Local Differences, Global Lessons: Insights from Organisation Policies for Legislation

Lucie-Aimée Kaffee*
Hugging Face
lucie.kaffee@huggingface.co

Pepa Atanasova*
University of Copenhagen
pepa@di.ku.dk

Anna Rogers
IT University of Copenhagen
arog@itu.dk

Abstract

The rapid adoption of AI across diverse domains has led to the development of organisational guidelines that vary significantly, even within the same sector. This paper examines AI policies in two domains, news organisations and universities, to understand how bottom-up governance approaches shape AI usage and oversight. By analysing these policies, we identify key areas of convergence and divergence in how organisations address risks such as bias, privacy, misinformation, and accountability. We then explore the implications of these findings for AI legislation, particularly the EU AI Act, highlighting gaps where practical policy insights could inform regulatory refinements. Our analysis reveals that organisational policies often address issues such as AI literacy, disclosure practices, and environmental impact, areas that are underdeveloped in existing legislative frameworks. We argue that lessons from domain-specific AI policies can contribute to more adaptive and effective AI governance at the global level. This study provides actionable recommendations for policymakers seeking to bridge the gap between local AI practices and regulations.

1 Introduction

The recent advancements in the performance of AI models[1] on a multitude of tasks, especially in zero-shot or few-shot scenarios Kocoń et al. [2023], Zhao et al. [2023], have accelerated their adoption across different domains. This rapid uptake has presented organisations with unprecedented challenges, necessitating the *swift development of organisational guidelines* for the use of AI to manage associated risks and opportunities. These guidelines reflect bottom-up governance approaches tailored to local needs and operational contexts.

From the top-down perspective, the most significant governance effort currently is the EU AI Act (`EU-AIA`) European Parliament [2024] (§2), which provides overarching frameworks for managing high-risk AI systems. However, while such frameworks establish broad standards, their top-down approach necessarily lacks the specificity required for effective implementation in diverse organisational settings. This creates gaps where organisations must independently interpret and address

*These authors contributed equally to this work.

[1]The policies discussed in this article mostly focus on the current systems commonly referred to as "generative AI", in that they can be used to generate synthetic texts/images etc. They are typically based on models of Transformer architecture pre-trained on large volumes of data (texts, images etc.). The exact definitions are rarely provided and are an active area of discussion in research (see references in Table 6).

Workshop on Regulatable ML at the 39th Conference on Neural Information Processing Systems (NeurIPS 2025).

risks, resulting in guidelines that emphasise practical challenges, such as AI literacy, bias mitigation, and environmental sustainability, which are underdeveloped in existing legislation. While it is expected and natural that top-down legislation is necessarily incomplete and leaves room for local interpretation and practices, legal endogeneity theory Edelman et al. [1999], Edelman [2007] also stipulates that the evolving 'local' practices may in time impact the 'global' legislation. This raises our main research question: *What AI-related issues and solutions have been identified in the current organisational policies, and what role may they play in the iterative development of industry practices and legislation?* To address this question, we contribute case studies of European-based news organisations and universities, considering their AI policies in the light of EU-AIA.

This paper is structured as follows. We conduct case studies of organisational AI guidelines from two domains: universities and news organisations. Through iterative coding of guidelines developed by organisations in each domain (§3), we examine discrepancies in how risks are classified, uses are prescribed, and the performance of AI models is perceived. We discuss the commonalities and discrepancies within each domain (news §4, universities §5).We then identify the policy items, for which a consistent implementation would conflict with the current market incentives of the AI industry, and which therefore may require a legislative intervention rather than industry standards. Based on this research, we provide actionable policy recommendations on digital inequity, disclosure of AI-generated content, bias, attribution, and environmental impact in the context of AI legislation (§6).

2 Related Work

In recent years, both news organisations and educational institutions have proactively developed policies to guide the ethical and effective use of artificial intelligence (AI) within their respective domains. While a survey of socio-technical challenges of generativ models is out of scope for this work, we identified the major potential source of risks for organizations, organized according to the taxonomy by Weidinger et al. [2022] in App., Table 6.

EU AI Act. For our purposes, the key factor of the EU AI Act is that it implements a risk-based approach to regulating AI, in which systems are categorised by their potential threat. Given the dynamic nature of AI development, the legislation will need updating over time Cantero Gamito and Marsden [2024]. Cantero Gamito and Marsden [2024] highlight the critical role of standards in the `EU-AIA`'s co-regulation strategy, emphasising the need for reform to keep pace with the rapid evolution of AI technologies. There is also an important distinction between "providers" and "deployers", each bearing distinct responsibilities. In the case of universities, they can be both deployers of AI systems, if they offer their own AI systems, as well as end-users, subscribing to other AI deployers services, such as 'ChatGPT'. For news organisations, in a majority of cases, the policies assume that they subscribe to an external AI deployer's systems. Downstream users and large-scale distributors of AI generated content, as could be the case for news organisations, do not currently have obligations under the `EU-AIA`. In this article, we focus on the `EU-AIA` and exclude discussions about related EU regulation, such as the *Directive on Copyright in the Digital Single Market*.

Scholars have raised multiple concerns about the Act's approach, e.g., arguing that the EU conflates trustworthiness with the acceptability of risk Laux et al. [2024] and identifying limitations and loopholes in the `EU-AIA` and proposed liability directives Wachter [2024]. These critiques underscore the importance of analysing gaps in the `EU-AIA` and pointing out possible improvements for future iterations, which is the purpose of this work. While previous studies have described organisational policies or analysed the gaps in the `EU-AIA`, to the best of our knowledge, this is the first attempt to inform analysis of the gaps in the `EU-AIA` by the existing and practically tested organisational policies as well as issues identified in academic research.

News Organisations. With the wide uptake of AI, news organisations have been impacted by the use and availability of AI models and systems, requiring them to create clear guidelines and policies on how to use this new technology Simon and Isaza-Ibarra [2023]. A comprehensive analysis of 52 news organisations across various countries reveals a concerted effort to address AI's implications on journalistic integrity, transparency, and accountability Simon and Becker [2023]. Previous work has compared the transparency provisions in the European Union's AI Act, particularly Article 50, and their alignment with news readers' expectations Piasecki et al. [2024]. Similar to our findings, the study highlights the necessity for clear disclosure when AI systems contribute to news content, as transparency is crucial for maintaining reader trust. de Lima-Santos and Ceron [2022] explore the current perceptions and future outlook of AI in news media, identifying key areas where AI technologies are being adopted, such as machine learning, computer vision, and natural language processing. While they find potential benefits of AI in enhancing news production and distribution, they also caution against challenges related to editorial standards and public trust. In the context of visual AI in news organisations, Thomson et al. [2024] emphasise the importance of clear guidelines to ensure clear boundaries between human-created and AI-generated images.

Education. The uptake of AI technologies also impacts educational Institutions, such as universities. Studies point out a set of challenges with the use of ChatGPT in the educational institutions such as the misuse of ChatGPT to trick online exams Susnjak and McIntosh [2024]; the higher correct answer rate on exams with the improvement of AI technologies de Winter [2023]; the integration of generative AI technologies into engineering education given the limitations of the training data quality Qadir [2023]; superficial or inaccurate responses, potentially misleading students and the risk of bias and discrimination Farrokhnia et al. [2024]. Hence, educational institutions are also actively formulating AI policies to navigate the integration of AI in academic settings. The 2024 EDUCAUSE Action Plan[2] and UNESCO Holmes et al. [2023] outline comprehensive guidelines for AI adoption in higher education, both focusing on ethical considerations and the impact on teaching and learning practices. Ghimire and Edwards [2024] conducted a survey of academic institutions, such as high schools, to collect information on current policies w.r.t. generative AI. This study gives an insight into the current lack of AI policies in many education institutions in the US. Slimi and Carballido [2023] emphasise the importance of stakeholders working together to develop AI policies in the education space. A set of studies have focused on the perspective of educators Pischetola et al. [2024], proposing policy implications based on a survey of teachers Chiu [2023]. Dotan et al. [2024] propose a "points to consider" approach for the responsible adoption of generative AI in higher education, emphasising alignment of AI integration with the unique goals, values, and structural features of higher education institutions.

