tripleC 22 (1): 307-328, 2024
http://www.triple-c.at

Tracing Class and Capitalism in Critical AI Research

Petter Ericson*, Roel Dobbe** and Simon Lindgren***

*Department of Computing Science, Umeå University, Umeå, Sweden, pettter@cs.umu.se, https://www.umu.se/personal/petter-ericson

**Department of Technology, Policy and Management, Delft University of Technology, Delft, The Netherlands, r.i.j.dobbe@tudelft.nl, https://bit.ly/roel-tudelft

***DIGSUM, Department of Sociology, Umeå University, Umeå, Sweden, simon.lindgren@umu.se, https://simonlindgren.com

Abstract: This article explores the rapidly developing field of Critical AI Studies and its relation to issues of class and capitalism through a hybrid approach based on distant reading of a newly collected corpus of 300 full-text scientific articles, the creation of which is itself a first attempt at properly delineating the field. We find that words related to issues of class are predominantly but not exclusively confined to a set of studies that make up their own distinct subfield of Critical AI Studies, in contrast to, e.g., issues of race and gender, which are more broadly present in the corpus.

Keywords: artificial intelligence, machine learning, digital capitalism, research, critical studies

1. Introduction

Artificial intelligence (AI), referring to machine learning, large language models, image generators, and assorted emerging and long-existing computational and algorithmic systems, is a term currently used by self-proclaimed futurologists and marketers alike, who have great success in building unjustified hype around these digital products and services. AI, as currently constituted, is inherently tied to digital capitalism, with both its technology and the data it runs on functioning primarily as commodities to be bought and sold. As the research area of Critical AI Studies is rapidly developing, and as notions of what it means to be 'critical' may vary, this article investigates to what degree the topics of class and capitalism do indeed come to the fore in this developing field.

Even if we define AI more conservatively, many ties to digital capitalism remain, notably between AI and automation and the mechanisation of the workforce. There are, for example, many attempts to directly replace or supplant labour with AI in a growing number of jobs, but AI technologies are also used for increased regimentation of the workforce through algorithmic management and AI tools that 'optimise' labour, e.g., in Amazon warehouses and the so-called gig economy (Jones 2021, Delfanti 2021, Ongweso Jr. 2021).

By inferring future performance and categorisations from past data, AI also strengthens existing societal power relations, reifying them and embedding them in naturalised form in technical infrastructures where they may be even harder to counter than in their social form. AI solutions may exacerbate and further polarise existing social divisions by relying on training and benchmarking on discretised data, demanding and proliferating sharp and 'objective' distinctions between different categories (Birhane et al. 2021).

Date of Acceptance: 12 February 2024
Date of Publication: 26 April 2024

Perhaps more tangibly, the consolidation of AI-related research and development among a small number of primarily US-based tech companies provides clear evidence that AI forms a next frontier to further centralise power and wealth accumulation through computational infrastructures, often reifying existing racist and white supremacist ideas and practises to breed new forms of global digital colonialism and capitalism(Whittaker 2021, Kak and Myers West 2023, Birhane et al. 2021, Couldry and Mejias 2020).

With this tight entanglement between capitalism, conservatism, and the current iteration of AI, it is perhaps no surprise that several strands of distinctly progressive AI-critical research are being pursued. In particular, AI has garnered a large amount of regulatory scrutiny, with many countries pursuing legislation to curb the excessive use of AI and to try to limit the harms and risks that may come from its deployment. Additionally, attempts to de-bias, correct, and otherwise counter AI's inherently conservative and structure-reaffirming tendencies are legion at conferences such as AIES and FAccT, as well as the major AI conferences such as AAAI and NeurIPS.

1.1. Research Aim

In light of this far-reaching and complex entanglement of AI and capitalism, we aim in this article to investigate to what degree and how issues of capitalism and class are articulated and positioned in the field of critical AI research. We do this through a systematic analysis of academic publications in this area. Approaches from natural language processing will be leveraged, alongside descriptive statistics, to provide distant readings (Moretti 2013), to map general structures and patterns in how these issues are dealt with in current critical AI research. We will apply these methods to a dataset of research articles sampled from the Scopus database to investigate the role of perspectives on capitalism and class.

Our assessment will draw on established criteria for what constitutes *critical* analysis – “critical ethics; critique of domination and exploitation; dialectical reason; ideology critique; critique of the political economy; struggles and political practice” (Fuchs 2022, 20). The goal is to provide an empirically grounded classification of how capitalism and class are acknowledged or ignored in the research.

Being a kind of meta-study, this article is partly influenced by the paper “The Values Encoded in Machine Learning Research” (Birhane et al. 2021), wherein the authors examined biases in machine learning research through an analysis of 100 much-cited papers from leading machine learning conferences. A key finding was that only a small minority of the papers (15 percent) linked their work to societal needs, and that only one percent discussed potential negative effects. The authors found values such as performance, generalisation, quantification, and efficiency were at the centre, leading to a centralisation of power. Furthermore, they found notable affiliations between the papers, major tech companies, and elite universities. In this paper, we devise a similar critique but with a somewhat different approach. Our study is not focused on mainstream and high-profile AI-research papers but on papers sampled and extracted based on their affiliation with a ‘critical’ perspective (see further the section on “Dataset Creation”). In other words, our study aims to contribute to a more comprehensive and diverse analysis of biases in AI research by focusing on perspectives explicitly positioning themselves as critique. Furthermore, our study leverages the combination of computation and interpretation offered by a distant reading approach.

2. Method

2.1. Dataset Creation

As stated above, this study's goal is to analyse to what degree and how issues related to capitalism and class become articulated and positioned within critical AI research. Initially then, there is a need for a conception of what critical AI research entails. It is only based on such a definition that we might be able to construct a reasonably representative set of publications to analyse. To achieve an operationalisation, we designed a search string for use with the Scopus indexing service through iterative experimentation. While Scopus is not exhaustive, it covers a significant segment of academic literature, which is why we chose it to construct our dataset. In the field of bibliometric research, Scopus is one of the premier databases, as it has broad coverage across a wide variety of disciplines (Falagas et al. 2008, Mongeon and Paul-Hus 2016), is frequently updated (Mingers and Lipitakis 2010), and has robust export features (Meho and Yang 2007).

As the field that we are trying to delineate is well underway to becoming known under the moniker of "critical AI studies" (Roberge and Castelle 2021, Lindgren 2023b, Jones 2023), we included that as one of the key terms in our query. But importantly, to capture articles that conceptually match with the field but do not necessarily use that specific term, we searched for a set of terms (for example 'critical theory', 'marx*', 'racis*', 'capitalis*') in conjunction with AI terms. The query used was the following:

> TITLE-ABS-KEY ((("social justice" OR "queer" OR "critical theory" OR "marx*" OR "feminis*" OR "decolonial" OR "*racis*" OR "*fascis*" OR "*capitalis*") AND ("artificial intelligen*" OR "data scien*" OR "machine learning")) OR "critical AI studies")

Briefly, then, we queried for articles that, in either title, abstract, or keyword, contain either the name of the research field itself ("critical AI studies") or a combination of two sets of key concepts: first, something pertaining to the "critical" part, i.e. some subject of critical study such as various structural biases and systems of oppression, or a critical approach, such as feminism, decoloniality, or similar; and then something pertaining to "AI": AI, machine learning or data science. Deploying this search query yielded 1212 articles in the Scopus index on June 22nd, 2023.

It should be noted that our sampling strategy inherently involves a certain element of potential circularity. This is because the terms used to delineate and extract articles in Critical AI Studies inevitably shape the results yielded. However, this is also a deliberate choice aligned with our research aims. Our study does not seek to determine *if* specific themes, such as class, capitalism, social justice, feminism, and decolonialism, occur within the literature on AI. Rather, it aims to analyse 'to what degree' and 'how' the particular topics of class and capitalism are discussed and positioned in relation to others. This involves examining their proportionality and positioning within the broader area of Critical AI Studies. It is essential to include these terms in our search strategy to capture a relatively broad spectrum of critical scholarshipon AI. This approach has allowed us to collect a dataset that hopefully reflects a certain degree of variety and depth of critical perspectives in the field. So, for the purposes of the search string, terms like 'class', 'capitalism', 'social justice', 'feminism', and 'decolonialism' are not just keywords but foundational concepts that we see and assume as having shaped the discourse of Critical AI Studies. Moreover, it should be noted that while the search string looks only in the title, abstract, and keywords of the articles, the study of the use

of these terms in the following uses the fulltext of the article to interrogate the use of these terms more deeply.

