

# Roskilde University

The Algorithmic Gut Feeling-Articulating Journalistic Doxa and Emerging Epistemic Frictions in Al-Driven Data Work

Jannie Møller, Hartley; Bonde Thylstrup, Nanna

Published in: Digital Journalism

DOI:

10.1080/21670811.2024.2319641

Publication date: 2025

Document Version Publisher's PDF, also known as Version of record

Citation for published version (APA):

Jannie Møller, H., & Bonde Thylstrup, N. (2025). The Algorithmic Gut Feeling–Articulating Journalistic Doxa and Emerging Epistemic Frictions in Al-Driven Data Work. *Digital Journalism*, *13*(3), 438-456. https://doi.org/10.1080/21670811.2024.2319641

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
  You may not further distribute the material or use it for any profit-making activity or commercial gain.
  You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact rucforsk@kb.dk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 03. Jul. 2025



# **Digital Journalism**



Routledge

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rdij20

# The Algorithmic Gut Feeling – Articulating Journalistic Doxa and Emerging Epistemic Frictions in Al-Driven Data Work

# Hartley Jannie Møller & Nanna Bonde Thylstrup

**To cite this article:** Hartley Jannie Møller & Nanna Bonde Thylstrup (01 Mar 2024): The Algorithmic Gut Feeling – Articulating Journalistic Doxa and Emerging Epistemic Frictions in Al-Driven Data Work, Digital Journalism, DOI: <u>10.1080/21670811.2024.2319641</u>

To link to this article: <a href="https://doi.org/10.1080/21670811.2024.2319641">https://doi.org/10.1080/21670811.2024.2319641</a>

9	© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
	Published online: 01 Mar 2024.
	Submit your article to this journal $\ensuremath{\ \ \ }$
lili	Article views: 392
Q <sup>L</sup>	View related articles ☑
CrossMark	View Crossmark data ☑
4	Citing articles: 1 View citing articles 🗗



#### RESEARCH ARTICLE

**3** OPEN ACCESS



# The Algorithmic Gut Feeling – Articulating Journalistic Doxa and Emerging Epistemic Frictions in Al-Driven Data Work

Hartley Jannie Møller<sup>a</sup> and Nanna Bonde Thylstrup<sup>b</sup>

<sup>a</sup>Department of Communication and Arts, Roskilde University, Roskilde, Denmark; <sup>b</sup>Department of Arts and Cultural Studies, Copenhagen University, København, Denmark

#### **ABSTRACT**

This article explores the epistemic practices and doxa of data workers in a news organisation in Denmark that is currently developing and experimenting with artificial intelligence (AI)-driven recommender systems, machine learning and natural language processing solutions. Previous literature on the changing epistemologies of digital journalism has focused on the increased role of metrics and the transformed practices inside newsrooms, as well as on how journalists perceive and articulate the computational. This article advances these studies by focusing on how data scientists perceive and articulate "the journalistic" when building AI systems for distributing news. Developing the notion of "the algorithmic gut feeling", the article highlights different frictions present in the articulations of the journalistic doxa in Al-driven data work concerning (1) how to algorithmically define ethics, (2) how to algorithmically categorise and understand relevance, and (3) how to algorithmically curate "a good mix" for the front page. The emerging frictions and algorithmic gut feeling are key to understanding how the doxa of data workers involved and deeply invested in "the good of journalism" at times also transforms journalistic epistemologies of what constitutes "news" and "the right mix" of content in the service of a democratic public.

#### **KEYWORDS**

Epistomologies; GutFeeling; Al; datawork; ethnographic; Doxa; domain

#### Introduction

News organisations are increasingly experimenting with artificial intelligence (AI)-driven solutions in the production and distribution of news. These initiatives are embedded within a longer trajectory of what Matt Carlson (2018) calls "measurable journalism", a term that encapsulates the cultural and material shifts to digital platforms in the pursuit of real-time, individualisable, quantitative data about audience consumption practices. These real-time data on users are increasingly deployed in the development of AI-driven distribution of content with recommender systems (recsys), machine learning (ML) models, and natural language processing (NLP) solutions.