3 Methodology

Selection of Organisational Policies for Analysis. We selected universities and news organisations as illustrative cases because they represent two sectors where AI adoption is already having a significant impact on core societal functions: the production and dissemination of knowledge, and the shaping of public discourse. Both have been notably proactive in developing organisational AI policies, enabling a comparative analysis that highlights domain-specific needs and offers insights for refining broader legislative frameworks. At the same time, they present an instructive contrast: universities apply AI in direct functions, some of which are classified as high-risk (e.g., automated grading), and operate under sector-wide regulatory oversight from ministries of education and accreditation bodies, making them useful for examining gaps in top-down regulation. By contrast, newsrooms use AI in content generation and editorial workflows, raising concerns about misinformation, bias, and synthetic media, and they largely rely on self-regulation and industry standards, exemplifying bottom-up governance approaches. To capture the diversity of organisational responses across different legal, cultural, and institutional contexts, we conduct a cross-country comparison, recognising that differences between technical and general research universities, or among news outlets with varying political perspectives, further enrich the analysis by showing how diverse institutional identities shape AI governance choices.

Selection of News Policies. We select 10 news outlets with publicly available policies on generative AI use from across the EU, Switzerland, and the UK, based on Simon and Becker [2023] and Cools [2023], which list news organisations' policies. The full list of news outlets, selected under the condition that they are based in Europe, is in App. Table 4. We translate each of the policies into English using Google Translate if they are not available in English. We access the latest version of the guidelines, which is documented along with the links to the policies in App., Table 7.

Selection of University Policies. We select 10 universities from across the EU, choosing one top-ranked university from each country[3] which has publicly available guidelines on generative AI use. [4]

[2] https://www.educause.edu/research/2024/2024-educause-action-plan-ai-policies-and-guidelines

[3] https://www.topuniversities.com/europe-university-rankings

[4] We selected only one outlet per country to prioritise for geographic coverage while avoiding overrepresenting a single country. This allows us to highlight cross-country differences and similarities while keeping the analysis balanced, manageable, and focused on identifying patterns that are broadly relevant across contexts.

Code	FT	ANP	Guardian	Parisien	Spiegel	SZ	BBC	Mediahuis	Ringier	VG
Suggested Uses										
Illustrations/Graphics	✅	✅	-	✅	✅	✅	-	-	-	✅
Image generation	📜	✅	-	✅	✅	-	-	-	-	📜
Article generation	📜	✅	-	📜	✅	📜	-	-	-	-
Data analysis	✅	-	✅	-	-	✅	-	-	-	-
Language tool	-	✅	✅	✅	-	-	-	-	-	-
Transcription/Translation	✅	✅	-	-	-	✅	-	-	-	-
Ideas (Content)	-	✅	-	-	-	-	-	-	-	-
Ideas (Marketing)	-	-	✅	-	-	-	-	-	-	-
Content Moderation	-	-	-	-	-	✅	-	-	-	-
Issued Warning and Rules										
Human oversight	📜	📜	📜	📜	📜	📜	📜	📜	📜	📜
Declaration of use	📜	-	📜	📜	📜	📜	📜	📜	📜	📜
Factual accuracy	⚠️	⚠️	⚠️	-	⚠️	-	📜	-	⚠️	-
Bias in AI	⚠️	⚠️	📜	-	-	-	-	📜	⚠️	-
Privacy and sensitive data	-	-	-	-	📜	-	📜	📜	📜	📜
Copyright	-	-	📜	⚠️	-	-	📜	📜	⚠️	-
AI literacy training	📜	-	-	-	-	-	📜	📜	-	-
Document AI use	📜	-	-	-	📜	-	📜	-	-	-

Table 1: AI-related practices in news organisations. ✅= suggested use, ⚠️= warning, 📜= rule.

The selected institutions are listed in App. Table 5. To reach our target of ten universities, we checked 27 in total, discarding 17 that did not have public guidelines (six universities from France, three from Spain, two from Sweden, two from Denmark, two from Italy, one from Finland, and one from Switzerland). The guidelines for Vie, UiO, KUL were automatically translated. We access the latest version of the guidelines, which is documented along with the links to the policies in App., Table 8.

Coding of Organisational Policies. Given the set of policies in the news organisations and universities described above, we perform iterative coding of these policies. From the identified categories in these policies, we focus on those mentioned by at least two organisations within the same domain. Additionally, we discuss points raised by individual organisations in sections §4 and §5. Where possible, we merge the codes identified across the two domains. App. Table 3 outlines the resulting codes and their definitions in the respective domain. The codes are grouped into two broad categories: suggested uses of AI within the domain and warnings or rules/requirements related to its use.

4 AI Policies in News Organisations

We describe the policies of 10 news organisations, as summarised in Table 1 towards their suggested uses of AI and issued warnings and rules w.r.t. AI use in the newsroom.

4.1 Suggested Uses

All news outlets surveyed encourage the use of AI in their work; however, they propose different degrees of use and applications. Where in Table 1 outlets do not mention any of the listed codes, they still encourage AI usage but do not explicitly list suggested uses in their policy. Further, all news outlets require human oversight for all or most AI use as well as labelling the output as AI generated.

The use of AI in news organisations has two directions. First, tooling to help automated processes in the day to day work of journalists such as **data analysis**, **language tools**, e.g., grammar correction, and **transcription and translation** of, e.g., interviews. Here, there seems to be very little differences in policies – either, these use cases are mentioned in the policies as allowed or encouraged, or they are not explicitly mentioned, none of the news organisations forbids this type of AI use. In a similar vein, Süddeutsche Zeitung (SZ) explicitly mention the use of AI for **content moderation**, showing the wide range of possible AI application in news organisations. The second set of proposed use cases of AI in news organisations is around the content of the published news, i.e., the use of AI for content production. Here policies diverge. Interestingly, while many news organisations explicitly allow the use of AI for the generation of **illustrations/graphics**, Financial Times (FT) and VG prohibit the use of AI for **image generation**, i.e., the creation and publication of photorealistic images whereas ANP, Le Parisien, and Der Spiegel allow use for image generation. Likewise, policies differ for **article**

generation, i.e., the use of AI to create full or long parts of texts for news articles. While ANP and Der Spiegel allow article creation under human supervision, Financial Times (FT), Le Parisien, and Süddeutsche Zeitung (SZ) explicitly state that their articles are exclusively written by humans. In the realm of ideation, only ANP encourages the use of AI to support journalists in finding ideas on headlines, articles, and even sources. The Guardian does not mention idea generation for news content, but they do encourage the use of AI for marketing campaign ideas.

4.2 Issued Warning and Rules

All news organisations mandate **human oversight**, requiring editorial approval before AI-generated content is published, and emphasise AI as a supportive tool rather than an independent generator. They also agree on the need for **declaration of use**, mandating clear labelling of AI-generated content.

Broader concerns such as **accuracy** and **bias** are widely acknowledged, but concrete measures are rare. Only the BBC requires checks for "accuracy and reliability," while The Guardian and Mediahuis propose safeguards against bias in models and training data.

For journalists, sensitive data handling is a crucial issue. **Privacy and sensitive data** policies appear in four organisations, banning entering information about sources, staff, or partners into AI systems (e.g., Ringier), and BBC extending this to privacy-respecting AI tools. **Copyright** is addressed by half of the outlets, though often narrowly (e.g., Le Parisien on AI images, Ringier on code). The Guardian stands out advocating fair compensation for creators whose data train AI models.

Several organisations support **AI literacy training** to promote responsible use: Mediahuis (staff accountability and qualification), BBC (transparency about AI use and data), and FT (AI for story discovery). Finally, accountability measures differ: Der Spiegel and BBC call for documentation, but only the FT requires an internal register to systematically track AI use.

5 AI Policies in University Education

Table 2 presents a summary of the suggested use cases of AI in teaching and learning as well as the risks and rules enacted in the corresponding institutions. The table presents the points raised by at least two universities, we also discuss points raised by individual universities in the following section.

5.1 Suggested Uses

Suggested uses of AI systems in education are framed from two primary perspectives: teachers and the students. The most commonly proposed application is **self-learning**, mentioned in seven organisational guidelines. Examples of such activities supported by AI systems include providing students with learning materials and resources, assisting in planning and monitoring progress, encouraging exploration of covered topics (TUM), and offering alternative perspectives (Aalto).

From the student perspective, AI systems are also frequently viewed as tools for enabling **personalised learning**, e.g. by recommending individualised learning plans, presenting material with explanations of varying difficulty levels, and enhancing accessibility for students with disabilities (ETH). Additionally, AI is suggested as highly effective for **coding tasks**, such as understanding concepts, breaking down problems into smaller parts, practising debugging skills, and receiving feedback on code. University guidelines further note that students can leverage AI to **gain an initial overview of a topic**, terms, or concepts. Moreover, AI can help students monitor their progress and **provide targeted feedback** on written work or ideas.