A manual inspection of this initial set of articles revealed, in fact, that it was overly broad and inclusive, containing many articles which would not normally be considered part of the field of Critical AI Studies. Thus, we carried out a second manual filtering step, wherein we applied a set of relatively simple and clear criteria to the articles' available metadata (title, abstract, and keywords). In particular, our interest is in articles that deal critically with the *practice* of AI, both inside and outside of academia. To minimise recency or authorship bias on the part of the author who did most of this filtering (Ericson), the order of articles was randomised before filtering. The criteria used were the following:

Criteria for inclusion:

- Articles with "critical AI studies" included in either the author or the index keywords were assumed to be in the field.
- Articles where the title or abstract strongly indicated a critical engagement with the practice of AI were included.
- Articles that strongly referenced theories and writers within the critical tradition, such as Marx, Foucault, or Deleuze and Guattari, were included.
- Articles where the title or abstract strongly indicated that they related to the impact of AI on society were included.

Criteria for exclusion:

- Articles where the title or abstract indicated that the references to "AI" were largely tangential, or as one of a set of buzzwords, or where one of the search terms had proved to hit an unintended synonym[1] were excluded.
- Articles where the title or abstract strongly indicated that the "AI" was a *tool* used in the published research rather than an object of study were excluded.
- Articles where the title or abstract strongly indicated that the study of AI in question was purely technical, implicitly or explicitly excluding any social or socio-technical aspects, were excluded.
- Proceedings abstracts and similar summaries and overview texts were excluded, as the mentioned articles themselves were assumed to be included or excluded in the filtering as appropriate.
- Full books were excluded, as they were more likely to be unavailable in full-text and to reduce the risk of including a particular chapter by itself and as part of an edited volume.

After this filtering step, our 1212 items were segmented into two groups, namely, 380 articles included for further study, and 832 articles excluded. In addition, the groups were further segmented by which criterion warranted their inclusion/exclusion and whether it applied to the keywords, title, or abstract. Due to some articles being inaccessible in fulltext format to the authors for various reasons (paywalls, no digital version, etc.), in the end 309 articles in their fulltext versions were downloaded. Out of these, nine articles turned out to be written primarily in languages other than English.

[1] E.g., "*capitalis*" matched to "capitalise" instead of "capitalism" or "anticapitalist"

Thus, in the end, 300 items were included as part of the final analysis, comprising a total of around 2.6 million words[2].

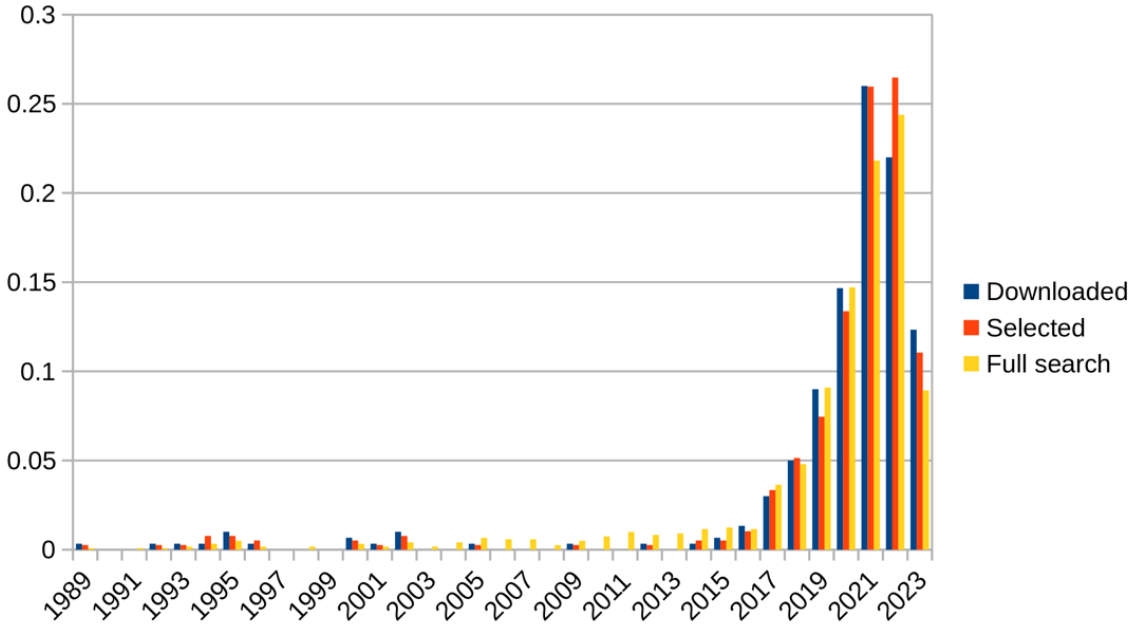

Figure 1: The proportion of items in the set published in each year (1989-2023), for each stage of the filtering: the full search results, the subset selected for further study, and those items available for download in fulltext.

To reveal any particular biases or patterns in the dataset, and in particular, how these patterns changed through the selection stages, we graphed the year of publication (figure 1), as well as the type of publication (figure 2), for the three selections: the full 1212 items of the Scopus search, the 389 articles chosen, and the 309 ultimately downloaded. For all three groups, there is an overwhelming amount of recent items (published in the last five years) in comparison to previous years, which is consistent both with the recent increase in scientific publications (in particular as indexed by Scopus) but more significantly with a recent massive influx in hype and interest in AI in general, and in particular in critical studies of AI. For document types, naturally there were a number of document types that were excluded through our application of the criteria as listed above. When we prepared the full text of these articles for further analysis, some additional exclusions were made. First, we omitted all publications in the “letters”-genre (i.e., shorter, more focused pieces, often with an expedited review process). Second, no items that were mere errata or retractions were included. Among the remaining types, the proportions stayed roughly equivalent through the selection process.

[2] The full list of considered articles is available from the corresponding author by request, but is excluded from this publication due to length constraints

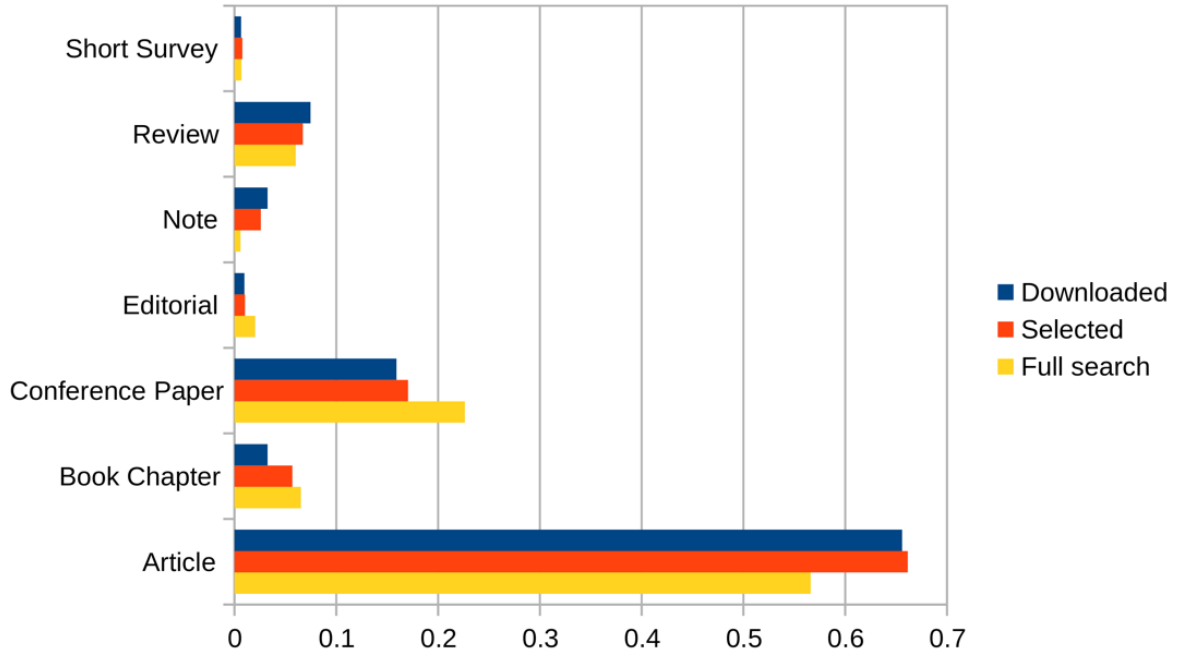

Figure 2: The proportion of items in the set of each type for each stage of the filtering: the full search results, the subset selected for further study, and those items actually available for download in full-text. Item types that only occurred in the full search results have been omitted (circa 5% of all items).

