Talking Past Each Other? Navigating discourse on Ethical AI: Comparing the discourse on ethical AI policy by Big Tech companies and EU policymakers

Comparing the discourse on ethical AI policy by Big Tech companies and EU policymakers

Cornelia, Evers
Independent Researcher
co.evers@outlook.de

ABSTRACT

This study examines the convergence of the European Commission (EC) and Big Tech companies' (Google and Microsoft) discourse on 'ethical' AI through critical discourse analysis and the concept of hegemonic discourse. The paper answers the question to what extent there is a hegemonic discourse on ethical AI between EU policymakers and Big Tech companies and whether this is impacted by the prospect of legally binding legislation, considering the possible impact of the 2021 AI Act Proposal of the European Commission. This analysis is relevant at an inflection point where previous literature notes superficial convergence between the approaches of public and private actors, indicating policy consensus. The scope of analysis however is limited to non-legally binding regulation and lacks regional focus. In the EU, the advent of legally binding AI regulation with the 2021 AI Act (AIA) Proposal marks a critical juncture: with agreement on the AIA in December 2023, ethics standards become part of market entry requirements to the EU Single Market and the underlying differences in approaching Ethical AI will have important ramifications on policy preferences, compliance, enforcement and thought leadership in the domain more broadly. I find that the European discourse on 'ethical' AI by the EC and Big Tech companies such as Google and Microsoft is largely hegemonic and depoliticised in non-legally binding settings from 2018-2021 due to shared assumptions on 'ethical' AI and absence of significant underlying social and political conflict. It evolves to non-hegemonic and repoliticised discourse through dislocation by the prospect of legally binding regulation, which pushes actors to reveal their genuine policy preferences that bear political and social conflictuality whilst both actor types take an instrumental approach to ethics.

CCS CONCEPTS

• **Science and Technology Studies**; • **Critical Algorithmic Studies**; • **AI and Ethics**;

KEYWORDS

AI Act, Ethical AI, Fairness, Transparency, STS, Socio-technical systems, European Commission, Big Tech, Discourse Analysis, Hegemonic discourse, depoliticization, repoliticisation

1 INTRODUCTION

The reproduction and exacerbation of social inequalities through AI systems put AI ethics on policy and academic agendas in the mid-2010s, making non-legally binding 'ethical' AI principles and governance structures proliferate. Literature notes superficial convergence between the approaches of public and private actors, indicating policy consensus. The scope of analysis is limited to non-legally binding regulation and lacks regional focus. In the EU, the advent of legally binding AI regulation with the 2021 AI Act (AIA) Proposal marks a critical juncture: with the official adoption of the AIA in December 2023, ethics standards become part of market entry requirements to the EU Single Market [1].

To understand whether the initial AIA proposal may have engendered policy conflicts or whether supposed convergence on 'ethical' AI between public and private policy actors prevails, this author examines the convergence of the European Commission (EC) and Big Tech companies' discourse on 'ethical' AI using Gramsci's concept of hegemonic discourse. This paper answers the question to what extent there is a hegemonic discourse on ethical AI between EU policymakers and Big Tech companies and whether it is impacted by the prospect of legally binding legislation. I find that the European discourse on 'ethical' AI by the EC and Big Tech companies such as Google and Microsoft is largely hegemonic and depoliticised in non-legally binding settings from 2018-2021 due to shared assumptions on 'ethical' AI and absence of significant underlying social and political conflict. It evolves to non-hegemonic and repoliticised discourse through dislocation by the prospect of legally binding regulation, which pushes actors to reveal their genuine policy preferences that bear political and social conflictuality. Previously minimised social and political conflicts on 'ethical' AI likely re-emerge to dominate the ongoing debates on 'ethical' AI governance in the EU. To reach this conclusion, I present the

emergence of AI ethics and the distinction between intrinsic and instrumental approaches to ethics, leading me to qualify the EC's and the Big Tech's approach to ethics as instrumentalised by broader economic growth or business development motives. The author links this literature to the supposed convergence on 'ethical' AI principles to reveal to what extent policymakers and companies share 'ethical' AI policy preferences. The theoretical framework introduces the concept of hegemonic discourse and Remling's concept of depoliticisation to better approach this question [2]. I use the European Commission and Big Tech Companies (Google and Microsoft) as case studies for Critical Discourse analysis of primary sources on their attitudes to both non-legally binding and legally binding 'ethical' AI governance within EU jurisdiction by examining their regulatory model preferences and their 'ethical' AI principles through the examples of transparency and fairness.

2 LITERATURE REVIEW

Ethical AI emerges in the mid-2010s amidst growing scrutinisation of AI's negative social consequences. Big Tech companies are considered as contributors to unequal power structures rather than drivers of economic growth. This shift of public mood towards Big Tech companies is called 'techlash' [3].

Emerging literature primarily focuses on AI-based discrimination caused by historically biased data, prejudiced risk variables and human bias. Public discourse highlights discriminatory outcomes in fields such as policing, social welfare services or employment [4, 5]. In reaction to this techlash, the contested concept of 'ethical' AI emerges. It comprises corporate, public policy, scholarly and civil society initiatives grappling to 'align' AI with fundamental values and ethics [6]. Tech ethics carry different and often conflicting connotations. They can be approached from a moral philosophy and justice perspective, but equally as a performative proclamation of public sector or corporate values [6]. Bietti distinguishes between intrinsic and instrumental approaches. Intrinsic ethics describes a justice-seeking process with independent moral value [7]. Instrumental approaches however value ethics for the results it produces, as a means for achieving a superior goal [7]. An organisation's ethical AI definition hence reflects their understanding of ethics. This paper adopts UNESCO's ethical AI definition, which connects "ethical values and principles to the challenges and opportunities linked to AI technologies, built on "international and national legal frameworks, human rights and fundamental freedoms" [8].

2.1 Supposed convergence in the debate on 'ethical' AI

Since 2018, Big Tech companies develop internal 'ethical' AI governance through principles, ethics boards, company codes of practice and academic expertise [9].

CEOs like Google's Sundar Pichai regularly call for legally binding regulation, which shall not stifle innovation [10]. The EU organises public consultations, creates the AI Watch research group and convenes the High-Level Expert Group on AI for elaborating the Ethical Guidelines on Trustworthy AI [11]. The EC's regulatory intention materialises with the 2020 Whitepaper on AI that culminates in the legally binding AI Act proposal. With increasing academic attention for AI ethics and non-legally binding 'ethical'

AI principles across private and public institutions, a discourse on supposed convergence on ethical AI emerges. Jobin, Ienaca and Veyena's seminal meta study reviewing 84 'ethical' AI principles declarations finds global convergence around the five principles of transparency, fairness, non-maleficence, responsibility, and privacy [12]. Hagendorff equally finds that 80% of guidelines converge around the principles of accountability, privacy, and fairness [13]. Nevertheless, both authors note that the principles' definitions may highly vary across actors. Most principles are approached through tech-solutionism, leading to systematic omissions of AI's underlying social and political embedding [13]. This is facilitated by their abstract nature and their limited effect on corporate or public governance [14]. Hence, it seems like 'ethical' AI principle convergence is limited to surface level wording but does not extend to diverging definitions and practices. This raises questions about the extent to which the broadly adopted discourse on 'ethical' AI convergence might prevent further scrutinisation of AI's place in society. Currently, this literature does not trace how convergence and differences in the discourse manifest across broader policy discourse and primary sources, notably at the prospect of legally binding regulation.

2.2 Contested AI Ethics

Beyond this debate on surface-level converging non-legally binding 'ethical' AI, there seems to be academic convergence on qualifying 'ethical' AI governance as an 'instrumental' approach to ethics. It is used by private and public actors as a means for achieving broader economic, or regulatory goals [7].