AI systems are also commonly suggested as **language tools**, particularly for grammar, spelling, and reference checks—often considered the safest and most permissible use of AI. For example, KTH identifies this as one of the three approved uses of AI, typically not requiring documentation in the declaration of use. Additional proposed applications for students include **idea generation, summarising academic literature, translation, student assignment assessment, search engine functionality, and image generation**. However, regarding translation, Vie cautions that AI may produce inaccuracies, particularly with new terminology or less common language combinations.

From the teacher's perspective, suggested uses of AI extend to **course design** activities, such as formulating learning objectives, drafting rubrics, planning workshops, designing assignments,

Code	UiO	Aalto	TUD	KUL	ETH	TUM	Vie	CUNI	DTU	KTH
Suggested Uses										
Self-learning	✅	✅	✅	-	-	✅	✅	-	✅	🔒
Course design	✅	✅	-	-	✅	✅	-	✅	✅	-
Personalisation	✅	✅	-	-	✅	✅	✅	-	✅	-
Coding	✅	✅	✅	🔒	-	✅	-	✅	-	-
Topic knowledge	✅	✅	✅	✅	-	-	✅	-	-	✅
Feedback	✅	✅	✅	-	✅	-	✅	-	-	-
Language tool	-	✅	✅	✅	✅	-	-	-	-	🔒
Ideas	✅	✅	✅	✅	✅	-	-	-	-	-
Summarisation	-	✅	✅	✅	-	-	✅	✅	-	-
Translation	✅	✅	-	-	-	-	⚠️	✅	-	-
Assessment	✅	-	✅	-	🔒	-	-	✅	-	-
Search	🔒	-	-	✅	-	-	-	✅	-	-
Image generation	-	-	-	✅	⚠️	-	-	✅	-	-
Issued Warning and Rules										
Teacher restrictions	📜	📜	-	📜	📜	📜	📜	📜	-	-
AI literacy training	⚠️✅	-	✅	⚠️✅	✅	⚠️✅	⚠️✅	✅	-	-
Human oversight	📜	📜	-	📜	📜	-	📜	📜	📜	📜
Privacy and sensitive data	📜	⚠️/📜	⚠️	⚠️/📜	📜	⚠️	⚠️	⚠️/📜	⚠️/📜	-
Factual accuracy	⚠️	⚠️	⚠️	⚠️	⚠️	⚠️	⚠️	-	⚠️	⚠️
Declaration of use	-	📜	⚠️/📜	📜	📜	-	📜	📜	📜	📜
Copyright	⚠️/📜	-	⚠️	⚠️/📜	📜	⚠️	⚠️/📜	⚠️/📜	⚠️	-
Bias in AI	-	⚠️	-	⚠️	⚠️	⚠️	-	-	⚠️	-
No prioritisation	⚠️	-	-	⚠️	-	⚠️	⚠️/📜	-	-	⚠️
Digital inequity	⚠️	-	⚠️	⚠️	-	-	-	⚠️	-	-
Knowledge cut-off	⚠️	⚠️	⚠️	⚠️	-	⚠️	-	-	-	-
Persuasiveness	⚠️	⚠️	⚠️	-	-	⚠️	-	-	-	⚠️
Source attribution	-	⚠️	⚠️	-	⚠️	⚠️	⚠️	-	-	-
Skills assessment	⚠️	-	⚠️	-	⚠️	⚠️	-	-	-	-
Environment	⚠️	-	⚠️	⚠️	-	-	-	-	-	-
AI over-reliance	⚠️	-	-	-	-	-	⚠️	-	-	-
Teacher load	-	-	⚠️	-	-	-	-	-	⚠️	-

Table 2: Suggested uses and issued *warnings and rules* regarding AI in university teaching and learning. ✅= encouraged use, ⚠️= cautionary warning, 📜= formal rule, 🔒= restriction.

creating questions, preparing courses, and even simulating a test student. Regarding assessment, ETH explicitly disallows the fully automated grading of student work, but most universities grant teachers the discretion to decide whether **AI use is restricted** in assignments and exams.

While many universities provide lists of potential AI applications in teaching and learning, some impose restrictions on those applications. For instance, KTH limits AI usage to predefined prompts and prohibits directly asking the system to generate specific answers or complete assignments. Similarly, KUL restricts the use of AI in coding tasks to generating components of larger assignments, and only if explicitly approved by the teacher.

5.2 Issued Warnings and Rules

The most common warnings and rules in university policies regarding AI include considerations of **privacy** and **sensitive data**, **copyright**, **factual accuracy**, and the **declaration of use**. While some universities encourage students and teachers to consider privacy and copyright concerns and warn of potential violations, others enforce strict rules regarding the types of data that can be input into AI systems to prevent such issues. For example, Aalto specifies that teachers may only submit student work to university-approved systems and prohibits entering other students' answers or personal information into external systems. To support these policies, Aalto, along with DTU, KUL, and ETH, provides access to Microsoft Copilot for both teachers and students, which is meant to ensure that submitted data is not stored or used for future model training. Regarding copyright, universities caution that AI can reproduce copyrighted material without proper acknowledgement. They also require users to avoid inputting proprietary information, such as the university's intellectual property, into AI tools.

Other warnings address risks associated with the quality of generated content, including **biases** in the content, **lack of prioritisation** of arguments, **absence of source attribution** (making verification of accuracy, plagiarism, etc., difficult), **limited knowledge due to cut-off dates**, and a **persuasive tone** that can mislead users about the correctness of the information. Risks also arise in educational scenarios where AI integration might **necessitate course reorganisation or additional assessments** to ensure learning objectives are met and student skills are accurately evaluated. AI usage may also contribute to **digital inequity**, stemming from disparities in access to paid versus free tools and variations in the quality of generated content based on user skills. **Over-reliance on AI tools** is another identified risk.

KUL explicitly highlights the lack of reproducibility as a concern, noting that output can vary between attempts. UiO is unique in warning that AI can produce inappropriate or offensive content. KUL also advises against "humanising" AI, emphasising that it is merely a tool.

Many universities stress the **importance of asking the right questions** and not settling for the first answer provided by AI. To support this, they offer guides for crafting effective prompts, and some institutions even provide dedicated courses on this topic. Teachers are encouraged to **set clear restrictions on AI use** within their courses and must communicate allowed uses transparently. Finally, most guidelines underscore that both teachers and students remain **fully responsible** for the content they incorporate into their work, regardless of AI use.

6 Insights from Organisation Policies for the EU AI Act

We now put the above findings from organisational policies in the perspective of EU AI Act European Parliament [2024] (`EU-AIA`). We recognise that these efforts towards AI governance are fundamentally different in scope and process through which they were created, and they are complementary. However, given the ongoing effort[5] to develop a more specific implementation guidelines for `EU-AIA`, we believe that the insights from the bottom-up policies could help to identify areas where more clarity would be appreciated. We relate these to known research challenges in Appendix Section **??**

6.1 News Organisations' Policies

Similarities. News organisations' policies align with the provisions introduced by the `EU-AIA` for example on the emphasises on *human oversight* with the `EU-AIA` Article 14. Article 50 of the `EU-AIA` describes transparency obligations for providers and deployers, aligning with the requirements to *disclose AI-generated content* by news organisations. For high-risk AI systems, Article 10 of the `EU-AIA` regulates *data governance*, a question that also is relevant to downstream users such as journalists. In the context of the `EU-AIA` this is limited to training, validation, and test data, whereas for news organisations the question of input of sensitive data into AI systems is crucial to preserve privacy of possible sources, reflecting the broader GDPR compliance required under the `EU-AIA`.

Gaps. The news' AI policies cover several points not covered by the `EU-AIA`. In particular, The Guardian emphasises *compensating those whose data is used for AI*, while the `EU-AIA` lacks explicit provisions for data creators' compensation outside of existing copyright regulations. Financial Times and Mediahuis include newsroom *AI training* in their policies. The EU Act currently only requires AI literacy training for developers and deployers of AI systems. If news organisations as downstream users of these systems are not considered as deployers, this requirement will not cover this type of distribution of AI-generated content. Multiple policies require addressing *bias in the AI systems* used by journalists, a topic that is yet to be comprehensively covered by legislation. The *internal AI usage register policy* by Financial Times is an additional documentation practice not specified by the Act but useful for accountability. It could enable retroactive checks on which content was created with which AI system and where the systems where used, which would improve transparency and accountability of this outlet, and hence potentially increase trust in it.