2.2. Analysis Methods

In analysing the corpus of articles, we largely employed a *distant reading* approach to get a broad understanding of the overall themes and patterns present in the texts to form the basis for our assessment of the handling of issues of capitalism and class in critical AI research.

Distant reading, as such, was coined by language scholar Franco Moretti and refers to the practice of analysing large bodies of textual data from a distance without delving into any close or detailed reading of individual texts. Moretti (2005) argues that this is to gain a macroscopic view of the corpus that can help identify overarching patterns, trends, and connections in the texts that may not be identified through close reading alone. While quantitative content analysis of text is a broadly used method in various fields, conceiving the analysis in terms of distant reading adds an additional layer of qualitative interpretation and understanding to the process. As argued by Lindgren (2020), the notion of distant reading allows for a conceptual understanding that sits better with hybrid methods than with purely statistical ones.

In our distant reading, we use a variety of techniques, described in the following sections. Briefly, we begin by constructing document vectors for each article in our dataset using doc2vec, after which we cluster the documents of the dataset to identify tendencies therein and select a subsample of specific interest. We use word2vec on this subsample to characterise the specific word associations present more clearly. Furthermore, we conduct a qualitative study on titles in the subsample and compare the relative use of various terms inside and outside the subsample. This is done to better understand the characteristics of the subsample and what distinguishes it from the rest of the dataset.

2.2.1. Document Vectors

As a first step of the distant reading, we used doc2vec (Le and Mikolov 2014), an unsupervised machine learning algorithm, for representing the full articles as vectors that were clustered using k-means before the model was reduced to 2D using Uniform Manifold Approximation and Projection (UMAP) (McInnes et al. 2020) and visualised as a scatterplot (figure 3). This analytical step helped us identify clusters of similar articles in the corpus, enabling us to see general topical patterns in critical AI research. As we discuss in the results section, we found two clusters (n = 95) to be of importance for our analysis.

2.2.2. Word Vectors

In a second step, we trained a word2vec model on the articles in the two clusters of interest (n=95). Word2vec – the model upon which doc2vec was built – is a method for learning word embeddings, which are vector representations of words in a high-dimensional space (Mikolov et al., 2013).

Word embeddings – conceived as a tool for distant reading – can help us understand the semantic relationships between words based on how they are contextually used within the corpus. Word embeddings, therefore, can be devised as tools for analysing discourses and ideologies. Lindgren (2020, 119-120) draws on the discourse theory of Laclau and Mouffe (1985) and explains how its key concepts can be mapped onto the logic of word embeddings:

> "Starting with the concept of *discourse*, it refers to the general fixation of meaning within a certain domain. So, approaching [the corpus] through word2vec and thereby getting knowledge about how different words cluster together as a consequence of [scholars'] language-use [..] is a means of mapping [..] relations among *elements* (i.e. words), engaging them as moments (i.e. words including their relational positions vis-à-vis other words). The structured totality of relational positions among discursive moments, as described by Laclau and Mouffe takes shape around a set of privileged signs around many other signs are organised. They name such key signs as *nodal points* [..]".

In this study, then, we trained a word2vec model on a subcorpus of articles, sampled as described above, with the intent of reading a 2D visualised view of that model as discursive space where critical scholarly perspectives on AI are articulated around certain 'nodal points' (Figure 4). Given our stated aim, to focus particularly on how issues around class and capitalism are handled and confronted (or not), we started our analysis by approaching the data based on a set of pre-decided terms (namely: *class, classes, capital, capitalism*), which functioned as entry points into the trained model through which related parts of the discourse could be uncovered and disentangled.

2.2.3. Preprocessing and Parameters

We trained the doc2vec and word2vec models using the Gensim library (Rehurek and Sojka 2011). Preprocessing included lowercasing, stop word filtering, removal of numerals and special characters, and a set of custom steps to filter out literature references from the full text. Importantly, we only retained nouns, adjectives, and verbs in the corpus based on which the models were trained. The models were trained at a vector size of 300, meaning that vectors of 300 dimensions represented each document and word in the text. More concretely, this means that each document, and later,

word, in the text, was transformed into a numerical representation consisting of 300 values. In Gensim, the default setting is 100, but setting the number higher can lead to more accurate models. When it comes to how many words before and after the key terms were considered as context, we used Gensim's default 'window size' of 5. This setting is generally considered to be sufficient for extracting syntactic meaning based on the immediate context of words while at the same time counteracting the diluting effect that a window size that is too large may lead to. Levy & Goldberg (2014, 3), in exploring the effects of different hyperparameters for word2vec, argue that "a window size of 5 is commonly used to capture broad topical content", which is also what we strive to do in our study.

2.2.4. Additional Approaches

Aside from the training and subsequent distance readings, as presented in the analysis section of this article, we drew upon some additional steps and measures to deepen and enrich our understanding of the dataset.

First, for the final subsample of 95 articles, we conducted a qualitative review of their titles and abstracts to get a more refined picture of what they were about in terms of what topics they engaged with and which analytical perspectives they favoured. Second, we also employed a direct full-text search of various terms to compare their relative prevalence within and outside the subsample to further nuance our understanding of the differences and similarities between the two sets of documents.

3. Analysis

3.1. Clusters of Articles

The first step of the distant reading, as described above, entailed analysing the patterns by which the articles in the corpus (n=300) are thematically related through the common use of concepts and language. The doc2vec model was clustered using k-means, where an iterative and exploratory process made us arrive at the assessment that setting k to 6 clusters provided a plot that was highly readable and exhibited a certain degree of explanatory power (figure 3). The axes in figure 3, showing a projection that reduces the multidimensional vector space to two dimensions for visualisation, do not represent any variables in the original data. As with all such projections, the axes, therefore, do not have any direct interpretation. Rather, what matters are the relative positions of points: points (articles) that are close in the high-dimensional space of the vector model are mapped to nearby points in the 2D figure to emphasise clusters.

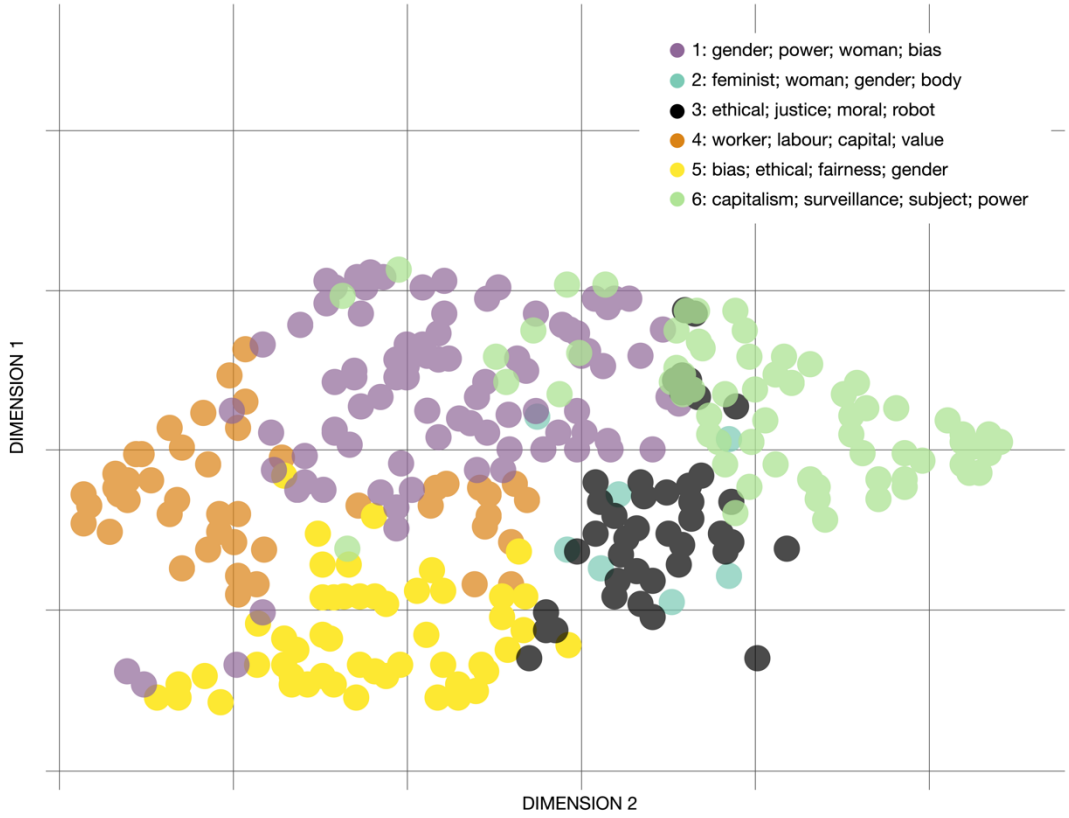

Figure 3: A UMAP 2D projection[3] of the doc2vec vectors, coloured by cluster.

