentially contributing to our understanding of the agenda-setting process of news media that shapes and influence the public perceptions of AI and its impacts on society.

As for *Malicious Actors & Misuse Risks*, right-biased and left-biased sources had the highest and second highest proportions of articles, accounting for 43.1% and 36.2% of their respective coverage, focusing on this risk category compared to outlets with other domain biases. Specifically, right-biased news outlets reported risks related to the misuse of AI by the "government" to "make AI woke", "push a leftist agenda", or "to clean the internet of conservative thought and replace it with leftist narratives", posing a threat to "free speech". Other reported risks focus on the potential misuse of AI's role in influencing the "public consumption and perception of news media", including "AI models favor[ing] leftist outlets like The Washington Post, NPR, and PBS" and "recommending news sources perceived as leftist-biased as the best". As for left-bias outlets, the reporting is focused on how the misuse of AI could "elevate extremist content" on the internet, "push far-right agenda on social media", and "enable far-right conspiracy theorists to push baseless claims about voter fraud", with a very few instances of articles reporting on the potential misuse of AI to advise "terror groups on biological weaponry".

Similarly, for the *Discrimination & Toxicity* risks, the proportion of articles published by right-biased (36.1%) and left-biased domains (24.9%) over-emphasize the coverage of this risk category, compared to center-bias and least-bias outlets. However, consistent with the *Malicious actors &*

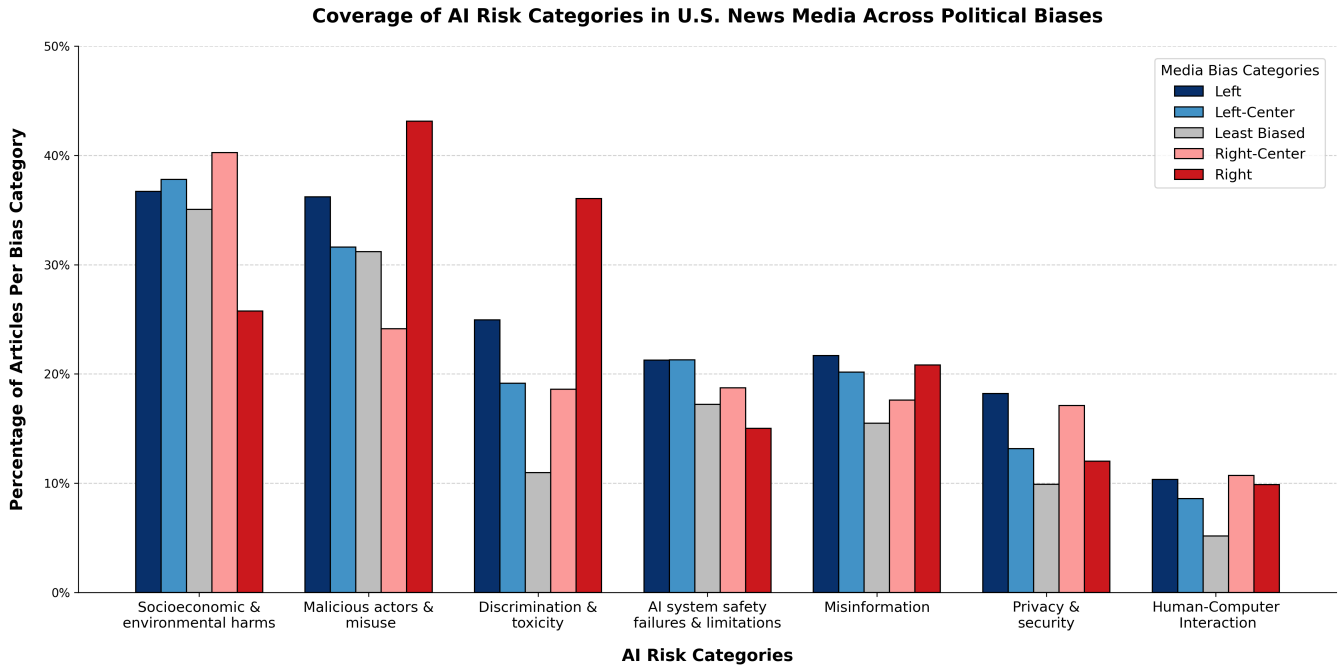

Figure 2: Proportion of news articles in our sample from U.S. domains across five media bias categories, as rated by Media Bias Fact Check (MBFC): Left, Left-Center, Least Biased, Right-Center, and Right.