6.2 University Education's Policies

Similarities. University AI guidelines closely align with the EU AI Act's provisions, particularly regarding transparency, human oversight, and data privacy. Universities' emphasis on protecting personal data and advising against uploading sensitive information to AI tools reflects Article 10 of the AI Act's data governance requirements and broader GDPR compliance. The requirement

[5] `https://artificialintelligenceact.eu/ai-act-implementation-next-steps/`

for human oversight in AI-driven assessments and exams mirrors Article 14's mandate for human monitoring in high-risk AI systems, with universities either prohibiting AI in exams entirely or implementing strict oversight to ensure AI doesn't replace independent student evaluation. Since the AI Act classifies education as a high-risk area (Article 6 Annex III), universities are addressing concerns about bias mitigation and fairness, particularly regarding potential grading biases and manipulation tactics in AI assessments, while also exploring applications like adaptive learning and dropout risk prediction—the latter presenting challenges in balancing improved performance with regulatory compliance requirements.

Gaps. University AI policies address critical issues absent from or not explicitly covered by the AIA, including the environmental impact and sustainability of AI use (highlighted by institutions like UiO, TUD, and KUL) and the risk of digital inequity among students arising from disparities in access to paid tools and variations in content quality based on user skills. While the AIA's Article 4 mandates AI literacy training for developers and deployers, university policies extend this responsibility to educators and students, encouraging critical evaluation of AI outputs and providing resources for effective AI interaction – a proactive approach that contrasts with the AIA's narrower focus on professional users. Overall, universities adopt a context-specific governance approach that addresses academic integrity, assessment reliability, and pedagogical challenges in ways that current EU regulation does not yet fully capture, complementing the AIA's legal framework for AI safety, transparency, and human oversight with education-specific considerations.

7 Policy Recommendations Based on Gaps Identified in This Work

To reiterate, according to the legal endogeneity theory, the local organisational practices and regulatory efforts are constantly impacting each other. The above practices in universities and news organisations highlight many themes, but of particular interest are the themes where the best practice recommendations for organisations depend on a higher transparency from the model providers. Since that currently goes against the market incentives[6], we identify the following as the areas most in need of further research and consideration for top-down governance efforts.

Expanding AI literacy training. The `EU-AIA` mandates AI literacy training for developers and deployers (Article 4), but it does not mandate AI literacy for students, teachers, journalists, the general public who generate or interact with AI-generated content, or even the media professionals or other users of AI models distributing their outputs on a large scale. Universities integrate AI literacy into academic policies, requiring students and educators to develop critical engagement with AI tools. Institutions like CUNI make a proactive step in this direction emphasising the education in AI ethics and responsible use. Newsrooms such as Financial Times and Mediahuis provide AI training for journalists, ensuring that AI-generated content is fact-checked and responsibly handled. Such training should include also critical reflection on the real functionality of the current AI models vs the marketing hype. While it is important to keep responsibility with the AI developers and deployers, supporting users on how to approach AI will be critical.

Policies addressing Digital Inequity. The `EU-AIA` does not explicitly address digital inequity, despite its potential to exacerbate social and economic disparities (e.g. due to unequal access to AI and different quality of the models available for different socio-economic and linguistic groups). This is particularly relevant in education, where university AI policies have highlighted concerns about disparities in access to paid vs. free AI tools, as well as differences in the quality of AI-generated content based on user skills. Other concerns include the temptation to use AI 'education' as a low-cost substitute for human teachers for the already underpriviledged groups, and siphoning of public resources to for-profit AI providers instead of building public AI infrastructure. All this requires more thought to develop more equitable education infrastructure and policies that consider socio-economic impact on various population groups.

Improved Transparency for Generated Content. While `EU-AIA` Article 52 mandates disclosure of AI-generated content, it does not specify how AI-assisted work should be attributed or audited by downstream users, or how the record of AI assistance should be kept. Universities enforce strict

[6]Some of these themes were part of AIA debates but got watered down by AI provider companies Wachter [2024]. However, the fundamental discrepancies in the organisational policies, such as whether or not the use of non-ethically-sourced model services is allowed.

AI citation rules, requiring students to disclose AI-generated content to uphold academic integrity. However, there is no standardised framework for disclosing AI use, which could aid AI literacy across sectors. For example, in student submissions (e.g. should it be a brief description, or a full prompt+output? How should the source system be specified?) An interesting practice is the internal AI usage registers in Financial Times, which allows editors to track which articles were AI-assisted.

Getting specific about 'bias'. Both news and university policies sometimes warn about the possibility of 'bias', but it is not clear what kinds of bias should be addressed, or how. This is a gap legislation could address by providing guidance (e.g. based on existing human rights and non-discrimination laws) for model providers, deployers, as well as downstream users and content distributors.

Attribution and compensation of sources of AI training data. EU has copyright laws, but AI training data poses new challenges currently tested in both US Vynck [2023], Grynbaum and Mac [2023] and in Europe Smith [2024]. In our sample, only The Guardian advocates for compensation of content creators whose data is used as part of AI training. In education, an equally important factor is source attribution, without which the students could be unwittingly plagiarising existing scholarly work. The question of data governance and compensation should be further investigated, taking into account concepts such as data collectives Hsieh et al. [2024].

Disclosing Environmental Impact of AI. The Act does not explicitly address the carbon footprint of AI models besides documentation, despite researchers' concerns about large-scale computational demands Dodge et al. [2022], Bouza et al. [2023], Luccioni and Hernandez-Garcia [2023], Li et al. [2023]. Some universities and news outlets highlight the environmental costs of training and running large models, yet there are no regulatory incentives to optimise for sustainability. One ongoing research direction is developing more efficient models Treviso et al. [2023], but if the more efficient models just get used more, this will not help. Mandating a visible disclosure to the users of how much water and energy each AI query costs, and where the resources come from, could help to discourage excessive use. Organisations could also have AI use by their employees as a part of their sustainability reporting.

Detection and enforceability. There are currently no reliable methods of detecting generated text, which makes any policies unenforceable. A promising solution is watermarking Jiang et al. [2024], Roman et al. [2024], but the providers of commercial LLMs have no incentive to provide a mechanism that could reduce the usage of their services Davis [2024], Gloaguen et al. [2024]. This is where the considerations of social impact Solaiman et al. [2023] should guide regulation mandating such transparency from the popular AI service providers.

Clarifying liability. In compliance with `EU-AIA`, providers of AI models may attempt to build in "safeguards" to avoid e.g. toxic outputs, and they will have to pass some evaluations to put the model on the market. But should something go wrong, e.g. seriously impacting the mental health of the user, it is currently not clear how much legal recourse the affected users would have. The question of AI liability European Comission [2025] will get more pressing with the amount of cases that pose the question of responsibility for the consequences of AI use Dao et al. [2022].

8 Conclusion

In conclusion, the rapid adoption of AI technologies across diverse domains has exposed significant gaps in governance, with multiple organisations scrambling to quickly develop their policies. Our comparative analysis of AI guidelines in universities and news organisations highlights both shared and domain-specific challenges, such as the need for clear accountability mechanisms, addressing biases, and managing domain-specific applications like personalised learning and content generation. We have also identified multiple challenges that are known in academic research, but not addressed by the current policies. These findings underscore the fragmented nature of current governance efforts and the critical need for cohesive policies that balance local organisational needs with broader societal and technological imperatives, while recognising and supporting areas where more research is needed for better policies. By identifying these gaps and challenges, this paper offers actionable insights for refining legislation and informs the critical future directions of research. Ultimately, bridging the disconnect between local practices, academic research, and global standards is essential for ensuring the safe, fair, and effective deployment of AI across diverse contexts.

References

Alfonso Amayuelas, Kyle Wong, Liangming Pan, Wenhu Chen, and William Yang Wang. Knowledge of knowledge: Exploring known-unknowns uncertainty with large language models. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, *Findings of the Association for Computational Linguistics: ACL 2024*, pages 6416–6432, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-acl.383. URL `https://aclanthology.org/2024.findings-acl.383/`.

Markus Anderljung, Joslyn Barnhart, Anton Korinek, Jade Leung, Cullen O'Keefe, Jess Whittlestone, Shahar Avin, Miles Brundage, Justin Bullock, Duncan Cass-Beggs, Ben Chang, Tantum Collins, Tim Fist, Gillian Hadfield, Alan Hayes, Lewis Ho, Sara Hooker, Eric Horvitz, Noam Kolt, Jonas Schuett, Yonadav Shavit, Divya Siddarth, Robert Trager, and Kevin Wolf. Frontier AI Regulation: Managing Emerging Risks to Public Safety, November 2023.

Pepa Atanasova, Oana-Maria Camburu, Christina Lioma, Thomas Lukasiewicz, Jakob Grue Simonsen, and Isabelle Augenstein. Faithfulness tests for natural language explanations. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 283–294, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.acl-short.25. URL `https://aclanthology.org/2023.acl-short.25/`.