We extracted top lists of commonly occurring words for each identified cluster of articles and subjected them to an interpretive reading starting from the top and moving down the list. Unsurprisingly, the top terms in all clusters included words such as 'artificial', 'intelligence', 'data', 'computers', 'machines', and so on. The objective of our readings of these lists, however, was to identify among the top words the most distinguishing words in terms of what focus the critical analysis of each given cluster of papers appeared to have. Relying on the condensed proxies of these frequency lists, keeping with the spirit of distant reading, rather than scouring the full texts as such at this point, some general patterns arose fairly quickly. In the legend of Figure 3, the four top distinguishing words, identified in this qualitative manner, for each cluster are listed.

It must be emphasised here that while there are differences between the clusters, there are also, naturally, many overlaps and a certain number of articles that could easily fit into two or more clusters. Indeed, there are many different clusterings possible, and while redoing the clustering step seems to indicate a certain amount of stability to the clusters presented here, the specifics can and do shift, particularly for those items that lie near the borders between cluster centroids.

Both clusters 1 and 2 appear to include articles with the common trait of dealing with the critical analysis of issues around gender and power. If one is to discern any differences between the two, it seems that cluster 1 is representative of a predominant focus on questions of AI bias, while cluster 2 may, to a larger degree, be marked by more common mentions of feminist perspectives and on issues of embodiment in re-

[3] The parameters were n_neighbors=8, spread=0.7, min_dist=0.2, metric='euclidean'

lation to AI. In cluster 1, there appears to be a prominent occurrence of words associated with the perspective of “data feminism” (Nasrin 2023, 141) and that engage with AI through the lens of “gender theories [that] have developed a nuanced set of tools through which to analyse issues around inclusion, exclusion and justice” (Jones 2018, 96). Articles in cluster 2 – the common words suggest – are, to a larger extent, positioned within the field of feminist AI research (Keith 1994) and seem to deal with posthuman perspectives on gender and embodiment as related to AI. Such scholarship takes an interest in “the feminist potentials [of dismissing] the separation of biology and technology [...,] merging the flesh and the machine through embodied narratives” (Ferrando 2014, 3). It appears, in fact, that cluster 5 is more about what could be labelled as mainstream AI ethics scholarship and perspectives rather than about the critical forms of AI studies that we purport to study here. The reason for this cluster still resulting from our dataset is that it, indeed, includes references to issues of gender and feminism, which matches our search terms for constructing the sample. As a side note, then, we see that when AI ethics scholarship bridges into the critical domain, it appears to do so most often by referring to gender equality and less so by referencing, for example, decolonialism and racism.

Clusters 3 and 5, while being considered through our sampling strategy as belonging to the area of critical AI research, are both, at the same time, overlapping with more techno-legal and policy-oriented scholarly discourses that centre around key terms such as ‘bias’ and ‘fairness’, which can have a somewhat corporate connotation from the perspective of critical theory (Lindgren 2023a). Cluster 3 does this while also engaging to a certain degree with issues of robotics, with characteristic contributions that “advocate caution against developing artificial moral agents” (Herzog 2021, 1) and that discuss in what ways autonomous systems are “troublesome in the ethical domain” (Paraman and Anamalah 2023, 1). Articles in Cluster 5, to the extent that they engage with specific cases, are mostly considering AI-related injustices that relate to gender. Most of all, however, these articles deal more broadly with issues of AI bias and fairness within a legalistic framework. While we, of course, cannot assess the specific level of ‘criticality’ of individual articles in this cluster, the key terms at the aggregated level align with what would seem to be the 'least critical' out of the topical clusters. Such an interpretation aligns with what Bassett and Roberts (2023, 80) write:

> “Critical studies of artificial intelligence pick up where normative models of ‘responsible AI’ end. The latter seeks to build acceptance for AI decision making by incorporating a form of ‘best practice’ into AI techniques and machine learning models to ensure ‘equity’ of some order (racial, gender, less often sexuality or disability, hardly ever class) through the avoidance of bias and ‘transparency’ in their operation. Responsible AI is, then, primarily a technical solution to technosocial problems.”

In line with this, we might assume that this cluster may align more with ‘corporate’ perspectives on such critical topics. However, looking closer at the articles, we find that many mentions of these concepts are, in fact, in the context of critiquing these very perspectives. For example, in one of the articles in the corpus, the authors write that:

> “It is commonplace for large tech companies to talk of AI ethics and notions like ‘responsible AI,’ and many companies have internal research and policy development around how such an ideal could be achieved [...] What ambitions like

> these entail is not always clear, and such initiatives are sometimes accused of being a kind of 'ethics washing', staving off regulation and marginalising issues that do not fit the corporate agenda (Furendal and Jebari, 2023, 36)."

Furthermore, in another one of these articles, the author contends that such corporate efforts aim "to reconcile capitalist AI production with ethics. However, AI ethics is itself now the subject of wide criticism" (Steinhoff 2023, 1). Most clearly, however, the issues of capitalism and class that we aim to map with this study are found in clusters 4 and 6, where the central topics revolve around capital, labour, power, value, subjectivity, and surveillance. As described in the section on methods, these two clusters were chosen for the second step of distant reading to get a richer picture of the discourse within these areas and what that may say about how issues around capitalism are dealt with in critical AI research. While xluster 4 is explicitly centred around notions of labour, capital, workers, and so on, the connection to capitalism and class in cluster 6 might appear less direct. The reason for including the latter in the closer analysis, however, is that the importance of the notion of surveillance (also in conjunction with'power' and subjectivity) ties in these strands of research strongly to the discourse on surveillance *capitalism* as developed in and around the writings of Zuboff (2019).

3.2. Word Embeddings

The word2vec model trained on the 95 articles in the clusters (4 and 6) representative of articles where issues of capitalism appear to be central to the analysis is visualised in figure 4. In this UMAP projection plot, clusters (using k-means) are differentiated by colour, and labels for key terms have been added beside the clusters for improved readability.

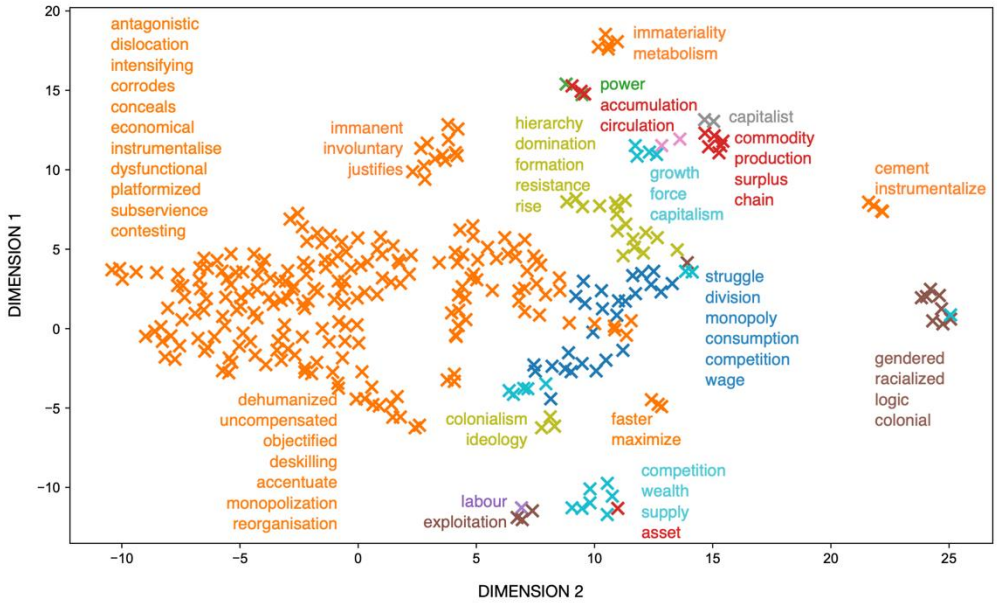

Figure 4: Clusters of words in critical AI articles (word2vec model).