Misuse Risks, the reporting of risks relevant to this category in right vs. left bias outlets include markers of politicized language. For instance, right-bias outlets mentioned risks pertaining to the false portrayal of "the founding fathers as people of color" by Gemini AI and how AI systems like ChatGPT "trained using Wikipedia, may produce responses skewed against conservatives and in favor of leftists". Another set of examples focused on the "bias in content moderation systems", including how the "negative comments about women are more likely to be flagged as hateful compared to the same comments about men". In contrast, left-bias media outlets reported on risks about the "inclusion of far-right and non-reputable sources" in AI training data which leads to the "dissemination of hate speech through AI responses". Furthermore, risks covered by left-bias outlets also focus on reporting the negative impacts of AI technologies experienced by people of color such as "financial discrimination" as a result of "mortgage approval algorithms discriminat[ing] against applicants of color", or the wrongful arrest of "eight month pregnant Porscha Woodruff" due an AI facial recognition tool flagging Porscha's face as a "carjacking suspect".

The Chi-square test results for both *Malicious Actors & Misuse Risks* and *Discrimination & Toxicity Risks* categories indicate that media coverage of these risk varies by the political bias of the reporting outlet, $\chi^2(4) = 84.87$, $p < 0.001$, and $\chi^2(4) = 116.52$, $p < 0.001$, respectively.

Emerging as one of the short-term and most severe global risks to human society (World Economic Forum Report 2024), *Misinformation Risks* received a consistent share of coverage by left (21.6%) and right (20.8%) biased news outlets, per the proportion of articles reporting this risk. However, the way these risks are reported on in news media seem to vary. Besides the need for AI systems to minimize hallucinations and improve factuality, as reported by several news articles, there are a number of articles reporting on the impacts of misinformation on democracy. For instance, some right-bias outlets voiced concerns about the risks of AI-generated content on electoral process and participatory policymaking, potentially "impacting democracy". This observation is also consistent with left-bias news outlets that view AI's capability to undermine democratic processes and accountability through AI-generated content that "spread unreliable information", "conspiracy theories", and "false information to [mislead] voters". Despite that relatively consistent coverage of misinformation risks across the four domain bias categories, Chi-square test shows that there is a statistically significant association ($\chi^2(4) = 13.55$, $p < 0.01$) between the coverage of misinformation risks and the domain bias of the outlets reporting on this risks category.

The remaining three risk categories receiving low coverage from news media, *AI Systems Safety, Failures, & Limitations*, *Privacy & Security*, and *Human-Computer Interaction* also show significant associations between the coverage of these risks and domain bias, $\chi^2(4) = 16.56$, $p = 0.002$, $\chi^2(4) = 37.40$, $p < 0.001$, $\chi^2(4) = 17.76$, $p = 0.001$, respectively. Unlike the previous categories, the coverage of these risks did not include politicized language, though we did observe some differences in the scope in which these risks are communicated in right vs. left biased outlets. For instance, for *AI Systems Safety, Failures, & Limitations* risks, articles published by left-biased outlets scoped the reporting on AI sys-

tems' lack of capability and robustness such as "ChatGPT lack[ing] a real moral compass and relies on crude guardrails that can be easily broken". In contrast, right-based outlets were centering their coverage for this risk category on the potentials of AI possessing dangerous capabilities that can cause mass harm or extinction "annihilat[ing] humankind in some sort of existential catastrophe" or "pose[ing] a risk of loss of control over human civilization". Similarly, left-biased outlets scoped their coverage of *Human-Computer Interaction* risks on the over-reliance on AI systems such as Doctors accepting the output of AI systems in medicine "without sufficient scrutiny", but less on how AI could transform human relations "potentially diminishing genuine human connections" and "deepen[ing] what some are calling an epidemic of loneliness", as observed in right-biased sources.

Collectively, our findings highlight how the coverage of AI risks varies by country and based on the political bias of news outlets reporting on these risks. We also illustrated through a few examples how the prevalence of AI risks, especially amongst left v.s. right-biased outlets, doesn't entail an equivalent scoping and communication of these risks by news media. While our analysis is focused on analyzing and describing the prevalence of AI risks in news reporting on AI, future work accounting for *how* AI risks and harms are framed, in the U.S. and abroad, in biased news outlets will help complement our findings and further inform how politically-biased media is shaping the public opinion and attitudes towards AI, and subsequently the public support (or opposition) of AI-related policies.