This article explores these developments through an ethnographic study of a news organisation in Denmark. Existing research emphasises that news organisations differ regarding the use of recommenders (Møller 2022); that personalisation projects reconfigure news values (Møller Hartley, 2013) and contradict public service values, such as universalism (Sørensen 2019), that they impact diversity (Möller et al. 2018); and that they can be designed to support democratic values (Helberger 2019). Less research has focused on the design processes of NLP and ML with Al in the context of news organisations, which is the focus and aim of this article. In this article, we investigate how data scientists and IT professionals employ what we call algorithmic gut feelings in the development of Al systems in news organisations. These gut feelings are important because they have an impact on the long-term "epistemic claims" of news journalism (Ekström and Westlund 2019).

We theoretically examine the implicit valorisations in data work through the lens of "doxa". In journalism studies, news values are typically seen as part of a doxa, which can be described as a "gut feeling" (Schultz 2007), meaning that journalists and editors find it difficult to explain why some stories are deemed important, while others are not. This sensemaking is part of their doxa, which Bourdieu also describes as "having a feel for the game" (1998, 81). We utilise the concept to understand and analyse the algorithmic gut feeling among data scientists in the process of developing Al systems compared to specific journalistic doxa. We combine the field theoretical approach with perspectives from science and technology studies to examine tensions between the contextual specificity of domain knowledge and the digital practices, platforms, and infrastructures that enable it (Hansen and Thylstrup 2023). More specifically, the article investigates how journalistic epistemologies are translated into Al systems and which domain frictions emerge in the translation process.

#### The Field of Automated News

Recent work on Al in journalism highlights the increasingly crucial role ML plays in defining and transforming journalism and journalistic values. One branch of literature has focused on large-scale data collection, algorithmic data analysis, and computational practices in the production and dissemination of news—a trend that goes by many different names, including computational journalism (Diakopoulos 2015), data journalism (Fink and Anderson 2015), data-driven journalism (Parasie 2015), robot journalism (Carlson 2015), and journalism as programming (Parasie and Dagiral 2013). These strands have, for instance, highlighted how audience measurement software shapes editorial choices and gatekeeping processes in the newsroom. Often, they also link these observations to broader tendencies of platformisation and centralisation of big tech ownership. Generic web metrics used by Google Analytics, for example, promote the commercialisation of news and the segmentation of audiences, which can adversely affect the diversity of information available and the decision-making processes in the newsroom.

In addition to these works, an important strand of research also delves into how algorithms shape journalistic epistemologies and ontologies of, for example, news values (Schjøtt Hansen and Hartley 2023) and how algorithmic technologies can be understood both as a way to support democratic values, such as diversity (Helberger

2019), as well as a point of friction with traditional public service values, such as universalism (Sørensen 2019). Christin (2020), for instance, shows how algorithms not only generate news feeds based on signals from networks and the preferences of advertisers but also assume agency in the physical space by presenting "algorithmic publics" to newsroom actors. She highlights that algorithmic metrics change the values and gut feelings of journalism because journalists become "disciplined by numbers" to veer in a more commercial direction (Ibid.). Focused more specifically on automated journalism, Carlson (2018) argues that the emergence of "algorithmic judgement" in relation to "journalistic judgement" impacts the ontologies of news and the legitimising discourses surrounding them. Studies focusing on personalisation and recommender systems show how the current wave of personalisation within the news industry builds on approaches and algorithmic models that are also used by large commercial platforms, such as Amazon, Google, and YouTube (Bodó 2019; Smith and Linden 2017).

Taina Bucher's (2017) work on how journalists and newsroom staff articulate the computational is an especially relevant reference for this article. Focusing broadly on both personalised recommender systems and automated news production, she finds three claims in these articulations: that machines do not have instincts, that democracy can never be personalised and that the computational is something to think with, rather than simulate. She further argues that what can and cannot be calculated is not merely a technical question; it is also a deeply social, cultural, political, and economic one.