Whilst these goals may be concurring or mutually exclusive, they are often combined. Economic instrumentalisation of AI ethics describes the commodification of the latter for maintaining contentious business models despite their potential material risks to fundamental rights or for using it as an enabler of economic growth. Big Tech companies hence ethicise the negative social and political consequences of their operations to legitimise existing business models as adequate for respecting fundamental rights [15]. Aligning business models with AI ethics results in the adoption of narrow business and reputation risk perspectives, which minimise underlying social and political challenges [16]. Similarly, the EU takes an economically instrumental approach by embedding AI ethics into an economic and geopolitical AI race narrative against the United States and China, which mobilises economic resources for its uptake [17]. AI ethics are framed as a means for prevailing in the economic and geopolitical race whilst maintaining the EU's commitments to fundamental rights under Art.2 of the Treaty on the EU. This commodification of AI ethics facilitates the omission of underlying social and political tensions from 'ethical' AI discourse, which is sustained by 'tech-solutionist' approaches to AI ethics by public and private actors. Ethical concerns are framed in narrow technical terms, presenting them as 'fixable' through research and product optimisation. Assuming that complex social and political challenges can be solved with technological fixes is called tech-solutionism [18]. Its consequence is that only a small circle of technical experts should engage with 'ethical' AI issues [19]. In the studied context, this paper adopts the assumption that AI ethics are tech- solutionist and economically instrumentalised.

Talking Past Each Other? Navigating discourse on Ethical AI

FAccT '24, June 03–06, 2024, Rio de Janeiro, Brazil

This framing is pervasive in North American STS and critical algorithmic literature studies. However, these concepts are generally underexplored in computer science and legal scholarship.

3 THEORETICAL FRAMEWORK

This research focuses on the EU and Big Tech companies operating under its jurisdiction. Existing literature finds convergence on 'ethical' AI principles to be superficial due to the presence of diverging definitions and assumptions on AI. Mittelstadt and Hagendorff demonstrate how underlying differences in interpreting core 'ethical' AI principles buries social and political conflict and avoids it from having to be vocalised in the debate on 'ethical' AI [13, 14].

Nevertheless, they do not consider instrumental approaches to AI ethics which is necessary to understand how social and political conflict is channelled through a common framework of underlying assumptions. Assessing the convergence of the debate in consideration of the presented limitation may bear important results for future policy development and conflict. Particularly, the turn towards legally binding regulation reveals an important gap in the study of 'ethical' AI governance: There is no assessment on how the supposedly convergent discourse on 'ethical' AI governance evolves over time and how it may change upon the prospect of legally binding regulation. As the latter puts constraints on AI market actors, it is expected that previously minimised divergences in the superficially convergent discourse may become more prevalent, through increased conflictuality motivated by the desire of reaching the most favourable legal framework. Consequently, I focus on re-assesing the presence of convergent non-legally binding 'ethical' AI discourse in the EU. This is the basis for understanding its evolution in light of legally binding regulation (AIA Proposal) and whether it impacts on the extent to which social and political conflict is expressed in the policy debate. Consequently, I ask the following research question: To what extent is there hegemonic discourse on ethical AI between EU policymakers and Big Tech companies and is it impacted by the prospect of legally binding legislation?

Convergence is approached through the concept of hegemonic discourse as defined by Gramsci [20]. This concept allows to account for discourse effects on power structures and the struggles of the actors to dominate 'ethical' AI governance. Hegemonic discourse describes a cohesive narrative that through its "power force wins the discursive marketplace" [21]. Through its dominating position in a given societal context, it constructs a distinct "common sense" understanding of phenomena, often supported by shared underlying myths and assumptions [22]. This paper considers hegemonic discourse on a spectrum and not as an absolute binary to account for discourse nuances. In the realms of digital development, myths are used to "construct and promote digital developments, communication policy and legitimate modes of governance that would not have been possible without the establishment of such a discourse" [23]. Consequently, myths as components of hegemonic discourse are crucial for determining how a society embraces AI. Furthermore, hegemonic discourse limits the possibility of discussing alternative conceptions of a phenomenon like Big Tech infrastructure power, which fits the literature on convergence yet

is not jointly discussed [23]. By sustaining a narrative within the limits of predominant hegemonic discourse, dislocations, that is disruptive elements to the dominating narrative are subsumed into it, which serves to maintain existing social, political, and economic power structures [2]. Dislocation describes how hegemonic discourse can be challenged through new events or emerging elements it can no longer explain or justify, hence discontinuing or shifting the hegemonic discourse [24]. This could depoliticise the discourse on a certain phenomenon. Remling's analysis of the EC's discourse on climate adaption, reveals how the EC constructs the phenomenon of climate change as "serious enough to warrant attention but without consequences that meaningfully change the status quo" [2]. I apply depoliticisation to the assessment of whether the non-legally binding policy discourse in the EU is hegemonic. I hypothesise that the discourse on 'ethical' AI and its non-legally binding policies by the EC and Big Tech companies form a hegemonic discourse that depoliticises underlying social and political questions. Following this logic, the AIA Proposal would be a dislocation to the hegemonic discourse and its hypothesised depoliticised nature, if it alters the manner in which social and political conflict is presented or omitted from the discourse. Whilst the AIA proposal is a product of the EC's continuous approach to ethical AI and is not a dislocation in itself, I assume it to be a dislocation to the hegemonic discourse of supposed convergence on 'ethical AI. It is expected that the AIA proposal forces actors to confront underlying social and political conflicts through the pressure of binding regulation, hence repoliticising the debate. Should this dislocation be confirmed, it will have wide reaching implications for future policy conflict. It is important to note that this paper will not attempt to prove that the AIA alone may produce this effect as other independent variables cannot be controlled for. It however may generate a hypothesis that can be scrutinised through further research to understand in how far policy debates may be dislocated by the advent of legally binding regulation.

4 METHODOLOGY

As per the case selection, I study EU jurisdiction due to its seminal policy initiatives on AI and its self- proclaimed leadership role in 'ethical' AI governance [25].

The EC is selected for analysis, as it has the initiative of legal acts and engages with AI ethics as a regulator since 2017. The publication of the EC's 2019 'Ethics Guidelines for Trustworthy AI' increased demands for regulatory intervention amidst public attention on discriminatory AI use across the EU [26, 27]. Following its 2020 Whitepaper on AI, marking the EC's legally binding regulatory intention, it is the first international organisation to develop legally binding legislation on AI with its 2021 AI Act Proposal [1]. Its legally binding nature forces Big Tech companies to closely engage with the EC through public consultations which provide important data to assess whether the AIA proposal dislocates the supposedly hegemonic discourse on 'ethical' AI. For comparison, I select 'Big Tech' companies, defined as the "five most prosperous and influential US technological companies of the global IT industry" [28]. They dominate the EU Tech market with combined global yearly revenues of more than 1,5 trillion US dollars in 2022 [29]. Moreover, they hold monopolistic market power by concentrating

global AI resources like cloud computing infrastructure, translating into bargaining power for shaping discourse on AI [30]. This market position increases the likelihood of competitors aligning with their 'ethical' AI approaches, making them the most relevant actors in comparison to the EC. Through purposive sampling, I select Microsoft and Google as 'Big Tech' case studies. Both are market leaders in the EU jurisdiction and intensely engage with EC policymakers on 'ethical' AI since the publication of the 'Ethics Guidelines for Trustworthy AI' [31]. They recently contributed to public consultations on the EC's Whitepaper on AI and the AIA proposal. Both companies undertake internal 'responsible AI' efforts by publishing Responsible AI Principles' and internal governance such as Microsoft's "Responsible AI Practices" and Google's regularly published "Policy Perspectives" [32, 33]. The main limit of this case selection is the small sample size, limiting the scope for making broad generalisations about Big Tech's approach to ethical AI.