6 Discussion

This research illustrates through a comparative analysis of AI risks covered in a cross-national sample of news media spanning 6 countries (the U.S., the U.K., India, Australia, Israel, and South Africa) the influence of national (as indicated by country) and political (as indicated by political orientations) variations in the coverage of AI risks in news media. Considering the role of news media in mapping AI risks in real-world contexts, our findings can help inform risk assessors and policy-makers about the importance of accounting for national and political nuances in media coverage of AI risks and potentially calibrate ongoing AI incident monitoring initiatives to also include these nuances when incorporating news media as part of risk-based regulatory practices.

In particular, by considering the cross-national variations in the analysis of AI risks reported on in the news media, our research articulates how various AI risks identified by the academic community (synthesized by the MIT domain risk taxonomy) are reported on and prioritized (i.e., which risks are deemed important) by the media per the coverage of these risks in each country. Moreover, insights from the comparative analysis of AI risks across countries can complement existing risk assessment practices looking to incorporate more diverse perspectives into the identification and prioritization of these risks, particularly for AI systems developed in the US (or other Western countries) with a global user base. This inclusion is especially important as a way to counteract the potential for expert bias in assessment practices (Bonaccorsi, Apreda, and Fantoni 2020; Crawford 2016).

Furthermore, even a well-documented and performed risk assessment in a Western country may not reflect the risks realized in or prioritized by another country. Thus, the choice of risk taxonomy for mapping AI risks and the national context in which these risks are identified does matter. While we encourage risk assessors to consider findings from our research as part of assessing AI systems that are developed in the Global North (e.g., the U.S.), but have users in other countries in the Global South (e.g., South Africa and India), we emphasize the importance of leveraging news media as a complementary source for risk assessments, while accounting for its national and political variations, in providing an analytical lens to help quantify the prioritization of AI risks across regions and countries. For instance, we find that the U.K. deems malicious actors and misuse risks as the most important category (see Figure 1), while it is not as important in South Africa. There, socioeconomic and environmental harms are reflected in our sample as a more pressing risk, which may reflect the high importance of social justice values especially in regards to diversity and ethnic neutrality in African countries (Mengesha, Belay, and Adams 2024). These findings may re-orient risk assessors to contribute region-specific socio-technical methods and evaluations that AI-developers can leverage to align LLMs to address the risks more salient in each nation.

Focusing on the U.S., our study also reveals clear differences regarding the prevalence of AI risk in media coverage across outlets with different political leanings. We find that right-biased outlets deviate in their coverage of AI risks from centered and left-biased outlets. Specifically, right-biased media have a clear focus on the risks related to malicious actors & misuse as well as discrimination & toxicity. Our exploratory analysis shows that right-biased media outlets identify these risks from a lens that supports their agenda on topics relevant to freedom of expression and the presence of a culture war that are also endorsed by right-wing politicians. We also find that right-biased media call out a perceived cancel culture and woke-movement that is also carried out and enforced with and by AI technologies. The perceived risk is then not about the discrimination of marginalized groups, but the alleged suppression of majority voices.

The differences between the prevalence of AI risks between U.S. news outlets with different political leanings also show signs of polarization. This is an indicator for an emerging politicization of AI risks, which is further confirmed by the different standpoints of different media stances towards the risk categories. There is a possibility that the politicized language used to cover AI risks is further fueled by recent developments in U.S. politics, where Big Tech advocates are promised a more active and open role in political procedures (e.g. Elon Musk heading the DOGE initiative). This in turn could lead into public discussions on AI governance that are likely to intersect with politics, which could have implications for bi-partisan efforts aiming to protect the public from the negative impacts of AI. Accordingly, we encourage future research to extrapolate on our findings and trace in more detail *how* AI risks are framed in news media, the impact of

such framing on public attitudes towards AI, and how the framing of AI risks could evolve with the political climate and democratic processes (e.g., elections) in the U.S.

Taking a step back, our research also finds that while the research community often focuses on potential risks of AI systems from technical (e.g., AI system safety failures and limitations) and socio-technical perspectives (i.e., implications of AI systems on society), per the prevalence of these risks in research papers included as part of the MIT Risk Repository (Slattery et al. 2024), news media reporting have a stronger focus on harms that tend to prioritize tangible societal implications of AI risks (such as socioeconomic & environmental harms; malicious actors & misuse; misinformation). This difference in the prioritization of reporting on some AI risks is evident by the prominence of Malicious actors & Misuse risks being ranked as the third most prevalent risk category in the corpus of academic papers included in constructing the MIT Risk Repository (i.e., 71% of academic papers mention this risk) compared to it being ranked second in our corpus of articles from news media. In addition, Misinformation risks plays a larger role in media reporting than in the MIT risk repository (ranked 7th based on its prevalence of 46% in academic papers). This finding supports our rationale that including news media coverage can support expert risk assessment practices in adding more nuances and societal importance indicators to the assessment practice. Collectively, these findings may warrant future research to explore how the difference in incentives and priorities between academic research and news media tend to influence what risks are discussed in the public sphere, which impacts are deemed important (and which aren't), and which impacts will ultimately be prioritized.