This article contributes to Bucher's line of research, but rather than asking how journalism articulates "the computational", we turn the question around and ask: How does the data science domain articulate "the journalistic"? In doing so, we move the focus out of the newsroom and into the IT department. Our choice is motivated by our observation of a case that illustrated the broader tendency of the "platformisation of the news" (van Dijck, Poell, and de Waal 2018, 49), which involves a growing interdependency between news organisations and the data and model infrastructures supplied by commercial platforms. Furthermore, our changed focus is motivated by the increasing importance of new entrants into the field in the form of data scientists (see also Chew and Tandoc 2022).

To understand the implications of these developments, it is necessary to expand the empirical realm beyond the role of journalism to engage ethnographically with peripheral actors in news organisations (Vulpius 2022). We therefore draw on the insights in the work of Bucher (2017), asking the following research question: How are journalistic epistemologies translated into AI systems, and which domain frictions emerge in the translation process?

# Linking Doxa, Domain Theory, and Journalistic Epistemologies

News journalism typically relies on the gathering, processing, and presentation of information: access to and evaluation of sources (fundamental practices of research), choice of styles of writing and talking (practices of communication), knowledge of visualisations, and editing into various formats (practices of presentation) (Carlson 2017). Referring to the work of Zelizer (1992), Carlson argues that news journalists claim authority as "credible spokespersons of 'real-life' events" from their knowledge of a world beyond people's everyday experiences, as well as in their audience-friendly and truthful presentation of news (Carlson 2017, 4). This results in what Ekstrøm, Ramsälv, and Westlund (2022, 765) has termed the "epistemic value of news" (news as valuable knowledge for the public) and the "epistemic practices". They show how metrics are used as a superior standard in deciding the epistemic value of news. This is expressed in strategies, guidelines, and discussions in the newsroom, while fulfilling reasonable truth claims is mainly taken for granted as an epistemic practice. Similar to Ekström, Ramsälv, and Westlund (2022), we focus on how different standards of epistemic value are prioritised and reconciled in a data-driven news culture, but rather than looking at the newsroom, we investigate the frictions concerning the epistemic value of news in the data science work developing ML models for the domain of journalism. More specifically, we study (1) how data workers' understandings of news and journalism as valuable knowledge are articulated in discussions during data work, and (2) how standards are enacted in the epistemic practices of professional data science judgements, when data workers decide and justify what is valuable journalism to publish with ML and Al systems? However, what we found lacking in the conceptualisation of epistemic practices is how to analytically approach the fact that agents in an intersection of two fields are deciding on the epistemic value of news and journalism, interpreting these according to their own domain-specific, taken-for-granted rules of the game. For this purpose, we turn to the concept of doxa as it is used in field theory.

According to Bourdieu (1998, 57), doxa is the taken-for-granted dimension of social practice, the seemingly natural, which we rarely make explicit and rarely question. Developing the Bourdieusian framework for the field of journalism, Schultz (2007) highlights that what journalists experience as their "journalistic gut feeling" entails both explicit news values—the dominant (orthodox) and dominated (heterodox)—as well as their silent, taken-for-granted, doxic counterparts. From a field perspective, doxa is closely linked to the concept of habitus. In general, Bourdieu (1998) uses the metaphor of having a feel for the game in his description of habitus: "Having a feel for the game is having the game under the skin; it is to master in a practical way the future of the game; it is to have a sense of the history of the game" (81). Apart from the embodied history, habitus also encompasses the "internalization of specific rules and practices" (Maares and Hanusch 2020, 11). Thus, our analytical focus is on how data workers, who are not part of the journalistic field by training, use their specific gut feelings to interpret how the Al systems can be developed according to the epistemic practices and the epistemic value of journalism.