Figure 4 shows the word associations that were learned by the word2vec model. The goal with this kind of analysis is to capture the semantic relationships between words based on a logic where words that appear in similar contexts have similar vector representations. To be able to visualise the vector representations in a two-dimensional figure, the high-dimensional vectors must be reduced. This is the same logic as in Figure 3, where the vectors represented sets of articles, rather than sets of words inside articles. Again, the UMAP algorithm used for this reduction, shows items – words, in this case – that are similar close to each other. As explained in relation to figure 3, the axes in these types of plots have no fixed meanings, apart from representing the two dimensions along which UMAP has chosen to show the structure of the data. As for interpreting the patterns shown in figure 4, words that are close to each other in the graph (i.e., same cluster and same colour) are semantically related. For instance, words like ”commodity”, “production”, and “surplus” are clustered together, suggesting that these concepts are closely related in the textual data the word2vec model was trained on. A visualisation such as this one creates conditions for a distant reading of the articles in question.

Making such a reading of this part of our corpus, using this plot as a heuristic tool, reveals a discursive landscape quite clearly aligning with central and Marxist concepts within critical theory. Even if the discourse can be decomposed into clusters in this way, most of them reflect standard terminology in Marxist culture and technology studies (Fuchs 2019). These include “ideology”, “labour”, ”monopoly”, ”logic”, ”commodity”, ”surplus”, and ”accumulation”, as well as notions of “resistance” and ”struggle'”, alongside many other related concepts that are symptomatic of analyses of the political economy. Notably, connections are also made to other struggles than those rooted in class (”gendered”, ”racialised”, ”colonial”) and, to some degree, to the context of digital capitalism (”platformised”). Note, importantly, that none of our sensitising concepts, as discussed in the methods section (*class, classes, capital, capitalism*), are visible in figure 4, as what is shown are all *other* terms that are among the most similar to these point of entry into the model.

The general insights gained from our distant readings, then, are, first, that notions of capitalism and class can be identified as being central in distinctive subfields (cf. digure 3) of the broader discourse of critical AI research and, second, that the discourse within those fields, while making connections to the digital, and to gender, race, and colonialism, largely reflect a distinctive and consistent terminology marked by critical concepts in Marxist analysis of the political economy.

3.3. Terms Inside and Outside of the Subsample

In table 1 below, we list a number of terms and their relative prevalence in documents taken from Clusters 4 and 6, compared to the rest of our corpus. It should come as no surprise that these relative frequencies differ between these two sets of documents, given that the clustering is taken from doc2vec, which records exactly these types of relative word frequencies. Moreover, in Section 3.1, we listed some of these terms to describe the tendencies of the clusters. In this section, we investigate these differences more fully, noting in particular that while some terms are more common within the subsample than outside of it (or vice versa), the magnitude of difference may not always be as expected. Notably, while we in some sense expect that “capital” and “capitalis*” are more commonly used within the 95-paper subsample, it is perhaps surprising that

the difference is as stark as indicated by the table (these specific terms being almost four times as common per document within the subsample compared to the rest of the dataset, while “[Mm]arx*” is three times as common, and “class” occurs almost twice as often). References to inequality are also more common within the subsample than outside, though by a smaller margin. References to both “discrimination” and “bias” are, however, significantly more common outside of the subsample than within it.

“Bias”, in particular, is almost five times as common outside the subsample as within it, which could indicate an aversion to using that term for researchers who deal more closely with issues of class and capitalism. Clearly, “bias” has come, in parts of the field of critical AI research, to connotate views of justice and fairness that are more individualistic and simplified compared to more structural analyses and explanations. Sometimes, deeper-cutting forms of systematic injustices and inequalities are obscured by a focus on surface-level statistical or computational bias. It can be argued, then, that truly critical analysis should go beyond bias to focus instead on "entrenched social values" and “even more naturalised and culturally sedimented understandings of the world and ourselves as human beings” (Bloom 2023, 35). A one-sided emphasis on “bias” may contribute to discursively constituting far-reaching problems in society, economy, and culture as issues of mere glitches to be adjusted rather than faced in more complex and multi-faceted ways.

Another pair of terms that occur much more frequently in documents outside of the subsample than within it are “fairness” and ”justice”. “Fairness” in particular occurs more than four times as often per document outside the subsample than within it. This pattern can be interpreted in terms of such notions referring to presumed universal values, which can be referenced without much further analysis or definition. This in contrast to references to “coloni” which are more than three times as common within the subsample than outside of it. Similarly, “solidari” is more than twice as common within the subsample.

Additionally, an interesting observation in terms of word choice can be made when looking for references to racial oppression. While searching for “raci*” shows no major difference between the subsample and the full dataset of papers, searching instead for ”racis*“ and ”racia*“ turns up a relative difference, where ”racis*“ (i.e., “racist” or ”racism'”) is more common outside the subsample, while ”racia*“ (”de/racial/ised”) is more common within it.

Theme	Term	Inside	Outside
Marxist/class	class	1.22	0.72
	capital	22.91	6.28
	capitalis[tm]	16.05	3.71
	[Mm]arx	5.73	1.89
	surplus	1.27	0.11
	solidari	0.5	0.19
	ideolog	2.04	1.03
Gender	gender	5.6	12.69
	queer	0.71	1.01
	misogyn	0.08	0.2
	sexis	0.6	0.76
	feminis	5.92	6.92
	patriarch	0.55	1.24
	[Dd]ata [Ff]eminism	0.15	1.13
Racism	raci	7.84	7.36
	racia	5.42	3.8
	racis	1.72	2.58
	coloni	8.25	2.52
Inequality/justice	inequality	1.89	1.2
	bias	3.14	14.26
	discriminat	2.83	4.5
	justice	3.75	8.39
	fair	2.83	9.47
	fairness	1.19	6.01

Table 1: The average amount of occurrences of various terms per document, in- and outside of the subsample. N.b. that “class” was required to be surrounded by whitespace, while the other terms were counted also as part of compound words (e.g. “bias” in ”debiasing”). Bracketed groups match either of the bracketed letters ([Mm]arx matches both “Marxist” and ”Marx”).

It is also interesting to note that references to ”gender”, ”queer”, ”misogyn*“ and ”sexis*'” are significantly more prevalent outside of the subsample than within it, as are references to ”feminis*”. Interestingly, doing a more detailed search for references specifically to “data feminism” (including references to Catherine D’Ignazio and Lauren Klein’s (2020) book of the same name) reveals one of the more stark differences between the in- and out-of-subsample groups: Within the subsample there are four articles in total that reference the term, with two of them making a single reference to it, one only using the book as a reference, and one[4] being a critique of the both the term and the book. In comparison, outside of the subsample, there are ample mentions of both the term and the book, with 33 articles in total employing the term in one sense or another, a total of 241 times.

However, references to ”patriarch*“ occur more frequently within the subsample than outside of it. How do we understand this? It seems premature both to analyse this

[4] Tacheva (2022)

as Data Feminism in particular and works building on it to be somehow opposed to anti-capitalist struggle or to a deeper engagement with class consciousness and to take the relatively lower number of references to misogyny and sexism within the subsample to indicate somehow less of a commitment to gender justice and equality from those authors. Instead, it is likely a question of emphasis and lineage.

3.4. Popular Themes

The insights gained above raise the further question about to what degree a critical theory perspective that emphasises capitalism and class is integrated more broadly in the larger corpus of critical AI articles, and to what degree different axes of oppression are tackled in common or by themselves. This relates, in particular, to the relative differences in (collective) emphasis uncovered in Section 3.3 between works focused on class and gender oppression.