7 Limitations

Although our study includes news articles from six countries across different geographical regions, it also has its limitations. First, our sample doesn't capture perspectives or discourse around AI risks by news domains with paid-subscriptions (see Data section). Rather, it relies primarily on news domains that are "freely available" (i.e., not behind paywalls), publicly accessible, and can be scraped, especially in countries beyond the U.S. Accordingly, our findings, based on analyzing news coverage from some of the most widely read and publicly accessed news outlets, are likely to capture and reflect a partial, yet broad, view of public concerns of AI risks as reported in these leading news sources in each of the six countries (Newman et al. 2024).

Second, our sample focuses primarily on articles written and published in the English language, excluding outlets in other regions (e.g., Latin America) and potentially missing other important and relevant news coverage, as well as local perspectives, on AI risks that are expressed in other languages beyond English that are native to some underrepresented countries in our sample, such as South Africa or India.

Third, analyzing the public discourse around AI from news outlets, especially in countries in the Global South, requires deep expertise in these countries to account for social, political, economic, and even legacy colonial ties which collectively have an influence on the discourse around how the impacts of AI are communicated and realized in communities in these countries (Dewitt Prat et al. 2024; Baguma et al. 2023). We, as a research team, do not claim to have this expertise, and, thus, took a descriptive approach to risk prevalence. However, country specific interpretations require cooperation with local scholars which is beyond the scope of this paper.

Lastly, despite the cross-national nature of our sample, we recognize that the six countries are not representative of the regions or political perspectives these countries belong to. For instance, although the U.K. is geographically located in Europe, it is no longer a part of the European Union which adopts a different perspective on AI risks than the U.K., as reflected in the EU AI Act (Montasari 2023; Cath et al. 2018; Akinola, Tunbosun, and Oladapo 2022).

8 Conclusion

This work highlights the importance of incorporating *national* and *political* variations embedded in the reporting of AI risks when considering news media as a complementary source in risk assessment and incidents monitoring practices. Through a comparative analysis of a cross-national sample of news media spanning 6 countries (the U.S., the U.K., India, Australia, Israel, and South Africa), we find that AI risks are prioritized differently across nations, as reflected in the prevalence of these risks from each country's media coverage. We further elaborate on our findings by considering the political orientation of news outlets in our analysis, particularly in the U.S. We find the reporting of AI risks by these outlets to contain politicized language across *Malicious Actors & Misuse Risks*, *Discrimination & Toxicity*, and *Misinformation* risks. Our research presents risk assessors, AI developers, and policymakers, with an analytical lens to help quantify the prioritization of AI risks based on the national and political variations of news media. Moreover, our findings may re-orient risk assessors' perspectives towards contributing region or country-specific evaluations that account for the socio-political contexts in various regions, especially in the Global South. In doing so, AI-developers can leverage these evaluations in their assessment frameworks to capture the societal (mis)alignment of AI systems, such as LLMs, with the values and risks most salient for each nation using these systems, without disregarding other globally emerging risks. Failing to account for the social and political contexts surrounding the identification, interpretation, and communication of AI risks may lead to assessors overlooking how these nuances are influencing the public perception of AI and its risks, and potentially hindering progress towards shaping more inclusive AI governance policies and regulations.

References

Akinola, O.; Tunbosun, O. A.; and Oladapo, B. 2022. Comparative Analysis Regulatory of AI and Algorithm in UK, EU and USA. *EU and USA (September 7, 2022)*.

Allaham, M.; Kieslich, K.; and Diakopoulos, N. 2024. Towards Leveraging News Media to Support Impact Assessment of AI Technologies. In *EvalEval Workshop at NeurIPS 2024*. https://arxiv.org/pdf/2411.02536.

Allaham, M.; Lokmanoglu, A. D.; Hart, P.; and Nisbet, E. C. 2025. Enhancing LLMs for Governance with Human Oversight: Evaluating and Aligning LLMs on Expert Classification of Climate Misinformation for Detecting False or Misleading Claims about Climate Change. In *International Workshop on AI Governance: Alignment, Morality and Law (AIGOV) 2025. AAAI Conference on Artificial Intelligence.* https://arxiv.org/abs/2501.13802.