What epistemic practices are at play for them when evaluating what constitutes a good Al model, for example, and what specific doxic values are used in these evaluations? As doxa is often tacit, it is difficult to examine and ask the informants directly about whether we were dependent on the ethnographic observations and concretely paid attention to moments when journalism and editorial goals were articulated in the process of building the models. Heterogeneous and interdisciplinary projects, such as the development project we focus on in this article, bring together the domains of data science and journalism, making explicit the differences between their doxa, including meanings and practices in workflows that otherwise often remain implicit. We concretely observed that these articulations are full of frictions,

indicating the need for a theoretical perspective for exploring how the theories, concepts, data, and hardware travel across scientific divides, in this case, journalism and data science.

Whereas sociological theories of doxa tend to focus on intra-field relations and frictions, science and technology studies scholars provide valuable theoretical and methodological frameworks for exploring "evaluations," the "logic of domains" (Ribes et al. 2019) and the "science frictions" (Edwards et al. 2011) that emerge when different domains are brought into contact. As David Ribes et al. (2019) notes, the appeal of machine learning is among other things its reputation as an effective science that that can be applied on a general level. This understanding is informed by a framing of data science as either 'emptied' of domain knowledge or as a method that has assembled enough specific domain knowledge. Pre-trained algorithms that are void of domain-specific content in the early layers may thus transition more smoothly to new domains, where adjustment and modification can take place in situ (Ribes et al. 2019). Thus, generative models today are often developed with a view to reusing across multiple other configurations of data and algorithms (Thylstrup et al. 2022). The concept of machine learning as 'empty' results in the practice of 'prospecting' within data science, involving 'the work of rendering data, knowledge, expertise and practices of worldly domains available or amenable to engagement with data scientific method and epistemology' (Slota et al. 2020, 1). At the same time, the work of finetuning also involves ongoing translation work between data scientists and the domain they are adapting the models to. These encounters sometimes result in what Paul Edwards et al. (2011) call "science frictions". Science friction refers to the time, energy, and human attention costs involved in transferring data between groups and organisations, as well as between machines. Each of these interfaces represents "a point of resistance where data can be garbled, misinterpreted, or lost" and are therefore often the site of "conflicts, disagreements, and inexact, unruly processes" (Edwards et al. 2011, 669).

Edwards et al. (2011) point out that focusing on these frictions is especially important "as datasets become increasingly commoditised, 'mined', and exchanged among distant disciplines", which is exactly the case with the PIN Project we observed. Such an approach allows us to challenge the binaries between data science and the domain of journalism by emphasising the inherently collaborative and heterogeneous nature of many data science practices. Thus, we do not perceive the doxas of the two domains of journalism and data science as belonging to completely separate and contradicting fields in struggles and conflicts over power but as intertwined agents from two fields, who are mutually "attuning" to the epistemic practices of the others. As Parmiggiani, Østerlie, and Almklov (2022) note, such practices "do not happen in clean rooms but involve various activities, negotiations, and actors beyond the analysts, such as project managers and product designers". Thus, from this perspective, although institutional settings' data science methods may be configured as potentially domain-independent, they almost always unfold within evolving institutional organizing principles and epistemic practices of scientific domains (Ribes et al. 2019; Parmiggiani, Østerlie, and Almklov 2022). This exchange between different epistemic practices also unfolded during our observations, as the data scientists sought to "attune" their approaches to the journalistic domain and its epistemologies.

#### The case: Al System Development at a Legacy Tabloid Newspaper

The analysis in this article builds on ethnographic fieldwork and in-depth interviews at Ekstra Bladet (EB), a part of the larger media conglomerate JP/Politikens Hus, from February to November 2022. EB is a national tabloid newspaper and is one of the most read in Denmark in its online version. It has around 300 employees, but a total of 2,100 people are employed by JP/Politikens Hus with the headquarters located in central Copenhagen.