The method of Critical Discourse Analysis is employed to analyse the use of language as a social practice and its effect on social phenomena across the discourse of the EC, Google, and Microsoft [34]. For assessing to what extent, the discourse of the actors on 'ethical' AI governance is hegemonic, I use the concept of congruence that I define through both lexical and semantic congruence. Lexical congruence refers to the degree of converging wording on 'ethical' AI that I operationalise through the concept of nomination. It consists of understanding how 'ethical' AI phenomena are defined and referred to and what aspirations this expresses in the construction of a policy strategy [35]. 'Ethical' AI nominations hence bear important implications for assessing semantic congruence. Semantic congruence, exploring whether the meanings of terms and their underlying assumptions align, is operationalised through the concepts of argumentation and perspectivization [35]. Argumentation is the analysis of arguments for understanding their use in persuading an audience of normative rightness in claims and narratives [35]. It allows to scrutinise the actors' argumentation strategies for justifying their 'ethical' AI governance approaches. Perspectivisation allows to understand from what perspective the studied nominations and arguments are expressed [35]. This requires acknowledging Big Tech's corporate perspective, driven by their business models and the EC's position as the EU's legislative regulator, pursuing concurring economic- and rights-based goals. The limitations of CDA for this analysis are the potential selection biases of the author in choosing to analyse certain discourse elements over others and the lack of peer-review regarding the identification of discourse themes. CDA is based on publicly available primary sources published by the actors between 2017 and June 2023. This timeframe is selected as it reflects the initial emergence of non-legally binding policy debates across the studied actors until the end of data collection in June 2023. For CDA on non-legally binding regulation, I refer to the EC's policy documents published since 2017 such as the 2019 Ethics Guidelines for Trustworthy AI, until the 2020 Whitepaper on AI which marks the turn for the assessment of discourse on legally binding regulation. The main document for assessing the EC's attitudes to legally binding regulation is the 2021 AI Act Proposal. Regarding Google and Microsoft, I consider their 'responsible' AI principles, policy documents, corporate blogs and videos discussing internal and external governance preferences for

'ethical' AI as a primary discourse source on non-legally binding regulation. The core documents for assessing attitudes to legally binding regulation are the submissions to the public consultations on the 2020 Whitepaper on AI and the 2021 AIA proposal. A detailed list of primary sources can be found in appendix 1. Using inductive theme identification, I initially familiarise myself with the texts, identify relevant passages about 'ethical' AI governance, and then re-evaluate and group retained text elements into categories and subcategories listed in appendix 2 [36]. Discourses on 'Ethical' AI Principles proclaimed by all studied actors and the AI Act were treated independently to identify distinct themes. Later, the findings were reintegrated to understand the implications of theme overlaps, particularly regarding nuances between discourse on non-legally binding and legally binding governance approaches.

5 EMPIRICAL RESULTS AND DISCUSSION

5.1 Hegemonic Discourse on non-legally binding regulation

This section demonstrates the actors' hegemonic approach to non-legally binding 'ethical' AI regulation. I find high lexical congruence in regulatory modes and abstract principles, yet minor semantic congruences. These results are due to varying degrees of policy preference specification and differences between rights-based and corporate approaches to ethics.

The EC considers non-legally binding 'ethical' AI regulation as a transitory form of regulation which precedes legally binding regulation. Its non-legally binding guidance is intended to prepare legally binding regulation, as the 2020 Whitepaper clearly underscores. Google and Microsoft however consider non-legally binding regulation to be sufficient in itself. Google highlights why stakeholder expertise and broader flexibility of self- and co-regulatory regimes is better than "static" regulation, a negatively connoted word for legally-binding regulation as it underlies the assumption that legislation is inherently rigid and incapable of evolution [37]. This opposes the Big Tech vision of fast-moving innovation and hence subtly reveals potential regulatory conflict. Interestingly, Google and Microsoft consistently frame themselves as governance experts and announced their respective 'responsible' AI principles and governance structures in 2018. Since, they engage in diverse research and development (R&D) and policy initiatives to operationalise them. Additionally, they engage with European institutions, academia, and other stakeholders to embed their vision of 'ethical' AI governance into large-scale regulation. Overall, there is high lexical congruence between the EC and Big Tech companies embracing non- legally binding governance, yet the vision of self-regulatory regimes semantically differs: whilst the EC sees it as a transition to legally binding regulation, the companies consider it as a sufficient governance mode which may raise policy conflict upon the prospect of legally binding regulation.

To understand the extent to which these lexically congruent approaches to non-legally binding regulation translate to hegemonic discourse on the principles and practices associated with 'ethical' AI, I scrutinise each actors' 'ethical' AI principles and supporting documents for qualitative comparison. My findings confirm those of Jobin, Ienca and Vayena and Hagendorff who find high lexical

Table 1: Comparative overview of 'ethical' AI principles of studied actors

Broad Category	Nomination (EC)	Nomination Google	Nomination Microsoft
Safety	Technical Robustness and safety	Safety	Reliability and safety
Privacy/Data Protection	Privacy and Data Governance	Privacy	Privacy and Security
Accountability	Accountability	Accountability	Accountability
Fairness	Diversity, Non-Discrimination and Fairness	Fairness	Fairness and Inclusion
Transparency	Transparency	N/A, grouped with other categories	Transparency
Social Mission	Societal and Environmental Wellbeing	Social Benefit	N/A, mention in other policy documents

congruence across proclaimed 'ethical' AI principles but incongruences in their semantic meaning [12, 13]. My comparison reveals similar nominations of high-level abstract principles for approaching 'ethical' AI governance, indicating a high degree of lexical congruence. Table 1 shows which nominations the actors choose for the principles identified in Jobin et al.'s meta study. Table 1 confirms their results as all categories are present in our sample. A noteworthy incongruence is that Google does not proclaim a transparency principle. It is included in the accountability principle and further primary sources, allowing the conclusion that the principle is nevertheless prioritised.

To assess to what degree these congruent lexical nominations extend to semantic congruence on the meaning of these 'ethical' AI principles, I focus on the principles of transparency and fairness. They are selected for analysis as they are central to each actor's 'ethical' AI governance approach, persist in their discourse on the AIA proposal and have the least overlap with other related policy debates. The principles of privacy or safety for instance are interlinked with other policy debates such as the EU General Data Protection Regulation (GDPR) or the EU Cybersecurity Act, which would necessitate to include these regulatory contexts more closely for a comprehensive analysis.

5.1.1 Transparency. All three actors allude to the principle of transparency. The EC is the most specific and prescriptive about what transparency means for 'ethical' AI governance and how it should be applied. Microsoft and Google remain more abstract, creating an impression of superficial discourse congruence.

In its 'Ethics Guidelines for Trustworthy AI', the EC presents three different concepts associated with transparency: first, they introduce the principle of data explicability, the AI system and its business model [38]. This understanding is linked to explainable AI (XAI) computer science literature, which focuses on how computing processes can be transparently explained to professionals and lay people [39]. Traditionally, XAI is limited to explaining computing processes and data treatment. The EC's extension of this notion to business models demonstrates its demand of opening Big Tech business models to broader scrutiny. Closely linked to the notion

of XAI and the principle of explicability is the notion of traceability the EC introduces: "The data sets and the processes that yield the AI system's decision, including those of data gathering and data labelling as well as the algorithms used, should be documented to the best possible standard to allow for traceability and an increase in transparency" [38]. Traceability refers to increasing explicability and transparency through the possibility of chronologically following algorithmic computing processes. These two transparency concepts benefit specialists who can utilise the additional information for conducting assessments of AI systems to determine their compatibility with ethics requirements. Regarding citizen-oriented transparency, the EC proposes that human users shall be informed if they interact with an AI system [38]. This demonstrates the EC's emphasis on human autonomy which is linked to the fundamental rights-based element of its discourse on 'ethical' AI. Microsoft does not provide governance recommendations but focuses on providing a definition of transparency and what goals it should achieve. Transparency is presented as a dual concept, consisting of operator's transparency about "how and why they are using AI, and also the limitations of their systems" [33]. Secondly, they emphasise that "transparency also means that people should be able to understand the behaviour of AI systems" [33]. This approach matches the EC's conception of XAI yet is presented in an abstract fashion less specific about implementation. Microsoft is more vocal on the goals transparency can fulfil by referring to mitigating unfairness in AI systems, helping to technologically correct them and gain user trust. Overall, Microsoft's approach is largely semantically congruent with the EC's approach, which may be caused by the abstract nature of its statements. Google is the least vocal about transparency and does not consider it amongst its 'responsible' AI principles. Whilst issues pertaining to transparency are discussed in other primary sources such as the yearly operationalisation reports on 'responsible' AI, it is amongst the least discussed principles. Commitments on transparency pertain to providing relevant explanations and appeal mechanisms for Google's AI product operations [40]. Overall, there is medium level semantic congruence on the actors' conception of transparency with highly varying degrees

of engaging with the principle. This could lead to the impression of general discourse congruence for a lack of more details about how particularly Big Tech companies conceive the principle to be implemented.