Angwin, J.; Larson, J.; Mattu, S.; and Kirchner, L. 2022. Machine bias. In *Ethics of data and analytics*, 254–264. Auerbach Publications.

Avin, S.; Gruetzemacher, R.; and Fox, J. 2020. Exploring AI Futures Through Role Play. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 8–14. New York NY USA: ACM. ISBN 978-1-4503-7110-0.

Baguma, R.; Namuwaya, H.; Nakatumba-Nabende, J.; and Rashid, Q. M. 2023. Examining potential harms of large language models (llms) in africa. In *International Conference on Safe, Secure, Ethical, Responsible Technologies and Emerging Applications*, 3–19. Springer.

Becker, H. A. 2001. Social impact assessment. *European Journal of Operational Research*, 128(2): 311–321.

Biden, J. R. 2023. Executive order on the safe, secure, and trustworthy development and use of artificial intelligence.

Bommasani, R.; Hudson, D. A.; Adeli, E.; Altman, R.; Arora, S.; von Arx, S.; Bernstein, M. S.; Bohg, J.; Bosselut, A.; Brunskill, E.; et al. 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.

Bonaccorsi, A.; Apreda, R.; and Fantoni, G. 2020. Expert biases in technology foresight. Why they are a problem and how to mitigate them. *Technological Forecasting and Social Change*, 151: 119855.

Braun, V.; and Clarke, V. 2012. *Thematic analysis.* American Psychological Association.

Brennen, J. 2018. An industry-led debate: How UK media cover artificial intelligence. Publisher: Reuters Institute for the Study of Journalism.

Brennen, J.; Howard, P. N.; and Nielsen, R. K. 2018. An industry-led debate: How UK media cover artificial intelligence. RISJ Fact-Sheet.

Brennen, J. S.; Howard, P. N.; and Nielsen, R. K. 2022. What to expect when you're expecting robots: Futures, expectations, and pseudo-artificial general intelligence in UK news. *Journalism*, 23(1): 22–38.

Bunz, M.; and Braghieri, M. 2022. The AI doctor will see you now: assessing the framing of AI in news coverage. *AI & SOCIETY*, 37(1): 9–22.

Burtell, M.; and Woodside, T. 2023. Artificial influence: An analysis of AI-driven persuasion. *arXiv preprint arXiv:2303.08721*.

Cath, C.; Wachter, S.; Mittelstadt, B.; Taddeo, M.; and Floridi, L. 2018. Artificial intelligence and the 'good society': the US, EU, and UK approach. *Science and engineering ethics*, 24: 505–528.

Check, M. B. 2025. Media Bias/Fact Check: The Most Comprehensive Media Bias Resource. Accessed on January 17, 2025.

Chinn, S.; Hart, P. S.; and Soroka, S. 2020. Politicization and polarization in climate change news content, 1985-2017. *Science Communication*, 42(1): 112–129.

Chuan, C.-H.; Tsai, W.-H. S.; and Cho, S. Y. 2019. Framing artificial intelligence in American newspapers. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 339–344.

Constantaras, E.; Geiger, G.; Braun, J.-C.; Mehrotra, D.; and Aung, H. 2023. Inside the suspicion machine. Technical report.

Crawford, K. 2016. Artificial intelligence's white guy problem. *The New York Times*, 25(06): 5.

De Wilde, P. 2011. No polity for old politics? A framework for analyzing the politicization of European integration. *Journal of European integration*, 33(5): 559–575.

Dewitt Prat, L.; NDLOVU LUCAS, O. N.; GOLIAS, C.; and LEWIS, M. 2024. Decolonizing LLMs: An Ethnographic Framework for AI in African Contexts. In *Ethnographic Praxis in Industry Conference Proceedings*, volume 2024, 46–85. Wiley Online Library.

Diakopoulos, N. 2015a. Algorithmic accountability: Journalistic investigation of computational power structures. *Digital journalism*, 3(3): 398–415.

Diakopoulos, N. 2015b. Algorithmic Accountability: Journalistic investigation of computational power structures. *Digital Journalism*, 3(3): 398 – 415.

Diakopoulos, N. 2025. Prospective Algorithmic Accountability and the Role of the News Media. In Moorman, M.; and Verdicchio, M., eds., *Computer Ethics Across Disciplines: Deborah G. Johnson and Algorithmic Accountability*.

Fast, E.; and Horvitz, E. 2017. Long-term trends in the public perception of artificial intelligence. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31. Issue: 1.