A qualitative review of the titles and abstracts of the 95 papers in the subsample generally indicates that many of these papers are in the genre of critical theory as such, rather than being AI papers that incorporate a capitalism/class perspective. In other words, taking the full dataset as a fair view of the field of Critical AI Studies, it appears that critical perspectives on gender (certainly) and race (partly) have been mainstreamed into the field more broadly than have perspectives drawing on a critical analysis of capitalism. A deeper insight into this pattern can be gained with the help of the heatmap plotted in figure 5.

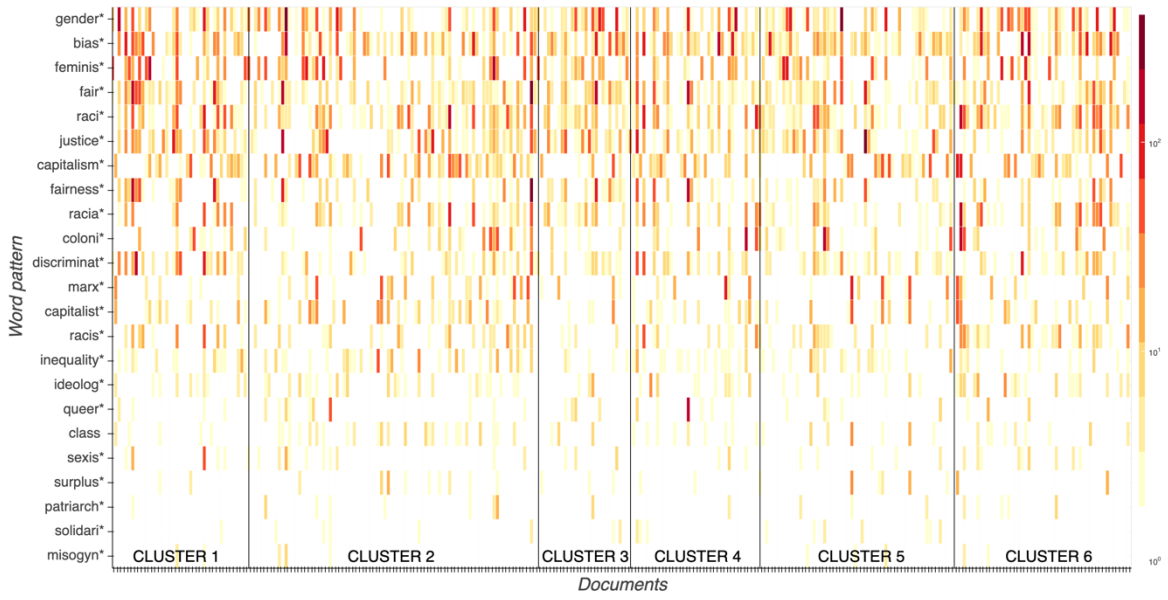

Figure 5: Heatmap based on counts of occurrences of a set of word patterns across the 300-document corpus. Documents are grouped by cluster.

As the figure shows, references to “gender*” and ”feminis*” are very frequent in a large proportion of the critical AI articles, indicating that – as presently constituted – critical AI research is largely focused on the dimension of gender. Furthermore, the high presence of ”bias” as compared to, for example, ”discrimination”, “inequality”, ”racis*”, ”misogyn*”, and so on, indicates that the problematic notion of bias, as discussed by Meredith Broussard (2023) and others, has a strong foothold in substantial parts of critical AI research. Naturally, some of these mentions may be part of texts that critique it. However, somewhat more critically precise and politically salient concepts such as

”sexis*“ are not as commonly used. Furthermore, references to ”raci*”, ”racia*”, and ”coloni*” are also fairly common, if not nearly as frequent as references to gender and feminism. Clearly, however, even if references to ”capitalism*“ (a fairly broad notion that can also be used quite descriptively) occur in a substantial number of articles, references to ”capitalist*”, “class”, and ”surplus*“ are much further down the list.

The methodological strategy for our qualitative review used a two-stage approach. First, we screened the articles based on their titles and abstracts to ascertain their relevance to our aims. Second, we conducted a more detailed examination of some papers where closer inspection was needed to better understand the concepts employed and the context of the research. Through this review, we identified two broad genres of critical AI papers. First of all, papers that can be considered to be “AI research” but with the added critical edge of race, gender, or decolonial theories. Examples of these are the papers “On the Ethics and Practicalities of AI Risk Assessment and Race” (Hogan et al. 2021), “Algorithmic Fairness and Structural Injustice: Insights from Feminist Political Philosophy” (Kasirzadeh 2022), “AI for Social Justice: New Methodological Horizons in Technical Communication” (Graham and Hopkins 2022), and “Asking More of Siri and Alexa: Feminine Persona in Service of Surveillance Capitalism” (Woods 2018).

In such studies, researchers explore how concepts of race, gender, and decolonial theories can be integrated into the field of AI research. This means, then, that such perspectives are incorporated as complementary tools in papers that have the advancement of AI technology or its applications in specific domains as their main focus. Such cases, then, on the one hand, are good examples of how critical theory can aid in analyses in other, more applied fields than its own. On the other hand, however, there is always the risk that any deeper-cutting critique gets suppressed by other aims. Such potential suppression can happen in situations where the primary focus of the research is on advancing AI technology or its applications rather than critically examining the underlying power structures and inequalities that may be perpetuated or exacerbated by these same advancements.

In a less integrated fashion, then, another genre of papers rather match a template of “Marxism-as-applied-to-AI”, and these mostly appear within the 95-paper subcorpus. This genre comes across as quite homogenous (cf. figure 4) and as more free-standing and independent in relation to scholarship about AI development and implementation. Examples of this type of paper are studies such as “Rethinking of Marxist Perspectives on Big Data, Artificial Intelligence and Capitalist Economic Development” (Walton and Nayak 2021) and “Data Capitalism and the Counter Futures of Ethics in Artificial intelligence” (Dixon-Román and Parisi 2020). This type of study is often characterised by fairly deep explorations of Marxist and other critical theories and concepts that are then applied to the case of AI. A potential risk to this style of analysis is that it may be lacking in empirical grounding and validation and might downplay real-world applicability and materialist politics in favour of focusing on developing concepts and theory.

As an aside, but giving some interesting complementary knowledge about the position of Marxist-influenced AI research within general AI scholarship more broadly, we returned to the Scopus database for a supplementary analysis. Here, we queried Scopus specifically in the subject areas of Social Sciences and Arts and Humanities for ((marx* OR capitalis*) AND (ai OR “artificial intelligence”)) in the abstract, title, or keywords, which yielded 330 results, while a similar query for only ai OR “artificial intelligence” yielded 49,254 results. This means, very roughly, that 0.7 % of AI research in these social sciences and humanities would appear to be drawing on Marx’s theories

or otherwise putting the notion of capitalism at the fore. Note, however, that both queries capture many articles that are not *about* AI as such, but rather *uses* AI to investigate some other topic, as discussed in section 2.1.

4. Discussion and Conclusions

The aim of this study was to analyse if and how class and capital(ism) are articulated and positioned in critical AI research. To respond to this aim, we created a corpus of critical AI research, in the form of articles indexed by Scopus. We then made this corpus the subject of largely computational, but also exploratory and qualitative, analysis to uncover prominent patterns both in terms of how class and capitalism appear as prominent, or not, in broader topical clusters of articles and how the language used in articles reflects the articulation and position of issues of, and perspectives on, class and capital. In doing this, our analysis focused on comparing patterns in articles positioned in topical clusters where the analysis of capitalism and class was central and articles that sprung from other areas within critical AI research.

Clearly, the background for us wanting to carry out this study, to get to know more about the degree to which the classical, Marxism-inspired strain of critical theory analysis plays a role in the emerging field of critical AI studies, is that we deem this perspective to be of crucial importance. This is because the recent uptake of generative AI models and subsequent investments to integrate these in a myriad of new applications and business models introduces a whole new set of societal risks, as these systems are inherently inscrutable, with nobody really understanding how they work and how they fail (Dobbe 2022, Bender et al. 2021). Under a veil of corporate marketing anthropomorphising and mystifying these systems' poor functioning, an agenda is rolled out that includes all of society in large-scale experiments that will trigger and solidify many new harms and injustices to individuals, marginalised groups as well as our collective information provision and democratic institutions.

Generally, the emerging field of Critical AI Studies, in contrast, does not take the presently prevailing conception of AI as a given. Instead, it attempts to, on the one hand, question assumptions that underlie most or all of current AI research and interrogate how its conception, deployment, construction and use reifies and reinforces unjust power relations and, on the other, consciously investigate alternative modes of AI development and use with alternative characteristics.