5.1.2 Fairness. Regarding fairness, I find highly congruent, almost exclusive, focus on bias in AI systems and tech-solutionist fixes to mitigate them. Upon inspection, the motivation for approaching bias slightly differs between the EC's rights-based approach and the Big Tech's corporate risk approach. This may impact on the operationalisation of fairness and raise policy conflicts as the discourse evolves to legally binding governance.

Following the EC, bias is caused by: "Inadvertent historic bias, incomplete data or 'bad' governance and describes its consequences as unintended (in)direct prejudice and discrimination against certain groups or people, potentially exacerbating prejudice, and marginalisation. Harm can also result from the intentional exploitation of (consumer) biases or by engaging in unfair competition, such as the homogenisation of prices by means of collusion or a non-transparent market" [38]. This definition encompasses human and technical sources of bias, and consequences, namely marginalisation and discrimination. Strikingly, it extends to market mechanisms such as unfair competition or price discrimination, demonstrating a genuine rights-based approach. This is absent from Big Tech definitions as it would undercut their business models. Google adopts a congruent definition of bias yet approaches it through a narrow engineering efficiency and corporate risk lens that omits market mechanisms. It contends that: "AI algorithms and datasets can reflect, reinforce, or reduce unfair biases. We will seek to avoid unjust impacts on people, particularly those related to sensitive characteristics such as race, ethnicity, gender, nationality, income, sexual orientation, ability, and political or religious belief" [40]. It acknowledges technical bias causes yet redirects discourse to the possibility that algorithms could reduce bias. Whilst absent from the EC's perception, it is present in Microsoft's discourse that considers fairness as a means for "making sure that the systems we develop and deploy, reduce unfairness in our society rather than keep things at the same level or make it even worse" [33]. This reframing of the narrative implies that AI is the solution to a problem it creates. Such relativisation of AI's negative risks may be part of a broader strategy to avoid regulatory binding legislation by reframing its ethical risks as societal opportunities.

Moreover, both Big Tech companies demonstrate their instrumental and corporate risk logic to AI ethics by framing fairness initiatives as matters of engineering efficiency. Google incorporates fairness "early in developers' machine learning workflow and throughout the product development lifecycle" as this would "avoid burning engineering cycles spent retrofitting technology if an issue emerges after launch or even much later. (...) This aligns with our product excellence mantra" [40]. This instrumentalised approach to fairness as a means to achieving product excellence and minimising efficiency losses demonstrates a highly corporate approach to AI ethics. Both Google and Microsoft acknowledge that 'unethical' products could lead to reputational and financial risks such as adversely affecting "revenues and operating results" [16]. Consequently, fairness serves the purpose of preserving business models and revenue. This approach results in companies

presenting mitigation practices that "the organisation is not yet doing but is capable of doing" through "relatively costless reforms that provide the veneer of ethical behaviour" [6]. My findings demonstrate that this facilitates a narrow vision of how complex principles such as fairness should be mitigated, mainly through a limited tech-solutionist approach. This allows companies to address fairness concerns through "relatively costless" technological reforms with existing resources without having to modify their underlying business models. Interestingly, the EC equally presents mitigation measures in mainly technical terms. The EC takes an abstract tech-solutionist approach to mitigating fairness by advocating for "identifiable and discriminatory bias to be removed in the collection phase", demonstrating the assumption that bias is primarily a technical problem that can be technically removed [38]. Computer science literature clarifies that removing characteristics from AI models is insufficient for reducing bias, in absence of a systemic scrutinisation of the dataset [5]. This approach omits how other computing processes such as statistical bias, faulty or overly precise inferences or shifting the sampling frame can equally be the source of discriminatory outcomes. The EC hence takes a reductive approach to technical bias mitigation. Google equally takes a fully tech-solutionist approach to addressing fairness. Particularly its yearly 'responsible' AI operationalisation reports highlight examples of fairness concerns the company identifies and how these are addressed. One noteworthy example is persisting gender bias in Google Translate that is addressed through dataset adaptation [40]. Beyond this example, Google scrutinises the implications of a wide product range, a level of transparency that Microsoft does not match. Microsoft is silent about mitigation techniques it takes and merely refers to its internal governance structures and how these implement such techniques [33]. Overall, the actors adopt a semantically congruent vision of fairness through the lens of AI bias and tech-solutionism, however with incongruences regarding the acknowledgement of market mechanisms and business models as bias sources, which may lead to policy conflict in a legally binding context.

5.1.3 Conclusion: Largely hegemonic discourse on non-legally binding regulation. In conclusion, I find high level lexical congruence in the nomination of non-legally binding 'ethical' AI principles. Moreover, I find largely semantically congruent discourse on the analysed principles of transparency and fairness. Despite apparent incongruences, I consider that the argumentation strategies of abstractness and unexplored contestation foster a largely hegemonic discourse on non-legally binding 'ethical' AI governance.

First, the discourse on transparency demonstrates how congruence is constructed through the contrast between specific policy propositions by the EC and abstract statements by Google and Microsoft. This abstractness makes it impossible to discern whether Big Tech companies agree with the EC's position. Therefore, the discourse of regulator and regulatee seems largely congruent yet omits and invisibilises potential incongruences that could manifest as policy conflicts. Secondly, the principles of transparency and fairness reveal incongruences between the EC's rights-based logic and the Big Tech's corporate risk-based lens. This conflicts with the EC's mention of market mechanisms and business models

Talking Past Each Other? Navigating discourse on Ethical AI

FAccT '24, June 03–06, 2024, Rio de Janeiro, Brazil

as producers of unfair AI bias and potentially untransparent systems. As these incongruences do not extend beyond a sentence of recommendations each yet would bear wide reaching implications for how to approach 'ethical' AI governance and the political economy of AI, they cannot be considered as consistent and frequent enough to undermine the overall perception of congruent discourse. This is reinforced as the EC does not stress such systemic economic considerations in other policy documents and because its approach to 'trustworthy' AI is grounded on said business models for the economic growth it expects to derive from AI development. Furthermore, the highly congruent tech-solutionist approach to mitigating ethics concerns across all three actors limits the scope for considering such deeper-seated political contestation. In conclusion, I qualify the discourse of all three actors on 'ethical' AI governance as hegemonic discourse, due to the lexical congruence on ethical principles and relatively high semantic congruence on how these principles are operationalised, despite the discussed mechanisms and limitations. This hegemonic discourse has the effect of reinforcing "a narrow and confined vision of the possibilities of regulatory change, and inhibit dialogue", particularly on underlying issues of structural injustice, the political economy of AI and the effect of its business models on ethics [23]. Instead, it centres the attention on "procedural fairness and fixable tweaks" [7]. As typical for hegemonic discourse, valuable alternatives are less likely to emerge [7]. In the context of 'ethical' AI governance, the hegemonic discourse contributes to the invisibilisation of underlying social and political contestation and conflict. In analogy to Remling's qualification of the EC's hegemonic discourse on climate adaptation as 'depoliticised' due to the absence of underlying social and political issues from the discourse, I consider the hegemonic discourse on 'ethical' AI governance as depoliticised. This raises the question to what extent the prospect of the legally binding AIA may present a dislocation to this hegemonic discourse that may 're-politicise' the discourse on 'ethical' AI by making political conflict resurface? As each three actors have an interest in aligning the AIA as closely as possible with their policy preferences, it is expected that the minor incongruences in the non-legally binding discourse will intensify to reveal the contested character of 'ethical' AI governance and potentially call the hegemonic discourse into question.