Filippucci, F.; Gal, P.; Jona-Lasinio, C.; Leandro, A.; and Nicoletti, G. 2024. The impact of Artificial Intelligence on productivity, distribution and growth: Key mechanisms, initial evidence and policy challenges.

Ganguli, D.; Lovitt, L.; Kernion, J.; Askell, A.; Bai, Y.; Kadavath, S.; Mann, B.; Perez, E.; Schiefer, N.; Ndousse, K.; et al. 2022. Red teaming language models to reduce harms: Methods, scaling behaviors, and lessons learned. arXiv. *URL: http://arxiv. org/abs/2209.07858 [accessed 2023-09-20]*.

Gilardi, F.; Kasirzadeh, A.; Bernstein, A.; Staab, S.; and Gohdes, A. 2024. We need to understand the effect of narratives about generative AI. *Nature Human Behaviour*, 1–2.

Hagerty, A.; and Rubinov, I. 2019. Global AI ethics: a review of the social impacts and ethical implications of artificial intelligence. *arXiv preprint arXiv:1907.07892*.

Hartmann, D.; De Pereira, J. R. L.; Streitbörger, C.; and Berendt, B. 2024. Addressing the regulatory gap: moving towards an EU AI audit ecosystem beyond the AI Act by including civil society. *AI and Ethics*.

Hautala, J.; and Heino, H. 2023. Spectrum of AI futures imaginaries by AI practitioners in Finland and Singapore: The unimagined speed of AI progress. *Futures*, 153: 103247.

Hernandes, R.; and Corsi, G. 2024. Auditing Google's Search Algorithm: Measuring News Diversity Across Brazil, the UK, and the US. *arXiv preprint arXiv:2410.23842*.

Hilgard, J.; and Li, N. 2017. *A Recap: The Science of Communicating Science*, volume 1. Oxford University Press.

Ittefaq, M.; Zain, A.; Arif, R.; Ala-Uddin, M.; Ahmad, T.; and Iqbal, A. 2025. Global news media coverage of artificial intelligence (AI): A comparative analysis of frames, sentiments, and trends across 12 countries. *Telematics and Informatics*, 96: 102223.

Jobin, A.; Ienca, M.; and Vayena, E. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9): 389–399.

Kieslich, K. 2024. *The role of public opinion on ethical AI principles and its implication for a common good-oriented implementation.* Ph.D. thesis, Universität Hohenheim.

Kieslich, K.; Diakopoulos, N.; and Helberger, N. 2024. Anticipating impacts: using large-scale scenario-writing to explore diverse implications of generative AI in the news environment. *AI and Ethics*.

Kieslich, K.; Došenović, P.; and Marcinkowski, F. 2022. Everything, but hardly any science fiction. Technical Report 7, Meinungsmonitor Künstliche Intelligenz.

Kieslich, K.; Helberger, N.; and Diakopoulos, N. 2024. My Future with My Chatbot: A Scenario-Driven, User-Centric Approach to Anticipating AI Impacts. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 2071–2085. Rio de Janeiro Brazil: ACM. ISBN 9798400704505.

Kieslich, K.; Helberger, N.; and Diakopoulos, N. 2025. Scenario-Based Sociotechnical Envisioning (SSE): An Approach to Enhance Systemic Risk Assessments.

Kieslich, K.; Lünich, M.; and Došenović, P. 2023. Ever Heard of Ethical AI? Investigating the Salience of Ethical AI Issues among the German Population. *International Journal of Human–Computer Interaction*, 1–14.

Kolt, N. 2025. Governing ai agents. *arXiv preprint arXiv:2501.07913*.

Leetaru, K.; and Schrodt, P. A. 2013. Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA annual convention*, volume 2, 1–49. Citeseer.

Lin, H.; Lasser, J.; Lewandowsky, S.; Cole, R.; Gully, A.; Rand, D. G.; and Pennycook, G. 2023. High level of correspondence across different news domain quality rating sets. *PNAS nexus*, 2(9): pgad286.

Madiega, T. 2021. Artificial intelligence act. *European Parliament: European Parliamentary Research Service*.

Mazeika, M.; Phan, L.; Yin, X.; Zou, A.; Wang, Z.; Mu, N.; Sakhaee, E.; Li, N.; Basart, S.; Li, B.; et al. 2024. Harmbench: A standardized evaluation framework for automated red teaming and robust refusal, 2024. *URL https://arxiv.org/abs/2402.04249*.