Many strands of this field are at least partially rooted in critical theory, drawing heavily on conceptions of structural and institutional power. In particular, works like *Data Feminism* (D'Ignazio and Klein 2020) and many others in feminist AI and data science draw on ideas from intersectional feminist studies to interrogate how patriarchal and other power structures manifest in AI and propose new approaches to AI and data science based on feminist principles, grounded in an ethics of care and based in solidarity among different groups. Similarly, attempts by scholars such as Abeba Birhane, Syed Mustafa Ali, Renata Avila Pinto and others to decolonise AI and the computational sciences more broadly are partly rooted in similar emancipatory struggles and draw on critical race theory in particular. One of the key contributions of our study is that we investigate the particular role of analyses of AI that see class and capitalism as their prime dimensions of analysis in relation to the more common forms of critical AI research that have increasingly come to incorporate feminist and decolonial perspectives.

A different mode of critical engagement with AI is geared more towards overall rejection, e.g., in the book *Resisting AI: An Anti-Fascist Approach to Artificial Intelligence* by Dan McQuillan (2022). In it, the author argues that since the very construction of

"artificial intelligence", it has been rooted in racist and sexist structures and in colonial and capitalist categorisations and epistemic assumptions on measurability and regimentation. Furthermore, he argues that AI is currently reinforcing, entrenching, and, in some cases, further intensifying capitalist and state-enacted exploitation and abuse. Thus, the whole "AI'" project is seen as tainted, and any technique and strategy to slow down, hinder, or sabotage its use is a valid tactic.

Based on our empirical analysis of how Critical AI Studies that focus on perspectives of class and capitalism relate – in terms of its articulations and positioning – to other strands of critical AI research, we have been able to come to the following conclusions:

- Articles focusing on capitalism and class stand out as a fairly distinct subgenre within the field of Critical AI Studies (figure 3).
- Within this subgenre, we identify a scholarly discourse which bears the characteristics of classic (post-)Marxist critical theory and which draws on an analytical vocabulary centred around class and capitalism (antagonism, struggle, labour, commodities, surplus, etc) (figure 4).
- An analysis of the language used in articles within this subsample, as compared to that in articles outside of it, shows traces of a discursive rift whereby many of the core terms in the subsample are relatively confined to those core papers, while concepts relating to gender, race, and (de)colonialism are more broadly present in the corpus as a whole (table 1, figure 5).

In sum, then, the empirical study that we have carried out to improve the understanding of which analytical strands within critical AI studies are prominent, as well as how they are integrated (or not), has shed light on one overarching and striking pattern. While on the one hand, themes that relate to gender, race, and (de)colonialism have – by comparison – found a certain degree of representation, and integration, with topics and concepts relating to what one might label "mainstream AI research". On the other hand, Marxist research on AI has not entered this mainstream to the same extent. As seen in figure 4, these discourses are certainly characterised by a pronounced (post-)Marxist critical theory undertone, replete with its analytical lexicon. Yet, the relatively confined usage of core terms, as contrasted in figure 5 and table 1, suggests a divide which risks creating or proliferating silos within the field. Therefore, we believe that it is crucial for the future and more holistic development of critical AI studies that analyses rooted in the frameworks of class and capitalism do not remain relegated to the peripheries. Instead – and ideally – they should be seamlessly integrated into the very mainstream of AI research. Doing so will provide avenues for understanding the potential for broader forms of solidarity by relating feminist, anti-racist and decolonial with anti-capitalist and class struggles. Only if this integration is achieved, we can ensure a comprehensive and actionable understanding of the implications of artificial intelligence and its underlying computational infrastructures, capturing their multifaceted and globally expressed socio-economic, ecological, and political repercussions.

References

Bassett, Caroline and Ben Roberts. 2023. Automation Anxiety: A Critical History. In *The Handbook of Critical Studies of Artificial Intelligence*, edited by Simon Lindgren, 79-93. Cheltenham: Edward Elgar.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610-623. Virtual Event Canada: Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Birhane, Abeba, and Olivia Guest. 2021 Towards Decolonising Computational Sciences. *Kvinder, Køn & Forskning* 2021 (1): 60-73. https://doi.org/10.7146/kkf.v29i2.124899

Birhane, Abeba, Pratyusha Kalluri, Dallas Card, William Agnew, Ravit Dotan, and Michelle Bao. 2022. The Values Encoded in Machine Learning Research. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 173-184. FAccT '22. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3531146.3533083

Birhane, Abeba, Vinay Uday Prabhu, and Emmanuel Kahembwe. 2021. Multimodal Datasets: Misogyny, Pornography, and Malignant Stereotypes." arXiv. https://doi.org/10.48550/arXiv.2110.01963

Bloom, Peter. 2023. The Danger of Smart Ideologies: Counter-Hegemonic Intelligence and Antagonistic Machines. In *The Handbook of Critical Studies of Artificial Intelligence*, edited by Simon Lindgren, 33-42. Cheltenham: Edward Elgar.

Broussard, Meredith. 2023. *More Than a Glitch: Confronting Race, Gender, and Ability Bias in Tech*. Cambridge, MA: The MIT Press.

Buitinck, Lars, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, et al. 2013. API Design for Machine Learning Software: Experiences from the Scikit-Learn Project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, 108-122.

Couldry, Nick, and Ulises A. Mejias. 2020. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford, CA: Stanford University Press. https://doi.org/10.1515/9781503609754

Delanty, Gerard, and Neal Harris. 2021. Critical Theory and the Question of Technology: The Frankfurt School Revisited. *Thesis Eleven* 166 (1): 88-108. https://doi.org/10.1177/07255136211002055

Delfanti, Alessandro. 2021. Machinic Dispossession and Augmented Despotism: Digital Work in an Amazon Warehouse. *New Media & Society* 23 (1): 39-55. https://doi.org/10.1177/1461444819891613

D'Ignazio, Catherine, and Lauren F. Klein. 2020. *Data Feminism*. Strong Ideas Series. Cambridge, Massachusetts: The MIT Press.

Dixon-Román, Ezekiel, and Luciana Parisi. 2020. Data Capitalism and the Counter Futures of Ethics in Artificial Intelligence. *Communication and the Public* 5 (3-4): 116-121. https://doi.org/10.1177/2057047320972029

Dobbe, Roel I. J. 2022. System Safety and Artificial Intelligence. arXiv. https://doi.org/10.48550/arXiv.2202.09292

Dobbe, Roel, and Meredith Whittaker. 2019. AI and Climate Change: How They're Connected, and What We Can Do about It. AI Now Institute, Medium, October 17.

Engstrom, Emma, and Karim Jebari. 2023. AI4People or People4AI? On Human Adaptation to AI at Work. AI & SOCIETY 38 (2): 967-968. https://doi.org/10.1007/s00146-022-01464-5

Falagas, Matthew E, Eleni I Pitsouni, George A Malietzis, and Georgios Pappas. 2008. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and Weaknesses. *The FASEB Journal* 22 (2): 338-342.

Ferrando, Francesca. 2014. Is the Post-Human a Post-Woman? Cyborgs, Robots, Artificial Intelligence and the Futures of Gender: A Case Study. *European Journal of Futures Research* 2 (1): 43. https://doi.org/10.1007/s40309-014-0043-8

Fosch-Villaronga, Eduard, Hadassah Drukarch, Pranav Khanna, Tessa Verhoef, and Bart Custers. 2022. Accounting for Diversity in AI for Medicine. *Computer Law & Security Review* 47: 105735. https://doi.org/10.1016/j.clsr.2022.105735

Fuchs, Christian. 2019. *Marxism: Karl Marx's Fifteen Key Concepts for Cultural and Communication Studies*. New York: Routledge. https://doi.org/10.4324/9780367816759

Fuchs, Christian. 2022. *Foundations of Critical Theory: Media, Communication and Society Volume Two.* London: Routledge. https://doi.org/10.4324/9781003199182

Furendal, Markus, and Karim Jebari. 2023. The Future of Work: Augmentation or Stunting? *Philosophy & Technology* 36 (2): 36. https://doi.org/10.1007/s13347-023-00631-w

Graham, S. Scott, and Hannah R. Hopkins. 2022. AI for Social Justice: New Methodological Horizons in Technical Communication. *Technical Communication Quarterly* 31 (1): 89-102. https://doi.org/10.1080/10572252.2021.1955151

Hall, Stuart. 1997. The Work of Representation. In *Representation: Cultural Representations and Signifying Practices*, edited by Stuart Hall, 13-64. London: SAGE.