5.2 Dislocation of hegemonic discourse through the prospect of legally binding regulation

To examine whether the prospect of legally binding regulation has an effect on the largely hegemonic discourse on non- legally binding 'ethical' AI governance, I assess the EC and Big Tech's attitude to legally binding regulation in the context of the 2021 AIA proposal.

This analysis reveals to what extent previously dissimulated underlying social and political questions may be repoliticised through the prospect of legally binding regulation. The 2021 AIA proposal is the first horizontal regulation on AI, encompassing all AI applications across the Single Market [41]. This study focuses on the 2021 EC proposal and not the final version at the moment of adoption in December 2023. It is a market entry regulation aimed at setting "harmonised rules on the development, placing on the Union market and the use of products and services making use of AI

technologies or provided as stand-alone AI systems" [1]. It equally aims to ensure all AI systems on the Single Market comply with the EU Charter of Fundamental rights [1]. The proposal takes a technologically neutral and risk-based regulatory approach. AI systems are classified according to the risk they pose to fundamental rights and accordingly need to fulfil differentiated obligations for entry to the Single Market. AI applications posing an unacceptable risk are prohibited and contain practices like social scoring. High-risk AI applications need to fulfil a set of requirements to gain access to the EU Single Market [1]. The majority of provisions in the AIA proposal concern obligations for such high-risk systems. Limited and minimal risk applications are solely subject to minor transparency obligations. Prior to its publication, all studied actors expressed themselves on their legally binding regulation preferences. These approaches are crucial to understand in how far the AIA proposal represents a dislocation to the previously hegemonic discourse.

5.2.1 Regulatory preferences for legally binding regulation. There is strong lexical and semantic congruence on a risk-based approach to AI regulation. However, there are important incongruences on underlying regulatory models between prescriptive approaches favoured by the EC and performance- based approaches favoured by Google and Microsoft.

These incongruences have important implications for the scope in which policy conflicts about the AIA proposal are discussed and reveals that matching each actor's perspectivisation, Big Tech companies favour regulation models that give them maximum discretion in implementing legally-binding 'ethical' AI governance, whereas the EC follows a public sector control rationale. Everyone favours a risk- based approach to regulation, meaning that the regulatory burden upon an AI system is proportional to the risk it poses to fundamental rights and safety. The EC adopts this vision starting with its 2020 Whitepaper on AI and the Big Tech companies since their public consultation submissions on the latter, indicating that the EC may have contributed to this policy preference. However, there are two important semantic incongruences regarding the scope of such regulation: Whilst the EC embraces a horizontal approach in the AIA and its previous communication, both Big Tech companies express themselves in favour of sectoral regulation: This vision of regulation is more fragmented and grants Big Tech companies more discretion in co-determining the rules that apply to their sector. Second, there are important incongruences between the EC and the Big Tech's preferred regulatory model: whilst the EC favours prescriptive regulation, based on management and technology standards, both Big Tech companies are in favour of performance-based standards. Prescriptive regulation consists of legal provisions that precisely detail how regulatory goals are to be achieved by regulates [42]. In regulatory literature, management- and technology-based regulation describe this prescriptive approach of the EC. The EC largely deploys this regulatory approach by prescribing how providers of AI systems shall govern the data underlying their systems, specificities of their datasets and pre- and post-market requirements that specify both process and outcome elements [43]. Contrarily, Big Tech companies favour performance- based regulation, focused on assessing regulatees on their capacity to reach certain regulatory targets without prescribing processes on how to reach them [43]. This may leave ethically

contentious processes out of regulatory sight [43, 44]. Google and Microsoft may prefer this option as it implies a limited regulatory burden compared to prescriptive approaches. Altogether, the afore-mentioned lexical and semantic incongruences on the approaches to legally binding regulation between the EC and Big Tech companies reveal highly diverging preferences of regulatory mode that will shape the subsequent analysis of whether the AIA proposal is a dislocation to the actors' hegemonic discourse on non-legally binding 'ethical' AI governance.

The EC, Google and Microsoft evolve in their discourse on 'ethical' AI governance upon the prospect of legally binding regulation since the publication of the 2020 Whitepaper on AI. Increased specificity in the Big Tech's expression of policy preferences on transparency repoliticises the previously hegemonic discourse by making political conflicts and concurrent rights more prevalent. Moreover, the previously hegemonic tech-solutionist discourse on fairness is dislocated through open policy conflict on the scope and operationalisation of fairness mitigation.

5.2.2 Transparency. The AIA proposal leads to a specification on how the EC envisions the principle of transparency in continuity of explicability and traceability it established in its non-legally binding 'ethical guidelines for trustworthy AI'.

Article 64 of the AIA proposal on 'access to data and documentation' grants market surveillance authorities the possibility to require "full access to the training, validation and testing datasets used by the provider, including through application programming interfaces ('API') or other appropriate technical means and tools enabling remote access" [1]. In the second paragraph, this right extends to access to the source code. These provisions enable transparency on the technical components of AI systems for market surveillance authorities which helps address the underlying information asymmetry between Big Tech companies and regulators. This approach is consistent with the notion of explicability and mitigates the information surplus of companies vis-à-vis regulators as they have more information about the operations of their systems. Incomplete information makes effective AI regulation more challenging for regulators and prolongs the delay with which they can react to emerging technological developments [44]. The EC hence remains consistent in its discourse and raises a tension with Big Tech companies by openly requiring access to their data and the systems with which their systems operate. Consequently, Google and Microsoft both see themselves forced to diverge from their previously abstract and unspecific discourse on transparency to express their contestation. Both companies do not deem the access to their datasets and source codes as adequate and cite security, data protection, trade secret and technological reasons to justify their preference. This creates policy conflict by forcing concurring ethics elements such as transparency and data protection to be weighed against each other. Google for instance contends: "In addition, there are serious privacy concerns regarding any sharing of datasets that contain personally identifiable information (PII). This includes the heightened exposure risk inherent in data being retained when it would otherwise have been deleted in line with GDPR's data minimisation principles, and the danger that the data transfer mechanisms enabling it are exploited by malicious actors" [45]. Beyond raising concerns about the proposal's compatibility

with the EU's General Data Protection Directive, this focus on the divulgation of PII reveals the contested nature of how transparency should be envisioned. Whilst one may argue that Google and Microsoft use data protection concerns as a pretext to avoid regulatory scrutiny of their business models and data, their reasoning nevertheless reveals underlying political tension that conflicts with the widely hegemonic discourse on transparency in non-legally binding settings.

5.2.3 Fairness. The AIA proposal, and particularly Article 10 on data governance approaches fairness and non-discrimination through a narrow focus on bias introduced to AI systems through training, validation, and testing data [1].