McCombs, M. E.; and Shaw, D. L. 1972. The agenda-setting function of mass media. *Public opinion quarterly*, 36(2): 176–187.

McGregor, S. 2021. Preventing repeated real world AI failures by cataloging incidents: The AI incident database. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 15458–15463.

Media Cloud. 2025. Collection 9272347. Accessed on January 17, 2025.

Meißner, F. 2024. Risks and opportunities of 'generative A.I.': How do news media cover ChatGPT? In *International Crisis and Risk Communication Conference Proceedings*. International Crisis and Risk Communication Association.

Mengesha, G. H.; Belay, E. G.; and Adams, R. 2024. Social justice considerations in developing and deploying AI in Africa. *Data & Policy*, 6: e65.

Metcalf, J.; Moss, E.; Watkins, E. A.; Singh, R.; and Elish, M. C. 2021. Algorithmic impact assessments and accountability: the co-construction of impacts. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*.

Mitchell, M.; Ghosh, A.; Luccioni, A. S.; and Pistilli, G. 2025. Fully Autonomous AI Agents Should Not be Developed. *arXiv preprint arXiv:2502.02649*.

Montasari, R. 2023. National artificial intelligence strategies: a comparison of the UK, EU and US approaches with those adopted by state adversaries. In *Countering Cyberterrorism: The Confluence of Artificial Intelligence, Cyber Forensics and Digital Policing in US and UK National Cybersecurity*, 139–164. Springer.

Nanayakkara, P.; Hullman, J.; and Diakopoulos, N. 2021. Unpacking the expressed consequences of AI research in broader impact statements. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 795–806.

Nechushtai, E.; Zamith, R.; and Lewis, S. C. 2024. More of the same? Homogenization in news recommendations when users search on Google, YouTube, Facebook, and Twitter. *Mass Communication and Society*, 27(6): 1309–1335.

Newman, N.; Fletcher, R.; Robertson, C. T.; Arguedas, A. R.; and Nielsen, R. K. 2024. Reuters Institute Digital News Report 2024. Report, Reuters Institute for the Study of Journalism, Oxford.

Nguyen, D.; and Hekman, E. 2022. A 'New Arms Race'? Framing China and the U.S.A. in A.I. News Reporting: A Comparative Analysis of the Washington Post and South China Morning Post. *Global Media and China*, 7(1): 58–77.

Nguyen, D.; and Hekman, E. 2024. The news framing of artificial intelligence: a critical exploration of how media discourses make sense of automation. *AI & SOCIETY*, 39(2): 437–451.

Nisbet, M. C.; Scheufele, D. A.; Shanahan, J.; Moy, P.; Brossard, D.; and Lewenstein, B. V. 2002. Knowledge, Reservations, or Promise?: A Media Effects Model for Public Perceptions of Science and Technology. *Communication Research*, 29(5): 584–608.

OpenAI. 2024. GPT-4o System Card. https://arxiv.org/abs/2410.21276. ArXiv:2410.21276 [cs.CL].

Ou-Yang, L. 2025. newspaper: Simplified Python Article Scraping & Curation. Accessed on January 17, 2025.

Ouchchy, L.; Coin, A.; and Dubljević, V. 2020. AI in the headlines: the portrayal of the ethical issues of artificial intelligence in the media. *AI & SOCIETY*, 35: 927–936.

Ouchchy, L.; Coin, A.; and Dubljević, V. 2020. AI in the headlines: the portrayal of the ethical issues of artificial intelligence in the media. *AI & SOCIETY*, 35(4): 927–936.

Pang, R. Y.; Santy, S.; Just, R.; and Reinecke, K. 2024. BLIP: Facilitating the Exploration of Undesirable Consequences of Digital Technologies. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–18.

Park, P. S.; Goldstein, S.; O'Gara, A.; Chen, M.; and Hendrycks, D. 2024. AI deception: A survey of examples, risks, and potential solutions. *Patterns*, 5(5).

Reuel, A.; Bucknall, B.; Casper, S.; Fist, T.; Soder, L.; Aarne, O.; Hammond, L.; Ibrahim, L.; Chan, A.; Wills, P.; et al. 2024. Open problems in technical ai governance. *arXiv preprint arXiv:2407.14981*.

Roberts, H.; Bhargava, R.; Valiukas, L.; Jen, D.; Malik, M. M.; Bishop, C. S.; Ndulue, E. B.; Dave, A.; Clark, J.; Etling, B.; et al. 2021. Media cloud: Massive open source collection of global news on the open web. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, 1034–1045.