The methodological framework of the article draws on the anthropologies of technology (Lanzeni et al. 2022). More specifically, although the article takes the specific technicalities of digitality seriously in empirical cultural research, we also approach digitisation principally as a relational matter that encompasses technical, social, and cultural dimensions (Klausner 2022). Moreover, the article develops a particular sensibility towards the "anticipatory practices" (Anderson 2007) of our interlocutors, which involves focusing on how they act in the present as they relate to the "proximate futures" of their own practices and the trajectories of their work (Dourish and Bell 2011). Concretely, we observed the Platform Intelligence in News (PIN) project, an Al innovation project partly funded by a Danish applied research foundation and partly funded by EB themselves.

The main purpose of the PIN Project was to develop responsible and effective transformers for the Danish news industry, with the dual purpose of developing new recommender systems and generating automated journalistic content using ML and NLP. A transformer model is a neural network that learns context, most often with ML, and thus learns the meaning of text or images by tracking relationships in sequential data like the words in this sentence, or in this case relationship between different journalistic words in specific articles. The project both co-developed new transformer models in Danish and used pre-existing transformer models, finetuning them for the news sector. In both cases, the manual annotation of Danish news articles from the news organisation played an important role, as these annotators created the corpus of identifiable words used to train the transformer models built by EB. The project made use of open-source platforms, primarily Hugging Face, which is a ML and data science platform hosting both NLP and ML models and data science discussion boards.

The PIN Project is thus both a development project in a news organisation carried out in collaboration with partners at universities and spanning computational disciplines (NLP and recommender systems), humanities, and social science researchers. This interdisciplinary and trans-institutional nature of the project meant that the head of innovation and research at EB and PI of the PIN project, who granted us access to carry out the fieldwork at EB, was also part-time employed at a university as an industrial post doc. However, the research activities were funded by the research foundation (and not the news organisation), and no NDAs were signed during the course of the research. The researchers working on the project had the academic freedom to define the parameters of their research within the overarching research question posed by the PIN Project, namely, how to develop fair AI systems in news organisations and how to methodologically and conceptually assess the ethical and political implications of such systems.

The nature of the project makes it an extreme case: it is the most comprehensive Al development project carried out in the Danish news industry, and because of its unfolding structure, it is not only commercially motivated but also involves purer research-oriented components. The specificities for each news organisation, their news values and commercial model and audiences notwithstanding, the influx of similar-minded projects across the news industries in Denmark and beyond makes our observations of the epistemological practices transferable to other contexts.

In addition to the people funded directly by the grant, the project also involved the participation of—and collaboration with—full-time employees at EB, including people working on the recommender models, front and back-end developers, and annotators coding the large datasets. Many of these were not formally part of the PIN Project but shared overlapping and intersecting workflows that made it challenging to analytically draw distinctions between PIN and beyond-PIN Project members. For these reasons, we will simply refer to our observations as related to the PIN Project, plus related activities and associated employees.

### **Our Methodological Approach**

Our observations concretely took place in the development department, located in an open-plan office one floor above the editorial section. The spatial location meant that we had less focus on the newsrooms at EB, although journalists were implicitly present for both the observations and interviews as the "users" of PIN Project-related systems implemented during this period. We participated in project management meetings as well as everyday work meetings two or three times per week during the observation period. In total, the ethnographic fieldwork consisted of more than 350 hours of observation both of the concrete every day and the project meetings, meetings with management, and meetings with other related projects going on at the same time. The project meetings were recorded, the files were transcribed, and we took pictures of the presentations and whiteboard drawings. These were used to contextualise our own field notes and were not analysed pr. se. We were unable to participate in all meetings, but the team we observed was helpful in explaining what we missed and also how the projects had moved forward between meetings. Moreover, we kept track by following a Slack channel on which the team shared information about internal developments interspersed with news from the broader fields of data science. The content of the channel was saved as screenshots on a closed university server and used as background information to understand some of the discussions unfolding between different developers and management during our observations. During the fieldwork, we undertook informal, day-to-day interviews, as well as eight in-depth, semi-structured interviews with developers, annotators, and project managers working within and beyond the PIN Project lasting between 60 and 90 min each. The formal interviews were transcribed while the informal day-to-day interviews were not recorded as they occurred across coffee machines, in the canteen or across tables, but they were noted down along the days we were there in shared closed university folder that both authors had access to.