Authors Guild Foundation. More than 15,000 Authors Sign Authors Guild Letter Calling on AI Industry Leaders to Protect Writers, 2025. URL `https://authorsguild.org/news/thousands-sign-authors-guild-letter-calling-on-ai-industry-leaders-to-protect-writers/`.

Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 610–623, 2021.

Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in neural information processing systems*, 29, 2016.

Lucía Bouza, Aurélie Bugeau, and Loïc Lannelongue. How to estimate carbon footprint when training deep learning models? a guide and review. *Environmental Research Communications*, 5 (11):115014, 2023.

Blake Brittain and Blake Brittain. Lawsuits accuse AI content creators of misusing copyrighted work. *Reuters*, January 2023. URL `https://www.reuters.com/legal/transactional/lawsuits-accuse-ai-content-creators-misusing-copyrighted-work-2023-01-17/`.

Marta Cantero Gamito and Christopher T Marsden. Artificial intelligence co-regulation? the role of standards in the eu ai act. *International Journal of Law and Information Technology*, 32(1): eaae011, 2024.

Stephen Casper, Xander Davies, Claudia Shi, Thomas Krendl Gilbert, Jérémy Scheurer, Javier Rando, Rachel Freedman, Tomek Korbak, David Lindner, Pedro Freire, Tony Tong Wang, Samuel Marks, Charbel-Raphael Segerie, Micah Carroll, Andi Peng, Phillip J.K. Christoffersen, Mehul Damani, Stewart Slocum, Usman Anwar, Anand Siththaranjan, Max Nadeau, Eric J Michaud, Jacob Pfau, Dmitrii Krasheninnikov, Xin Chen, Lauro Langosco, Peter Hase, Erdem Biyik, Anca Dragan, David Krueger, Dorsa Sadigh, and Dylan Hadfield-Menell. Open problems and fundamental limitations of reinforcement learning from human feedback. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL `https://openreview.net/forum?id=bx24KpJ4Eb`. Survey Certification, Featured Certification.

Thomas KF Chiu. The impact of generative ai (genai) on practices, policies and research direction in education: A case of chatgpt and midjourney. *Interactive Learning Environments*, pages 1–17, 2023.

Hannes Cools. Towards guidelines for guidelines on the use of generative ai in newsrooms. *Generative AI in the Newsroom (Medium)*, 2023. URL https://generative-ai-newsroom.com/towards-guidelines-for-guidelines-on-the-use-of-generative-ai-in-newsrooms-55b0c2c1d960.

David Dao, Hugo W., Alf-André Walla, VocalFan, Eric Rubiel, Fabian Schreiber, Jan Jaworski, Liu Liu, Nate Williams, Nick Pawlowski, Pawan Sasanka Ammanamanchi, Stanislaus Stadlmann, Steffen Kühne, and Tom Diethe. daviddao/awful-ai: Awful ai - 2021 edition (v1.0.0), 2022.

Wes Davis. OpenAI won't watermark ChatGPT text because its users could get caught, August 2024. URL https://www.theverge.com/2024/8/4/24213268/openai-chatgpt-text-watermark-cheat-detection-tool.

MF de Lima-Santos and W Ceron. Artificial intelligence in news media: Current perceptions and future outlook. journalism and media, 3 (1), 13–26, 2022.

Joost CF de Winter. Can chatgpt pass high school exams on english language comprehension? *International Journal of Artificial Intelligence in Education*, pages 1–16, 2023.

Leon Derczynski, Erick Galinkin, Jeffrey Martin, Subho Majumdar, and Nanna Inie. Garak: A Framework for Security Probing Large Language Models, June 2024.

Jesse Dodge, Taylor Prewitt, Remi Tachet des Combes, Erika Odmark, Roy Schwartz, Emma Strubell, Alexandra Sasha Luccioni, Noah A Smith, Nicole DeCario, and Will Buchanan. Measuring the carbon intensity of ai in cloud instances. In *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency*, pages 1877–1894, 2022.

Ravit Dotan, Lisa S Parker, and John Radzilowicz. Responsible adoption of generative ai in higher education: Developing a "points to consider" approach based on faculty perspectives. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, pages 2033–2046, 2024.

Lauren B Edelman. Overlapping fields and constructed legalities: The endogeneity of law. *Private equity, corporate governance and the dynamics of capital market regulation*, 55:58, 2007.

Lauren B Edelman, Christopher Uggen, and Howard S Erlanger. The endogeneity of legal regulation: Grievance procedures as rational myth. *American journal of Sociology*, 105(2):406–454, 1999.

European Comission. Liability Rules for Artificial Intelligence, 2025. URL https://commission.europa.eu/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en.

European Parliament. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance), June 2024. URL http://data.europa.eu/eli/reg/2024/1689/oj/eng.

Mohammadreza Farrokhnia, Seyyed Kazem Banihashem, Omid Noroozi, and Arjen Wals. A swot analysis of chatgpt: Implications for educational practice and research. *Innovations in Education and Teaching International*, 61(3):460–474, 2024.

Joseph Saveri Law Firm and Matthew Butterick. GitHub Copilot investigation, 2022. URL https://www.saverilawfirm.com/.

Paul Formosa, Sarah Bankins, Rita Matulionyte, and Omid Ghasemi. Can ChatGPT be an author? Generative AI creative writing assistance and perceptions of authorship, creatorship, responsibility, and disclosure. *AI & SOCIETY*, September 2024. ISSN 1435-5655. doi: 10.1007/s00146-024-02081-0.

Aashish Ghimire and John Edwards. From guidelines to governance: A study of ai policies in education. In *International Conference on Artificial Intelligence in Education*, pages 299–307. Springer, 2024.

Thibaud Gloaguen, Nikola Jovanović, Robin Staab, and Martin Vechev. Black-Box Detection of Language Model Watermarks. In *ICML 2024 Workshop on Foundation Models in the Wild*, July 2024. URL `https://openreview.net/forum?id=SgrEUyZia6`.

Jeremy Gray. OpenAI Claims it is Impossible to Train AI Without Using Copyrighted Content, January 2024. URL `https://petapixel.com/2024/01/09/openai-claims-it-is-impossible-to-train-ai-without-using-copyrighted-content/`.

Michael M. Grynbaum and Ryan Mac. The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work. *The New York Times*, December 2023. ISSN 0362-4331. URL `https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html`.

Melissa Heikkilä. This artist is dominating AI-generated art. And he's not happy about it., September 2022. URL `https://www.technologyreview.com/2022/09/16/1059598/this-artist-is-dominating-ai-generated-art-and-hes-not-happy-about-it/`.

Michael Townsen Hicks, James Humphries, and Joe Slater. Chatgpt is Bullshit. *Ethics and Information Technology*, 26(2):1–10, 2024. doi: 10.1007/s10676-024-09775-5.

Wayne Holmes, Fengchun Miao, et al. *Guidance for generative AI in education and research*. UNESCO Publishing, 2023.

Sara Hooker. On the Limitations of Compute Thresholds as a Governance Strategy, July 2024.

Jane Hsieh, Angie Zhang, Seyun Kim, Varun Nagaraj Rao, Samantha Dalal, Alexandra Mateescu, Rafael Do Nascimento Grohmann, Motahhare Eslami, and Haiyi Zhu. Worker data collectives as a means to improve accountability, combat surveillance and reduce inequalities. In Rosta Farzan, Claudia López, Daniel Cardoso Llach, Daniele Quercia, Maryam Mustafa, Shuo Niu, and Marisol Wong-Villacrés, editors, *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing, CSCW Companion 2024, San Jose, Costa Rica, November 9-13, 2024*, pages 697–700. ACM, 2024. doi: 10.1145/3678884.3681829. URL `https://doi.org/10.1145/3678884.3681829`.

Ben Hutchinson, Vinodkumar Prabhakaran, Emily Denton, Kellie Webster, Yu Zhong, and Stephen Denuyl. Social biases in NLP models as barriers for persons with disabilities. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5491–5501, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.487. URL `https://aclanthology.org/2020.acl-main.487/`.

Umar Iqbal, Tadayoshi Kohno, and Franziska Roesner. Llm platform security: Applying a systematic evaluation framework to openai's chatgpt plugins. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 7, pages 611–623, 2024.

Zhengyuan Jiang, Moyang Guo, Yuepeng Hu, and Neil Zhenqiang Gong. Watermark-based detection and attribution of ai-generated content. *arXiv preprint arXiv:2404.04254*, 2024.

Brian Judge, Mark Nitzberg, and Stuart Russell. When code isn't law: rethinking regulation for artificial intelligence. *Policy and Society*, page puae020, 05 2024. ISSN 1449-4035. doi: 10.1093/polsoc/puae020. URL `https://doi.org/10.1093/polsoc/puae020`.