Roe, J.; and Perkins, M. 2023. 'What they're not telling you about ChatGPT': exploring the discourse of AI in UK news media headlines. *Humanities and Social Sciences Communications*, 10(1): 753.

Schattschneider, E. E. 1957. Intensity, visibility, direction and scope. *American Political Science Review*, 51(4): 933–942.

Scheufele, D. A. 1999. Framing as a theory of media effects. *Journal of communication*, 49(1): 103–122.

Scheufele, D. A.; and Lewenstein, B. V. 2005. The Public and Nanotechnology: How Citizens Make Sense of Emerging Technologies. *Journal of Nanoparticle Research*, 7(6): 659–667.

Schäfer, M. S. 2017. *How Changing Media Structures Are Affecting Science News Coverage*, volume 1. Oxford University Press.

Shelby, R.; Rismani, S.; Henne, K.; Moon, A.; Rostamzadeh, N.; Nicholas, P.; Yilla-Akbari, N.; Gallegos, J.; Smart, A.; Garcia, E.; et al. 2023. Sociotechnical harms of algorithmic systems: Scoping a taxonomy for harm reduction. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 723–741.

Slattery, P.; Saeri, A. K.; Grundy, E. A.; Graham, J.; Noetel, M.; Uuk, R.; Dao, J.; Pour, S.; Casper, S.; and Thompson, N. 2024. The ai risk repository: A comprehensive meta-review, database, and taxonomy of risks from artificial intelligence. *arXiv preprint arXiv:2408.12622*.

Solaiman, I.; Talat, Z.; Agnew, W.; Ahmad, L.; Baker, D.; Blodgett, S. L.; Daumé III, H.; Dodge, J.; Evans, E.; Hooker, S.; Jernite, Y.; Luccioni, A. S.; Lusoli, A.; Mitchell, M.; Newman, J.; Png, M.-T.; Strait, A.; and Vassilev, A. 2023. Evaluating the Social Impact of Generative AI Systems in Systems and Society. ArXiv:2306.05949 [cs].

Sun, S.; Zhai, Y.; Shen, B.; and Chen, Y. 2020a. Newspaper coverage of artificial intelligence: A perspective of emerging technologies. *Telematics and Informatics*, 53: 101433.

Sun, S.; Zhai, Y.; Shen, B.; and Chen, Y. 2020b. Newspaper coverage of artificial intelligence: A perspective of emerging technologies. *Telematics and Informatics*, 53: 101433.

Ulken, A. 2005. Question of Balance: Are Google News search results politically biased.

Uuk, R.; Gutierrez, C. I.; Guppy, D.; Lauwaert, L.; Velasco, L.; Slattery, P.; and Prunkl, C. 2024. A Taxonomy of Systemic Risks from General-Purpose AI.

van der Heijden, J. 2021. Risk as an approach to regulatory governance: An evidence synthesis and research agenda. *Sage Open*, 11(3): 21582440211032202. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

Vergeer, M. 2020. Artificial Intelligence in the Dutch Press: An Analysis of Topics and Trends. *Communication Studies*, 71(3): 373–392.

Ward, M. D.; Beger, A.; Cutler, J.; Dickenson, M.; Dorff, C.; and Radford, B. 2013. Comparing GDELT and ICEWS event data. *Analysis*, 21(1): 267–297.

Weidinger, L.; Rauh, M.; Marchal, N.; Manzini, A.; Hendricks, L. A.; Mateos-Garcia, J.; Bergman, S.; Kay, J.; Griffin, C.; Bariach, B.; et al. 2023. Sociotechnical safety evaluation of generative ai systems. *arXiv preprint arXiv:2310.11986*.

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.; Haas, J.; Legassick, S.; Irving, G.; and Gabriel, I. ???? Taxonomy of Risks posed by Language Models. Publication Title: 2022 ACM Conference on Fairness, Accountability, and Transparency Volume: 22.

World Economic Forum Report. 2024. Global Risks Report 2024. Accessed: 2025-04-15.

Zeng, Y.; Klyman, K.; Zhou, A.; Yang, Y.; Pan, M.; Jia, R.; Song, D.; Liang, P.; and Li, B. 2024. AI Risk Categorization Decoded (AIR 2024): From Government Regulations to Corporate Policies. Version Number: 1.

Zhang, Z.; Lei, L.; Wu, L.; Sun, R.; Huang, Y.; Long, C.; Liu, X.; Lei, X.; Tang, J.; and Huang, M. 2023. Safetybench:

Evaluating the safety of large language models with multiple choice questions. *arXiv preprint arXiv:2309.07045*.