Particularly the third paragraph with requirements for data sets raises contestation amongst Google and Microsoft. It contends that: "Training, validation, and testing data sets shall be relevant, representative, free of errors and complete. They shall have the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons on which the high-risk AI system is intended to be used" [1]. Imposing the properties of data sets is prescriptive yet lacking specific definitions of what terms such as "relevant, representative, free of errors and complete" mean, creates legal uncertainty, and provides companies discretion in choosing processes for attaining these technological goals. It is hence a hybrid management-based regulatory approach. Moreover, by approaching fairness and non-discrimination through a narrow *ex- ante* focus on technical bias mitigation presents an incongruence with the EC's previous non-legally binding approach which is more holistic, considering fairness and non-discrimination through a broader socio-technical lens, including human bias, design choices and institutional choice and governance structures [38]. Even though the proposal contains a provision obliging providers to consider "specific geographical, behavioural or functional setting within which the high- risk AI system is intended to be used", this does not establish a clear link with bias and the potential need of adapting data sources and management practices according to the deployment context [1]. For instance, the requirement of data set 'representativeness' legitimises the assumption that bias is a result of unrepresentative data which obfuscates the multiple causes of AI unfairness and discrimination necessary to consider for a holistic approach. The example of AI in predictive policing tools shows that social and organisational bias amongst the police or political practices such as urban redlining equally facilitate the reinforcement of discrimination patterns for marginalised communities [4].

Overall, confining solutions to the ambit of technological fixes risks obscuring the complex socio-technical AI lifecycle and deployment contexts. Google and Microsoft voice their disagreement with this policy and cite both technical and socio-technical reasons to express their conflicting visions on bias mitigation. These underly the incongruent approach to the regulatory model by criticising the prescriptiveness of the EC's approach and advocating for performance-based standards. Regarding technical feasibility, Google says that "certain requirements of the regulation such as that datasets be 'free of errors and complete', demands a level of perfection that is not technically feasible" [45]. Microsoft similarly proclaims that "this requirement will be unworkable in some scenarios and inadequate in others" [46]. They explain that it is

Talking Past Each Other? Navigating discourse on Ethical AI

FAccT '24, June 03–06, 2024, Rio de Janeiro, Brazil

impossible to create a 'complete' dataset, as missing data points or errors are not fully removable due to undetectable self-reporting mistakes [46]. Moreover, Microsoft importantly recognises that "present provisions alone will not allow to fulfil the regulatory goal of reducing unfair and discriminatory outcomes produced by AI systems without broader consideration of industry practices and awareness for the highly contextual nature of fairness and the fact that it is never possible to fully "de-bias" an AI system or "guarantee" its fairness". Ensuring that training, testing, and validation datasets are appropriately tailored to the task at hand is only part of the solution to promoting AI fairness and non-discrimination" [46]. This statement denounces the EC's narrow focus on technological bias and the disregard for assessing its contextual deployment. They stress that "model design and continuous real-world testing throughout the deployment lifecycle are equally important" and overall assume that "Article 10 — or the AI Act as a whole cannot provide the level of fairness and non-discrimination in AI systems to which it aspires" [46]. By declaring the EC's approach to fairness as incompatible with fulfilling its goal, and its scope too limited, it reveals an underlying tension between tech-solutionist and socio-technical solutions. Whilst Microsoft remains within the confines of tech- solutionism by mainly focusing on technological mitigation measures to address fairness, its position reveals an underlying political tension present in the broader academic literature between STS and corporate governance approaches. In the given context, Microsoft's concerns seem to be less congruent with the spirit of STS literature and the notion of AI as socio-technical systems but rather motivated by shifting the focus to a less prescriptive regulatory model, focused on performance-based regulation. The company says that "outcomes-based approach to requirements is more likely to achieve the AI Act's goals" and presents a list of alternative performance metrics [46]. One such metric is that "high risk AI systems should provide a similar quality of service for relevant demographic groups impacted by the system" [46]. By recentring the focus on material consequences of AI systems rather than process specifications, Microsoft creates an argumentative structure to justify their regulatory preference that would give them more discretion and liberty in how to attain the regulatory goal. Google similarly advocates for performance-based governance and insists on the inclusion of 'industry best practices' as the appropriate process for achieving the desired regulatory outcome [45]. This emphasis on performance standards and industry best practices is a means for Big Tech companies to define the meaning and processes with which the regulatory goals are to be attained, granting them maximum control over regulation in their position as regulates.

5.2.4 Conclusion: Repoliticisation of discourse on 'ethical' AI through dislocation. Overall, the largely hegemonic discourse on non-legally binding 'ethical' AI governance is dislocated by the prospect of the legally binding AIA proposal through repoliticisation. Repoliticisation means that previously omitted yet underlying social and political conflicts become apparent in the actors' discourse that formulate specific policy preferences in the competition for the most favourable legally binding regulatory outcome.

Beyond the highly incongruent regulatory model preferences, I identify two mechanisms that drive the dislocation of the largely hegemonic discourse on 'ethical' AI governance: the shift from abstract and vague statements to specific policy preferences and open policy conflict within the highly hegemonic tech-solutionist approach to fairness and non-discrimination. Concerning regulatory model preferences, the discourse on legally binding regulation is largely incongruent between the EC and Big Tech companies.

Whilst all actors agree on a risk-based approach, there are wide-reaching incongruences between the EC's preference for horizontal and prescriptive regulation and the Big Tech's sectoral and performance-based regulation models. Initially, all actors shared a largely congruent discourse on principle-based non-legally binding regulatory regimes, with the only difference that the EC regarded such self-regulation as transitionary whereas Big Tech companies considered it as a stand-alone option. I hypothesise that in absence of a concrete prospect of legally binding regulation, this divergence did not lead to policy conflict as underlying regulatory model preferences were congruent in the time period from 2018 until 2020. The latter year marks a discourse turning point with the EC's Whitepaper on AI. It formulates its legally binding regulatory intention and its preference for risk-based, horizontal and prescriptive governance. This initiates the dislocation of the largely hegemonic approach to regulation as Big Tech companies present their incongruent preferences in response. They surprisingly persist in their highly incongruent discourse after the presentation of the AIA proposal, which sustains open policy conflict and hence repoliticises the debate about what regulatory model is best fit for 'ethical' AI governance.

Second, Big Tech companies that remained abstract in their discourse on transparency during non-legally binding regulation, become more specific upon the prospect on legally-binding regulation. Whilst this trend is not surprising, given that previously their vision is less decisive for broader governance, it contributes to repoliticising the discourse by making contentions questions and policy conflict more present in the discourse. Such conflicts manifest through confronting concurring fundamental rights, for instance tensions between maintaining data protection and enabling transparency. The move from abstract to specific policy preferences also means that conflicting regulatory goals of broader digital regulation are addressed: The EC's effort to reduce information asymmetry between Big Tech companies and the public sector by demanding access to company data underlying their business model raises relevant questions about information power and the growing concentration of digital infrastructure and data in the hands of highly monopolistic Tech companies [9]. The open policy conflict about such a provision that exemplifies underlying political tensions extending well beyond AI regulation only emerges through the presentation of the AIA proposal. Finally, hegemonic discourse on AI governance is dislocated through the repoliticisation of the hegemonic tech-solutionist lens to fairness and non-discrimination. Whilst the narrow focus on tech-solutionism persists in the context of the AIA proposal, the EC interestingly evolves to an even narrower focus on data bias, instead of the previously vaster socio-technical and market-centred approach to bias in AI systems. The narrow focus on data bias and its tech-solutionist operationalisation through data set requirements leads to open policy conflict with Google and Microsoft. The latter refer to technological infeasibility and a reductive vision of bias and fairness to shift the debate to industry led standards and processes. This leads to the emergence of two

distinct policy conflicts: The diverging operational understanding of a complex phenomenon like fairness, and second a persisting conflict on whether prescriptive or performance-based regulatory models are more adapted for 'ethical' AI governance.

6 CONCLUSION

Overall, this paper finds that the European discourse on 'ethical' AI by the European Commission and Big Tech companies such as Google and Microsoft is largely hegemonic and depoliticised in non-legally binding settings from 2018-2021, yet evolves to non-hegemonic and repoliticised discourse upon being dislocated by the prospect of non-legally binding regulation, which pushes actors to reveal their genuine policy preferences [1].