Rohan Khera, Melissa A. Simon, and Joseph S. Ross. Automation Bias and Assistive AI: Risk of Harm From AI-Driven Clinical Decision Support. *JAMA*, 330(23):2255–2257, December 2023. ISSN 0098-7484. doi: 10.1001/jama.2023.22557.

Eugene Kim. Amazon warns employees not to share confidential information with ChatGPT after seeing cases where its answer 'closely matches existing material' from inside the company, January 2023. URL `https://www.businessinsider.com/amazon-chatgpt-openai-warns-employees-not-share-confidential-information-microsoft-2023-1`.

Jan Kocoń, Igor Cichecki, Oliwier Kaszyca, Mateusz Kochanek, Dominika Szydło, Joanna Baran, Julita Bielaniewicz, Marcin Gruza, Arkadiusz Janz, Kamil Kanclerz, Anna Kocoń, Bartłomiej Koptyra, Wiktoria Mieleszczenko-Kowszewicz, Piotr Miłkowski, Marcin Oleksy, Maciej Piasecki, Łukasz Radliński, Konrad Wojtasik, Stanisław Woźniak, and Przemysław Kazienko. Chatgpt: Jack of all trades, master of none. *Information Fusion*, 99:101861, 2023. ISSN 1566-2535. doi: https://doi.org/10.1016/j.inffus.2023.101861. URL `https://www.sciencedirect.com/science/article/pii/S156625352300177X`.

Harsh Kumar, Jonathan Vincentius, Ewan Jordan, and Ashton Anderson. Human creativity in the age of llms: Randomized experiments on divergent and convergent thinking, 2024. URL `https://arxiv.org/abs/2410.03703`.

Tamera Lanham, Anna Chen, Ansh Radhakrishnan, Benoit Steiner, Carson Denison, Danny Hernandez, Dustin Li, Esin Durmus, Evan Hubinger, Jackson Kernion, et al. Measuring faithfulness in chain-of-thought reasoning. *arXiv preprint arXiv:2307.13702*, 2023.

Johann Laux, Sandra Wachter, and Brent Mittelstadt. Trustworthy artificial intelligence and the european union ai act: On the conflation of trustworthiness and acceptability of risk. *Regulation & Governance*, 18(1):3–32, 2024.

Pengfei Li, Jianyi Yang, Mohammad A. Islam, and Shaolei Ren. Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models, October 2023.

Yao Lu, Max Bartolo, Alastair Moore, Sebastian Riedel, and Pontus Stenetorp. Fantastically Ordered Prompts and Where to Find Them: Overcoming Few-Shot Prompt Order Sensitivity. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8086–8098, Dublin, Ireland, May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-long.556.

Alexandra Sasha Luccioni and Alex Hernandez-Garcia. Counting Carbon: A Survey of Factors Influencing the Emissions of Machine Learning, February 2023.

Marta Marchiori Manerba, Karolina Stanczak, Riccardo Guidotti, and Isabelle Augenstein. Social bias probing: Fairness benchmarking for language models. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 14653–14671, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.emnlp-main.812. URL `https://aclanthology.org/2024.emnlp-main.812/`.

Guillermo Marco, Julio Gonzalo, M.Teresa Mateo-Girona, and Ramón Del Castillo Santos. Pron vs prompt: Can large language models already challenge a world-class fiction author at creative text writing? In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 19654–19670, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.emnlp-main.1096. URL `https://aclanthology.org/2024.emnlp-main.1096/`.

R. Thomas McCoy, Shunyu Yao, Dan Friedman, Mathew D. Hardy, and Thomas L. Griffiths. Embers of autoregression show how large language models are shaped by the problem they are trained to solve. *Proceedings of the National Academy of Sciences*, 121(41):e2322420121, October 2024a. doi: 10.1073/pnas.2322420121.

R. Thomas McCoy, Shunyu Yao, Dan Friedman, Mathew D. Hardy, and Thomas L. Griffiths. When a language model is optimized for reasoning, does it still show embers of autoregression? An analysis of OpenAI o1, October 2024b.

Dhruv Mehrotra. Perplexity Is a Bullshit Machine. *Wired*, 2024. ISSN 1059-1028. URL `https://www.wired.com/story/perplexity-is-a-bullshit-machine/`.

Melanie Mitchell. Why AI is Harder Than We Think. *arXiv:2104.12871 [cs]*, April 2021. URL `http://arxiv.org/abs/2104.12871`.

Moin Nadeem, Anna Bethke, and Siva Reddy. StereoSet: Measuring stereotypical bias in pretrained language models. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 5356–5371, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.acl-long.416. URL `https://aclanthology.org/2021.acl-long.416/`.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022.

Vishakh Padmakumar and He He. Does writing with language models reduce content diversity? In *The Twelfth International Conference on Learning Representations*, 2024. URL `https://openreview.net/forum?id=Feiz5HtCD0`.

Aklovya Panwar. Generative AI and Copyright Issues Globally: ANI Media v OpenAI | TechPolicy.Press, January 2025. URL `https://techpolicy.press/generative-ai-and-copyright-issues-globally-ani-media-v-openai`.

Stanislaw Piasecki, Sophie Morosoli, Natali Helberger, and Laurens Naudts. Ai-generated journalism: Do the transparency provisions in the ai act give news readers what they hope for? *Internet Policy Review*, 13(4), 2024.

Magda Pischetola, Maria Hvid Stenalt, Liv Nøhr, Danielle Elizabeth Hagood, and Morten Misfeldt. Desirable and realistic futures of the university: a mixed-methods study with teachers in denmark. *International Journal of Educational Technology in Higher Education*, 21(1):29, 2024.

Giovanni Puccetti, Anna Rogers, Chiara Alzetta, Felice Dell'Orletta, and Andrea Esuli. AI 'News' Content Farms Are Easy to Make and Hard to Detect: A Case Study in Italian. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15312–15338, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.acl-long.817.

Junaid Qadir. Engineering education in the era of chatgpt: Promise and pitfalls of generative ai for education. In *2023 IEEE Global Engineering Education Conference (EDUCON)*, pages 1–9. IEEE, 2023.

Anna Rogers and Sasha Luccioni. Position: Key Claims in LLM Research Have a Long Tail of Footnotes. In *Forty-First International Conference on Machine Learning*, June 2024. URL `https://openreview.net/forum?id=M2cwkGleRL`.

Robin San Roman, Pierre Fernandez, Alexandre Défossez, Teddy Furon, Tuan Tran, and Hady Elsahar. Proactive detection of voice cloning with localized watermarking. *arXiv preprint arXiv:2401.17264*, 2024.

Nikhil Sharma, Q Vera Liao, and Ziang Xiao. Generative echo chamber? effect of llm-powered search systems on diverse information seeking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2024.

Chenglei Si, Diyi Yang, and Tatsunori Hashimoto. Can llms generate novel research ideas? a large-scale human study with 100+ nlp researchers. *arXiv preprint arXiv:2409.04109*, 2024.

Felix M Simon and Luisa Fernanda Isaza-Ibarra. Ai in the news: reshaping the information ecosystem? *Oxford Internet Institute, University of Oxford*, 2023.

FM Simon and KB Becker. Policies in parallel? a comparative study of journalistic ai policies in 52 global news organisations. *SocArXiv*, 2023.

Zouhaier Slimi and Beatriz Villarejo Carballido. Navigating the ethical challenges of artificial intelligence in higher education: An analysis of seven global ai ethics policies. *TEM Journal*, 12 (2), 2023.

Dylan Smith. GEMA Files Copyright Lawsuit Against OpenAI in Germany, November 2024. URL `https://www.digitalmusicnews.com/2024/11/13/gema-openai-lawsuit/`.

Irene Solaiman, Zeerak Talat, William Agnew, Lama Ahmad, Dylan Baker, Su Lin Blodgett, Canyu Chen, Hal Daumé III, Jesse Dodge, Isabella Duan, et al. Evaluating the social impact of generative ai systems in systems and society. *arXiv preprint arXiv:2306.05949*, 2023.

Karolina Stańczak, Sagnik Ray Choudhury, Tiago Pimentel, Ryan Cotterell, and Isabelle Augenstein. Quantifying gender bias towards politicians in cross-lingual language models. *Plos one*, 18(11): e0277640, 2023.

Teo Susnjak and Timothy R McIntosh. Chatgpt: The end of online exam integrity? *Education Sciences*, 14(6):656, 2024.