Herzog, Christian. 2021. Three Risks That Caution Against a Premature Implementation of Artificial Moral Agents for Practical and Economical Use. *Science and Engineering Ethics* 27 (1): 3. https://doi.org/10.1007/s11948-021-00283-z

Hintz, Arne, Lina Dencik, and Karin Wahl-Jorgensen. 2019. *Digital Citizenship in a Datafied Society*. Cambridge, UK; Medford, MA: Polity Press.

Hogan, Neil R., Ethan Q. Davidge, and Gabriela Corabian. 2021. On the Ethics and Practicalities of Artificial Intelligence, Risk Assessment, and Race. *The Journal of the American Academy of Psychiatry and the Law* 49 (3): 326-334. https://doi.org/10.29158/JAAPL.200116-20

Jones, Emily. 2018. A Posthuman-Xenofeminist Analysis of the Discourse on Autonomous Weapons Systems and Other Killing Machines. *Australian Feminist Law Journal* 44 (1): 93-118. https://doi.org/10.1080/13200968.2018.1465333

Jones, Maurice. 2023. Mind Extended: Relational, Spatial, and Performative Ontologies. *AI & SOCIETY*. https://doi.org/10.1007/s00146-023-01769-z

Jones, Phil. 2021. *Work Without the Worker: Labour in the Age of Platform Capitalism*. London: Verso.

Kak, Amba, and Sarah Myers West. 2023. AI Now 2023 Landscape: Confronting Tech Power. *AI Now Institute*. https://ainowinstitute.org/2023-landscape

Kasirzadeh, Atoosa. 2022. Algorithmic Fairness and Structural Injustice: Insights from Feminist Political Philosophy. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 349-356. Oxford United Kingdom: ACM. https://doi.org/10.1145/3514094.3534188

Keith, William. 1994. Artificial Intelligences, Feminist and Otherwise. *Social Epistemology* 8 (4): 333-340. https://doi.org/10.1080/02691729408578760

Laclau, Ernesto, and Chantal Mouffe. 1985. *Hegemony and Socialist Strategy: Towards a Radical Democratic Politics.* London: Verso.

Le, Quoc, and Tomas Mikolov. 2014."Distributed Representations of Sentences and Documents. In *Proceedings of the 31st International Conference on Machine Learning*, edited by Eric P. Xing and Tony Jebara, 32: 1188-1196. Proceedings of Machine Learning Research. Bejing, China: PMLR. https://proceedings.mlr.press/v32/le14.html

Levy, O., & Goldberg, Y. 2014. Dependency-Based Word Embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 302-308. https://doi.org/10.3115/v1/P14-2050

Lindgren, Simon. 2020. *Data Theory: Interpretive Sociology and Computational Methods*. Cambridge: Polity.

Lindgren, Simon. 2023a. *Critical Theory of AI.* Cambridge: Polity.

Lindgren, Simon. ed. 2023b. *The Handbook of Critical Studies of Artificial Intelligence*. Cheltenham: Edward Elgar.
McInnes, Leland, John Healy, and James Melville. 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv. https://doi.org/10.48550/arXiv.1802.03426
McQuillan, Dan. 2022. *Resisting AI: An Anti-Fascist Approach to Artificial Intelligence*. Bristol: Bristol University Press.
Meho, Lokman I, and Kiduk Yang. 2007. Impact of Data Sources on Citation Counts and Rankings of LIS Faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology* 58 (13): 2105-2125.
Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. Distributed Representations of Words and Phrases and Their Compositionality. *Advances in Neural Information Processing Systems* 26.
Mingers, John, and Evangelia Lipitakis. 2010. Counting the Citations: A Comparison of Web of Science and Google Scholar in the Field of Business and Management. *Scientometrics* 85 (2): 613-625.
Mongeon, Philippe, and Adèle Paul-Hus. 2016. The Journal Coverage of Web of Science and Scopus: A Comparative Analysis. *Scientometrics* 106: 213-228.
Moretti, Franco. 2005. *Graphs, Maps, Trees: Abstract Models for Literary History*. London: Verso.
Moretti, Franco. 2013. *Distant Reading*. London: Verso.
Nasrin, Sohana. 2023. New Ways of Activism: Design Justice and Data Feminism. *Social Movement Studies* 22 (1): 140-144. https://doi.org/10.1080/14742837.2021.1967132
Ongweso Jr, Edward. 2021. Amazon's New Algorithm Will Set Workers' Schedules According to Muscle Use. *Vice*, April. https://www.vice.com/en/article/z3xeba/amazons-new-algorithm-will-set-workers-schedules-according-to-muscle-use
Paraman, Pradeep, and Sanmugam Anamalah. 2023. Ethical Artificial Intelligence Framework for a Good AI Society: Principles, Opportunities and Perils. *AI & SOCIETY* 38 (2): 595-611. https://doi.org/10.1007/s00146-022-01458-3
Rehurek, Radim, and Petr Sojka. 2011. Gensim-Python Framework for Vector Space Modelling. NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic 3 (2).
Roberge, Jonathan, and Michael Castelle. 2021. *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies*, edited by Jonathan Roberge and Michael Castelle. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-56286-1
Steinhoff, James. 2023. AI Ethics as Subordinated Innovation Network. *AI & SOCIETY*. https://doi.org/10.1007/s00146-023-01658-5
Tacheva, Zhasmina. 2022. Taking a Critical Look at the Critical Turn in Data Science: From 'Data Feminism' to Transnational Feminist Data Science. *Big Data & Society* 9 (2) https://doi.org/10.1177/20539517221112901
Walton, Nigel, and Bhabani Shankar Nayak. 2021. Rethinking of Marxist Perspectives on Big Data, Artificial Intelligence (AI) and Capitalist Economic Development. *Technological Forecasting and Social Change* 166: 120576. https://doi.org/10.1016/j.techfore.2021.120576
Whittaker, Meredith. 2021. The Steep Cost of Capture. *Interactions* 28 (6): 50-55. https://doi.org/10.1145/3488666
Woods, Heather Suzanne. 2018. Asking More of Siri and Alexa: Feminine Persona in Service of Surveillance Capitalism. *Critical Studies in Media Communication* 35 (4): 334-349. https://doi.org/10.1080/15295036.2018.1488082
Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.

About the Authors

Petter Ericson
Petter Ericson is a postdoctoral research fellow in the research group for Responsible AI at the Department of Computing Science at Umeå University, Sweden. His research is focused on the political implications of technology, and on exploring alternative technological and computational paradigms with the potential to distribute, rather than accumulate, power. He is also interested in facilitating clear and honest communication between seemingly distant groups of people interacting with the same socio-technical systems, and in clarifying issues of common interest, as well as of conflict, between them.

Roel Dobbe
Roel Dobbe is an assistant professor in the Information and Communication Technology section of the Faculty of Technology, Policy, and Management at Delft University of Technology. His research embraces system safety as a lens to understand how harm arises in data-driven, algorithmic and artificial intelligence systems, leaning on the history of complex systems subject to software-based automation. Roel has a PhD in Electrical Engineering and Computer Sciences from the University of California Berkeley (2018) and a MSc in Systems and Control from Delft University of Technology (2010). He is an active contributor to the establishment of governance practices for algorithmic and AI systems in public organisations, including in public administration, energy systems and healthcare. He also serves as a board member to Foundation PublicSpaces, a Dutch coalition of public organisations in public media, cultural heritage, festivals, museums and education working together to reclaim the internet as a force for the common good and advocating for and building a new internet that strengthens the public domain.

Simon Lindgren
Simon Lindgren is Professor of Sociology at Umeå University. He is also the director of DIGSUM, an interdisciplinary research centre studying the social dimensions of digital technology, and the editor-in-chief of the *Journal of Digital Social Research*. His research is about the transformative role of digital communication technologies (internet and social media), and the consequences of datafication and algorithms, with a particular focus on politics and power relations. He uses combinations of methods from computational social science and network science, and analytical frameworks from interpretive sociology and critical theory. Lindgren's books include "Critical Theory of AI" (2023). "Data Theory" (2020), "Digital Media and Society" (2017; 2022), and "New Noise" (2013).