Whilst the author cannot confirm that the AIA proposal is the independent variable producing this change, the results allow for this hypothesis that should be further scrutinised in subsequent research. The initial assessment of non-legally binding regulation reveals that both the EC as a lawmaker and Big Tech companies as the most relevant private actors in the AI field take a highly lexically congruent approach to their 'ethical' AI principles, by identifying largely similar principles as relevant. Their definitions and operationalisation however reveal minor semantic incongruences, through the mechanisms of abstractness and unexplored contestation. Nevertheless, these incongruences are minor and the discourse on non-legally binding regulation is regarded as hegemonic, which is further sustained by the actors' shared instrumental approach to AI ethics. This hegemonic discourse has the effect of reinforcing a narrow policy scope which through the absence of underlying social and political issues from the discourse leads to depoliticisation of 'ethical' AI. This depoliticised hegemonic discourse on non-legally binding 'ethical' AI governance is dislocated by the prospect of the legally binding AIA proposal through the repoliticisation of the previously hegemonic discourse. Repoliticisation means that previously omitted yet underlying social and political conflicts become apparent in the discourse of the actors that formulate specific policy preferences in the competition to achieve the most favourable legally binding regulatory outcome. Beyond the highly incongruent regulatory model preferences, I identify two mechanisms that drive discourse dislocation: the shift from abstract and vague statements to specific policy preferences and open policy conflict within the highly hegemonic tech-solutionist approach to fairness and non-discrimination.

These findings are limited by the idiosyncrasy of the selected case and selection biases inherent to CDA methodology. I encountered difficulty discerning Big Tech's genuine policy preferences despite potential desirability bias in their primary sources. This bias might even be an explanatory variable for the initial hegemony of non-legally binding discourse and could be explored in future research. Furthermore, the restricted scope of this research to the initial 2021 AIA proposal which does not consider subsequent versions as well as the final version of the AIA adopted in December 2023 is a limit of this paper. Subsequent research could trace how the policy conflict shown in this paper is resolved or not by tracing the actors' discourse throughout the trialogue process and the final AIA. Moreover, it may be fruitful to trace the discourse of other

relevant actors such as the European Parliament and show nuanced results on the mechanisms of hegemony.

Despite these limitations, this paper successfully joins the concept of hegemonic discourse through the novel concept of lexical and semantic congruence to the literature on supposedly converging 'ethical' AI governance by considering shared underlying mechanisms such as an economically motivated instrumental approach to AI ethics. At a critical juncture in AI policy moving towards legally binding regulation, this paper identifies ensuing policy conflicts and their place within a narrow 'ethical' AI policy context focused on tech-solutionist mechanisms. The move from supposed hegemony to politicised discourse indicates upcoming policy conflict and the likelihood of previously ignored social and political tensions at the centre of digital governance. The repoliticisation of the discourse on 'ethical' AI within the EU following the introduction of the AI Act may have important implications for its ensuing codification and implementation across the EU. The results of this paper point towards conflictuality, which will likely be observed in the drafting and agreement about the Conformity Assessments under the AI Act throughout the transition period prior to full application by 2027. These assessments for high-risk AI systems (Article 3, AI Act) are to be conducted either internally or by accredited third parties to determine what regulatory burden AI providers will fall under [1]. Currently, European institutions are consulting with industry actors and other stakeholders on how to standardise and manage such assessments, which posits a ground on which the rising repolitcisation of the discourse on 'ethical' AI between Big Tech companies and the European Commission may continue. Companies will have an interest to influence the procedures and standards that determine risk classification and hence the regulatory burden falling on them. Their priorities may differ from the EC, which needs to balance its commitment to public safety with procedural decisions regarding its administrative structure that impact on implementatibility. A new arena for these tensions may be the planned EU AI Office that will mainly be tasked with enforcing the AI Act. Such persisting conflictuality may further entrench the repoliticisation of 'ethical AI' in the public discourse as the question of what understandings of ethics will prevail in such a standardisation process, coupled with the management practices it enables, will be decisive for the effective enforcement of the AI Act. As such development unfold, it will be fruitful to further trace the discourse of the studied actors in subsequent research.

ACKNOWLEDGMENTS

REFERENCES

[1] European Commission. 2021. Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=$celex%3A52021PC0206
[2] Remling E. 2018. Depoliticizing adaptation: a critical analysis of EU climate adaptation policy. Environmental Politics. 27 (3). 477-497. https://doi.org/10.1080/09644016.2018.1429207
[3] Phan, T., Goldenfein, J., Mann, M., & Kuch, D. 2021. Economies of Virtue: The Circulation of "Ethics" in Big Tech. Science as Culture, 1–15. https://doi.org/10.1080/09505431.2021.1990875
[4] O'Neil, C. 2016. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Google Books.https://books.google.be/books/about/Weapons_of_Math_Destruction.html?id=$NgEwCwAAQBAJ&redir_esc$=$y

[5] Eubanks, V. 2018. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor Macmillan. https://us.macmillan.com/books/9781250074317/automatinginequality

[6] Green, B. 2021. The Contestation of Tech Ethics: A Sociotechnical Approach to Technology Ethics in Practice. Journal of Social Computing, 2(3), 209–225.https://doi.org/10.23919/jsc.2021.0018

[7] Bietti, E. 2020. From ethics washing to ethics bashing. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. https://doi.org/10.1145/3351095.3372860

[8] UNESCO. 2021. Recommendation on the ethics of artificial intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000380455

[9] Hu. L. 2021. Tech Ethics: Speaking Ethics to Power, or Power Speaking Ethics? Journal of Social Computing, 2(3), 238–248. https://doi.org/10.23919/jsc.2021.0033

[10] Pichai, S. 2023. Google CEO: Building AI responsibly is the only race that really matters. Financial Times. https://www.ft.com/content/8be1a975-e5e0-417d-af51-78af17ef4b79

[11] Harasimiuk, D., Braun, T. 2021. Regulating Artificial Intelligence: Binary Ethics and the Law. Taylor & Francis.

[12] Jobin, A., Ienca, M., & Vayena, E. 2019. The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

[13] Hagendorff, T. 2020. The Ethics of AI Ethics: An Evaluation of Guidelines. Minds and Machines, 30(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8

[14] Mittelstadt, B. 2019. Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4

[15] Hedlund, M. 2023. Ethicisation and Reliance on Ethics Expertise. Res Publica. https://doi.org/10.1007/s11158-023-09592-5

[16] Metcalf, J., Moss, E., & Boyd, D. 2019. Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics. Social Research: An International Quarterly, 86(2), 449–476. https://par.nsf.gov/servlets/purl/10112016

[17] Ulnicane, I. 2022. Artificial Intelligence in the European Union. Library.oapen.org; Taylor & Francis. https://library.oapen.org/handle/20.500.12657/52622

[18] Morozov, E. 2013. To Save Everything, Click Here: The Folly of Technological Solutionism. https://www.hachettebookgroup.com/titles/evgeny-morozov/to-save-everything-click-here/9781610393706/?lens$=$publicaffairs

[19] Greene, D., Hoffmann, A., & Stark, L. 2019. Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning.https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/849782a6-06bf-4ce8-9144-a93de4455d1c/content

[20] Gramsci, A. 1971. Selections from the Prison Notebooks. Italica, 73(3), 453. https://doi.org/10.2307/479844

[21] Place, K. R., & Vardeman-Winter, J. 2013. Hegemonic discourse and self-discipline: Exploring Foucault's concept of bio-power among public relations professionals. Public Relations Inquiry, 2(3), 305–325. https://doi.org/10.1177/2046147x13494965

[22] McHale, J.P., Zompetti, J.P., & Moffitt, M.A. 2007. A Hegemonic Model Of Crisis Communication: Truthfulness and Repercussions for Free Speech in Kasky v. Nike. Journal of Business Communication. https://journals.sagepub.com/doi/10.1177/0021943607306137