TJ Thomson, Ryan J Thomas, and Phoebe Matich. Generative visual ai in news organizations: Challenges, opportunities, perceptions, and policies. *Digital Journalism*, pages 1–22, 2024.

Marcos Treviso, Ji-Ung Lee, Tianchu Ji, Betty van Aken, Qingqing Cao, Manuel R. Ciosici, Michael Hassid, Kenneth Heafield, Sara Hooker, Colin Raffel, Pedro H. Martins, André F. T. Martins, Jessica Zosa Forde, Peter Milder, Edwin Simpson, Noam Slonim, Jesse Dodge, Emma Strubell, Niranjan Balasubramanian, Leon Derczynski, Iryna Gurevych, and Roy Schwartz. Efficient Methods for Natural Language Processing: A Survey. *Transactions of the Association for Computational Linguistics*, 11:826–860, July 2023. ISSN 2307-387X. doi: 10.1162/tacl_a_00577.

Gerrit De Vynck. 'Game of Thrones' author and others accuse ChatGPT maker of 'theft' in lawsuit. *Washington Post*, September 2023. ISSN 0190-8286. URL `https://www.washingtonpost.com/technology/2023/09/20/openai-lawsuit-martin-picoult-franzen/`.

Sandra Wachter. Limitations and loopholes in the eu ai act and ai liability directives: what this means for the european union, the united states, and beyond. *Yale Journal of Law and Technology*, 26(3), 2024.

Laura Weidinger, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, Courtney Biles, Sasha Brown, Zac Kenton, Will Hawkins, Tom Stepleton, Abeba Birhane, Lisa Anne Hendricks, Laura Rimell, William Isaac, Julia Haas, Sean Legassick, Geoffrey Irving, and Iason Gabriel. Taxonomy of risks posed by language models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, page 214–229, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450393522. doi: 10.1145/3531146.3533088. URL `https://doi.org/10.1145/3531146.3533088`.

Fangzhou Wu, Ning Zhang, Somesh Jha, Patrick McDaniel, and Chaowei Xiao. A New Era in LLM Security: Exploring Security Concerns in Real-World LLM-based Systems, February 2024.

Yue Zhang, Yafu Li, Leyang Cui, Deng Cai, Lemao Liu, Tingchen Fu, Xinting Huang, Enbo Zhao, Yu Zhang, Yulong Chen, Longyue Wang, Anh Tuan Luu, Wei Bi, Freda Shi, and Shuming Shi. Siren's song in the ai ocean: A survey on hallucination in large language models, 2023. URL `https://arxiv.org/abs/2309.01219`.

Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. A survey of large language models. *arXiv preprint arXiv:2303.18223*, 2023.

Kaijie Zhu, Jindong Wang, Jiaheng Zhou, Zichen Wang, Hao Chen, Yidong Wang, Linyi Yang, Wei Ye, Yue Zhang, Neil Zhenqiang Gong, et al. Promptbench: Towards evaluating the robustness of large language models on adversarial prompts. *arXiv e-prints*, pages arXiv–2306, 2023.

Appendix

Ethical Considerations Statement

This study analyses AI policies from ten news organisations and ten universities to identify gaps in the EU AI Act that could be clarified for its implementation and point out possible research directions. All data used in this research is derived from publicly available policy documents, ensuring transparency and compliance with ethical research standards. We strived to present any interpretations or critiques of the policies in a constructive manner to inform policymakers, AI practitioners, and institutional stakeholders.

This study strives to respects intellectual property rights by citing all sources and representing policy content appropriately. Since the analysis pertains to institutional policies rather than individual data, no personally identifiable information is processed or collected.

Finally, we acknowledge that AI governance is an evolving field. To the best of our knowledge, our findings reflect the state of AI policies at the time of analysis and should be interpreted in light of ongoing regulatory and institutional developments. We encourage further interdisciplinary dialogue to refine AI governance frameworks in alignment with ethical, legal, and societal expectations.

Known research challenges not covered in `EU-AIA` or organisational policies

Finally, let us consider the set of sociotechnical challenges that is more broadly known from the existing academic literature (App., Table 6), but that we have not found to be addressed in sufficient detail in either organisational policies we considered or `EU-AIA`:

- ***Definition.*** The organisational policies do not typically define what kind of 'AI' is being addressed, and the definition proposed in `EU-AIA` has been criticized by researchers Hooker [2024].
- ***Enforceability.*** Many policies we considered require **declaration of AI use**, yet there are no robust detection mechanisms to verify compliance for generated texts Puccetti et al. [2024].
- ***Unsafe outputs.*** Only one university in our sample (UiO) mentioned the possibility of exposing students to inappropriate outputs from AI models.
- ***Misleading marketing claims.*** AI providers are constantly advertising new models claimed to be ever better at 'reasoning', 'understanding' and other constructs of questionable validity for the current AI Mitchell [2021]. Many policies we examined seem to be influenced by 'fear of missing out', manufactured by such narratives. More stringent requirements of transparency for claimed benchmark results could alleviate this problem.
- ***Explainability.*** We saw no policies directly addressing the fact that the current AI systems are not interpretable, which has implications for their use (especially where decisions could have significant consequences, e.g. student grading or news fact-checking).
- ***Un-equitable quality of service.*** Some university policies mention the need for AI literacy training, but we did not find that in news, and `EU-AIA` also does not discuss that for the users of AI systems.
- ***Risks to skills.*** It is possible that over-reliance on AI systems could damage basic competences or creativity of its users, but most policies we examined do not seriously consider this factor.
- ***Risks to workforce skills.*** Only 3 universities and no news organisations considered this point, and it is not addressed in `EU-AIA` beside documentation.

Code	Description	News	Edu
Suggested Uses			
Illustrations/Graphics	Using AI to generate illustrations and graphics for publication.	✓	
Image generation	Creating photorealistic visuals, diagrams, or representations.	✓	✓
Article generation	Generating long texts for publication.	✓	
Data analysis	Processing and analysing large datasets.	✓	
Language tool	Grammar checks, proofreading, and rephrasing.	✓	✓
Transcription/Translation	Transcribing or translating content.	✓	✓
Ideas (Content)	Brainstorming story ideas, headlines, sources.	✓	
Ideas (Marketing)	Generating ideas for marketing campaigns.	✓	
Ideas (University)	Brainstorming project/curriculum planning.		✓
Content moderation	Filtering spam, hate speech, fake news.	✓	
Self-learning	Supporting independent learning with adaptive resources.		✓
Course design	Drafting rubrics, assignments, teaching materials.		✓
Personalisation	Tailoring content to learner needs/preferences.		✓
Coding	Debugging, code snippets, concept clarification.		✓
Topic knowledge	Explaining concepts, simulating discussion, suggesting readings.		✓
Feedback	Giving feedback on essays, projects, code.		✓
Summarisation	Condensing academic texts or lectures.		✓
Assessment	Assisting with assignment evaluation.		✓
Search	Guiding research or literature review.		✓
Issued Warning and Rules			
Human oversight	Require human review of AI output.	✓	✓
Declaration of use	Disclosure of AI tool use.	✓	✓
Document AI use	Internal tracking of AI experiments.	✓	
Factual accuracy	Encourage fact-checking of outputs.	✓	✓
Bias in AI	Mitigating biased or discriminatory content.	✓	✓
Privacy and sensitive data	Discouraging input of private or sensitive data.	✓	✓
Copyright	Awareness of content and input IP risks.	✓	✓
AI literacy training	Educating users on ethical/effective AI use.	✓	✓
No prioritisation	Preventing biased prioritisation of outputs.		✓
Knowledge cut-off	Warning about outdated training data.		✓
Persuasiveness	Warning about convincing but incorrect output.		✓
Source attribution	Lack of citations and plagiarism risk.		✓
Digital inequity	Unequal access and capabilities across users.		✓
Skills assessment	AI use may obscure student ability.		✓
Environment	Awareness of AI's environmental cost.		✓
Teacher load	Increased burden for educators monitoring AI.		✓

Table 3: Codes used for annotating the policies of news organisations (News) and universities (Edu). ✓indicates that the code was present in that domain.

Name	Abbreviation	Country
The Guardian	Guardian	UK
ANP	ANP	Netherlands
Mediahuis	Mediahuis	Belgium / Netherlands
Norwegian Tabloid VG	VG	Norway
Le Parisien	Parisien	France
Financial Times	FT	UK
Süddeutsche Zeitung	SZ	Germany
Der Spiegel	Spiegel	Germany
Ringier	Ringier	Switzerland
BBC	BBC	UK

Table 4: Selected news organisations.

Name	Abbreviation	Country
Technical University of Munich	TUM	Germany
Delft University of Technology	TU Delft	Netherlands
KTH Royal Institute of Technology	KTH	Sweden
Aalto University	Aalto	Finland
Technical University of Denmark	DTU	Denmark
KU Leuven	KUL	Belgium
Swiss Federal Institute of Technology Zurich	ETH	Switzerland
Charles University	CUNI	Czech Republic
University of Vienna	Vie	Austria
University of Lisbon	UdL	Portugal
University of Oslo	UiO	Norway

Table 5: Selected universities.

Challenge & Summary	Risk for the org.
What is regulated: what kind of models even fall under the policy? Definitions can be based on training compute European Parliament [2024], data Rogers and Luccioni [2024], performance Anderljung et al. [2023] etc.	Guidelines not scoped appropriately
Detectability & enforceability: can we detect when AI models' usage violates the policy? Particularly, when generated content is used without disclosure? At present, no Puccetti et al. [2024].	Guidelines not enforceable
Factual errors: the current models cannot reliably reject queries for which they do not have enough information Amayuelas et al. [2024], and may output plausible-sounding but false results that are hard to identify and check Zhang et al. [2023], Hicks et al. [2024]. Retrieval-augmented generation still has this problem Mehrotra [2024].	Losing credibility and reputation
Unsafe models: in spite of attempts to force the models to follow certain content policies Ouyang et al. [2022], the models can still violate them Derczynski et al. [2024], and this training can even decrease the quality in some aspects Judge et al. [2024], Casper et al. [2023].	Exposing employees to toxic outputs
Privacy and security risks: employees using non-local generative AI models may expose sensitive data from themselves and their organizations to the entity controlling such models Kim [2023] or third-party attackers Wu et al. [2024]. Platform plugins may also increase vulnerabilities Iqbal et al. [2024].	Exposing sensitive data
Misleading marketing claims: employees may believe the claims of AI model "capabilities" and trust the machine too much Khera et al. [2023], even though the benchmark results may be compromised by methodological problems and test set contamination Rogers and Luccioni [2024].	Degradation in the outputs of the organization
Transparency & accountability: the social and legal norms on disclosing AI "assistance" and taking responsibility for the resulting text have not yet settled. The popular providers of these models do not accept responsibility for any faults in the output[7].	Public blame for any missteps
Bias & inequity: The social biases in AI systems are well-documented Bolukbasi et al. [2016], Nadeem et al. [2021], Marchiori Manerba et al. [2024], Stańczak et al. [2023], Hutchinson et al. [2020], Bender et al. [2021], Sharma et al. [2024], and the use of such models may reinforce misrepresentations in the society.	Discrimination, ethical code violation
Explainability: checking model outputs would be easier if they were accompanied by rationales, the current interpretability methods are not faithful to the model's actual decision-making Lanham et al. [2023], Atanasova et al. [2023].	Trusting unreliable solutions
Brittleness: Generative models perform worse outside of their training distribution McCoy et al. [2024a,b]. For language models, this includes changes in both language (idiolects, dialects, diachronic changes), content (e.g. evolving world knowledge), and slight variations in prompt formulation and examples Zhu et al. [2023], Lu et al. [2022].	Employees wasting time and/or getting poor results
Risks to creativity: AI systems may generate unseen sequences of words, but their "creativity" is combinatorial, often lacking diversity, feasibility, and depth Si et al. [2024], Padmakumar and He [2024], and further degraded in languages other than English Marco et al. [2024]. Exposure to AI assistance could *decrease* human creativity and diversity of ideas in non-assisted tasks Kumar et al. [2024].	Degradation in the outputs of the organization and its existing human resources
Credit & Attribution: AI systems are commonly trained on copyrighted textsGray [2024] without author consent. This practice triggered multiple lawsuits Brittain and Brittain [2023], Vynck [2023], Firm and Butterick [2022], Grynbaum and Mac [2023], Panwar [2025], protests from the creators Heikkilä [2022], Authors Guild Foundation [2025], and questions about the credit for the author of an "assisted" text Formosa et al. [2024].	Legal exposure, violating plagiarism policies/principles
Carbon emissions: The current AI systems are environmentally costly for both training and inference Luccioni and Hernandez-Garcia [2023], Dodge et al. [2022], Bouza et al. [2023], Li et al. [2023], and workflows that significantly rely on them would have adverse effect on climate action.	Not meeting sustainability goals

Table 6: Major sociotechnical challenges for organizations relying on the current AI technology

News Organisation	Policy Link	Version
Guardian	https://www.theguardian.com/help/insideguardian/2023/jun/16/the-guardians-approach-to-generative-ai	June 2023
ANP	https://drive.google.com/file/d/1-3sAJtkOJrdIGw-gZFqYDEGQQNw13e0U/view	April 2023
Mediahuis	https://www.independent.ie/editorial/editorial/aiframeworkguide140623.pdf	May 2023
VG	https://www.vg.no/informasjon/redaksjonelle-avgjorelser/188	April 2023
Parisien	https://www.cbnews.fr/medias/image-groupe-echos-parisien-s-engage-face-intelligence-artificielle-generative-76799	May 2023
FT	https://www.ft.com/content/18337836-7c5f-42bd-a57a-24cdbd06ec51	May 2023
SZ	www.ethz.ch/en/the-eth-zurich/education/ai-in-education.html	Dec. 2024
Spiegel	https://www.sueddeutsche.de/projekte/artikel/kolumne/kuenstliche-intelligenz-ki-e903507/	June 2023
BBC	https://www.bbc.co.uk/supplying/working-with-us/ai-principles/	Feb. 2024
Ringier	https://www.ringier.com/ringier-introduces-clear-guidelines-for-the-use-of-artificial-intelligence/	May 2023

Table 7: Links to AI policies and their versions for each news organisation.

University	Guidelines Link	Version
TUM	www.prolehre.tum.de/fileadmin/w00btq/www/Angebote_Broschueren_Handreichungen/ProLehre_Handreichung_ChatGPT_EN.pdf	Feb. 2023
TU Delft	www.tudelft.nl/teaching-support/educational-advice/assess/guidelines/ai-chatbots-in-unsupervised-assessment	June 2023
KTH	www.kth.se/profile/wouter/page/chatgpt-pragmatic-guidelines-for-students-september-2023	Sep. 2023
Aalto	www.aalto.fi/en/services/guidance-for-the-use-of-artificial-intelligence-in-teaching-and-learning-at-aalto-university	Aug. 2023
DTU	www.dtu.dk/english/newsarchive/2024/01/dtu-opens-up-for-the-use-of-artificial-intelligence-in-teaching	Jan. 2024
KUL	www.kuleuven.be/english/genai	-
ETH	www.ethz.ch/en/the-eth-zurich/education/ai-in-education.html	Dec. 2024
CUNI	www.ai.cuni.cz/AI-12-version1-ai_elearning_en.pdf	June 2023
Vie	www.studieren.univie.ac.at/en/studying-exams/ai-in-studies-and-teaching/	Sep. 2024
UdL	www.conselhopedagogico.tecnico.ulisboa.pt/files/sites/32/ferramentas-de-ai-no-ensino-v8-1.pdf	Nov. 2023
UiO	www.uio.no/english/services/ai/	-

Table 8: Links to AI guidelines and their versions for each university.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly state that the paper examines AI policies from news organizations and universities to identify convergence/divergence patterns and inform AI legislation refinements, which in turn accurately reflects the paper's contributions.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The paper acknowledges limitations, including geographic coverage decisions (one outlet/university per country), the temporal nature of the analysis, and scope limitations in the corresponding Methodology Section descriptions(§3) and Ethical Considerations Statement (Appendix).

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer:[NA]

Justification: This paper is an empirical policy analysis study and does not include theoretical results or proofs.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The paper provides detailed methodology for policy selection (§3), coding procedures (§3), and includes all source links in Table 7 and Table 8, enabling reproduction of the analysis.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in

some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: All analyzed policies are publicly available with URLs provided in Table 7 and Table 8. The coding framework is described in Table 3.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA]

   Justification: The paper does not include computation experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: The paper does not include computation experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Justification: The paper does not include computation experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We follow the NeurIPS Code of Ethics and discuss ethical considerations in the appendix (*Ethical Considerations Statement*).

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We mention societal impact both as policy recommendations as well as point out ethical implications in the appendix (*Ethical Considerations Statement*).

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper analyses existing public policies and does not release models, code, or datasets requiring safeguards.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We link to the policies used for the analysis in this paper.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: The paper does not introduce new datasets, models, or code assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper analyses publicly available policy documents and does not involve human subjects research.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: No human subjects research was conducted.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: The core methodology involves manual analysis of policy documents through iterative coding.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.