[23] Brevini, B. 2021. Creating the Technological Saviour: Discourses on AI in Europe and the Legitimation of Super Capitalism. 145–159. https://doi.org/10.16997/book55.i

[24] Höpfner, M. 2011. Hegemonic discourse and change. ECPR. https://ecpr.eu/Events/Event/PaperDetails/9052

[25] Lauchlan, S. 2020. The EU pitches ethical AI regulation and 'a single European data space'. Diginomica.com. https://diginomica.com/eu-pitches-ethical-ai-regulation-and-single-european-data-space-try-not-be-too-underwhelmed

[26] Amnesty International. 2021. Dutch childcare benefit scandal an urgent wake-up call to ban racist algorithms. Amnesty International. https://www.amnesty.org/en/latest/news/2021/10/xenophobic-machines-dutch-child-benefit-scandal

[27] Niklas, J. 2015. Profiling the Unemployed in Poland: Social and Political Implications of Algorithmic Decision Making. Panopticon Foundation. https://www.researchgate.net/publication/320935514_Profiling_the_Unemployed_in_Poland_Social_and_Political_Implications_of_Algorithmic_Decision_Making

[28] Sdia. 2022. Definition for Big Tech. https://sdialliance.org/dictionary/big-tech/#:$\sim$text$=$Other%20definitions,%2C%20Meta%2C%20Amazon%20and%2 0Microsoft.

[29] Poletti, T. 2023. Big Tech hit $1.5 trillion in sales last year, but the days of mega growth seem to be over. MarketWatch. https://www.marketwatch.com/story/big-tech-hit-1-5-trillion-in-sales-last-year-but-the-days-of-mega-growth-seem-to-be-over-b1472a4c

[30] Whittaker, M. 2021. The Steep Cost of Capture. Papers.ssrn.com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id$=$4135581

[31] Amaro, S. 2023. Amazon, Apple and more: EU names the Big Tech platforms set to face tougher regulatory scrutiny. CNBC. https://www.cnbc.com/2023/04/25/amazon-apple-and-more-eu-names-companies-set-to-face-tougher-regulatory-scrutiny.html

[32] Google. 2021. Consultation on the EU AI Act Proposal Google's submission. European Commission. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/F2662492_en

[33] Microsoft. n.d. Responsible AI principles from Microsoft. https://www.microsoft.com/en-us/ai/responsible-ai

[34] Fairclough, N. (1993). Critical discourse analysis and the marketization of public discourse: the universities. Discourse & Society, 4(2), 133–168. https://www.jstor.org/stable/42888773

[35] Reisigl, M. (2017). The Discourse-Historical Approach. Taylor & Francis. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315739342-4/discourse-historical-approach-martin-reisigl

[36] Mogashoa, T. (2014). Understanding Critical Discourse Analysis in Qualitative Research.ResearchGate.https://www.researchgate.net/publication/264215347_Understanding_Critical_Discourse_Analysis_in_Qualitative_Research

[37] Google. 2020. Building a responsive regulatory framework for AI. https://ai.google/static/documents/building-a-responsible-regulatory-framework-for-ai.pdf

[38] European Commission. 2019. Ethics Guidelines for Trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[39] Ankarstad, N. 2020. What is Explainable AI (XAI)? An introduction to building trust in machine learning through global and local variable importance using post-hoc explanations. Medium. https://towardsdatascience.com/what-is-explainable-ai-xai-afc56938d513

[40] Google. 2022. 2022 AI Principles Progress Update. https://ai.google/static/documents/ai-principles-2022-progress-update.pdf

[41] European Parliament.2023. Artificial Intelligence act. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

[42] Freeman, J., & Kolstad, C. D. 2006. Prescriptive Environmental Regulations versus Market-Based Incentives. Moving to Markets in Environmental Regulation, 3 16. https://doi.org/10.1093/acprof:oso/9780195189650.003.0001

[43] Coglianese, C., & Starobin, S. 2020. Management-Based Regulation. Policy Instruments in Environmental Law. https://scholarship.law.upenn.edu/faculty_scholarship/2226/

[44] Yeung, K. (2017). Algorithmic Regulation: A Critical Interrogation. Papers.ssrn.com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id$=$2972505

[45] Google. 2021. Consultation on the EU AI Act Proposal Google's submission. European Commission. https://ec.europa.eu/info/law/better-regulation/have- en

[46] Microsoft. 2021. Microsoft's Response to the European Commission's Consultation on the Artificial Intelligence Act. https://blogs.microsoft.com/wp-content/uploads/prod/sites/73/2021/09/microsoft-response-to-the-european-commission-consultation-on-the-artifical-intelligence-act.pdf</bib>

A APPENDICES

A.1 PRIMARY SOURCES CDA LISTED PER ACTOR

A.1.1 European Commission (Non-legally binding primary sources)

- Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of The Regions (2018)
- Ethics Guidelines for Trustworthy AI (2019)
- Communication: Building Trust in Human Centric Artificial Intelligence (2019)
- White Paper on Artificial Intelligence: A European approach to excellence and trust (2020)
- Ethics by design and Ethics of Use Approaches for Artificial Intelligence (2021)
- Coordinated Plan on Artificial Intelligence 2021 Review (2021)

A.1.2 European Commission (Legally binding primary source)

- Proposal for a regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts (2021)

A.1.3 Google (Non-legally binding primary sources)
- Responsible AI Practices (n.d.)
- Perspectives on Issues on AI Governance (2019)
- Responsible Development of AI (2019)
- Building a responsive regulatory framework for AI (2020)
- Marian Croak's vision for responsible AI at Google (2021)
- 2022 AI Principles Progress Update (2022)
- Google Research, 2022 & beyond: Responsible AI (2023)
- Google CEO: Building AI responsibly is the only race that really matters (2023)
- A policy agenda for responsible AI progress: Opportunity, Responsibility, Security (2023)
- A Policy Agenda for Responsible Progress in Artificial Intelligence (2023)

A.1.4 Google (Sources on legally binding texts)

- Consultation on the AI Inception Impact Assessment Google's submission (2020)
- Consultation on the EU AI Act Proposal Google's submission (2021)

A.1.5 Microsoft (Non-legally binding primary sources)
- A Balancing Act: Regulating AI to boost responsible innovation in Europe (2021)
- Governing AI: A Blueprint for the Future (2023)
- Microsoft #Tech Talk on Responsible AI (2023)

A.1.5 Microsoft (Sources on legally binding texts)
- Response of Microsoft Corporation to the European Commission's Inception Impact Assessment on Artificial Intelligence Legislation (2020)
- Microsoft's Response to the European Commission's Consultation on the Artificial Intelligence Act (2021)

A.2 THEMES IDENTIFIED THROUGH CDA

Table 2: Non-legally binding regulation

Theme	Presence amongst actors	Details
Nature of AI: Dual Nature framing of AI	Present in the discourse of all three actors	Refers to dual Nature framing of AI Refers to dual use nature, risk as inherent to innovation
Attitudes to governance	Present in the discourse of all three actors	AI Principles (comparison), not stifling innovation, policy alignment
Principle of transparency	Commitment by all actors to varying degrees	Explainability (XAI), traceability, black box problem
Principle of fairness	Commitment by all actors to varying degrees	Human and technical bias, technical mitigation strategies

Table 3: Legally binding regulation

Theme	Presence amongst actors	Details
Regulatory preferences	Present in the discourse of all three actors	Whitepaper on AI (2020) and AIA Act (2021) with their respective public consultations
Principle of transparency	Commitment by all actors to varying degrees	Article 10 AIA Proposal, transparency on business model and source codes
Principle of fairness	Commitment by all actors to varying degrees	Article 64 AIA Proposal, representative, complete and error free datasets