Methodologically we follow the ethnographic approach of Pink et al. (2016) and both authors have independently and systematically gone through the extensive bank of empirical data, making notes about what is related to doxa in the specific domain and how this compared to and differed from journalistic doxa and epistemic values and practices (see above). We met every other week during the observations to discuss and compare field notes, meaning that the analysis developed from a collaborative and inductive approach and over quite a long period of time, making the observations more and more focused as time progressed (Pink and Morgan 2013). In the final phase of analysis, we coded the material for situations in which frictions between the domains emerged, and this allowed for three overall frictions in the data work to emerge in the material, which forms our analysis into three parts presented below.

# Analysis: Articulating the Journalistic in Data Work

Each part of the analysis represents different aspects of the frictions present in the articulations of the journalistic in the data work at EB. Overall, we observed that the data scientists were aware of the journalistic epistemic practices, but frictions emerged when they had to translate these into their own domain doxa. The frictions concern interpreting (1) how to define ethics, (2) how to categorise and understand relevance, and (3) how to curate the right content for the front page. In our analyses, we relied on previous literature on news work concerned with doxic news values and news judgement in the journalistic domain, which allows for a comparison between the doxa in the two domains and the frictions emerging in the contact between them.

# Articulating Ethics as Fairness and Frictions of "Fluffiness"

An important pillar of the PIN Project was the development of an ethical approach to the implementation of Al-driven ML in news production. The articulation of ethics in the project drew on state of the art in data science, as well as communication and sociological theories on filter bubbles and bias. Workshop agendas were often tailored to the members present. A kick-off workshop that involved project members and EB employees framed the project's raison d'être with reference to a gap in research on ethics in Danish large language models (LLMs), noting that state of the art in existing Danish-language LLMs display little or no focus on "bias, fairness, and explainability". However, our observations showed that the data scientists in the PIN Project perceived journalism as a distinct domain with its own set of ethical concerns. One ambition of

Table 1. Overview of epistemic practices inherent in the doxa and gut feelings in the two domains.

Domain epistemic practices	Doxa/Journalistic gut feeling	Doxa/Algorithmic gut feeling
Ethics	Related to sources, facts and objectivity in the news product	Related to accuracy, fairness of machine learning models and non-bias in data
Categorising Relevance	A value-based judgement to informing, "relevant" to the public, while also catering to specific segments of users	Accuracy and probability in calculating and predicting user-needs in and with data
Curating the "right mix"	According to a "good mix" of content and topics balancing marketing and journalistic values.	Numerical and durable mix of content and topics, adaptive to user needs and preferences.

the project was, for example, to articulate a working definition of ethics that could not only be operationalised by data scientists but also fit within the broader framework of journalism as an important societal pillar. In this phase, the three values the data scientists specifically worked with were accuracy (of the models they built), efficient use of resources (compute power and economic resources spent), and utility bias. The researchers from the NLP component in the PIN Project would often articulate the focus on ethics as an opportunity to advance state of the art within the field of NLP more broadly, referring to the "gap in research" and how the activities of the PIN Project could contribute to filling it. The examples illustrate that despite the mission of including the journalistic practice, ethics as an epistemic practice was articulated through data science doxa, more specifically though a focus on fairness and bias in the data work (see Table 1).

The encounters between epistemologies of ethics in data science and journalism often elicited discussions among project members working with data science about how to implement more "fluffy" notions of journalistic ethics into computational items. As one data scientist pointed out in an interview, ethical considerations as articulated by the domain of journalism were "difficult to translate into the everyday workflows of data science" (Interview 3). Another data scientist, in a separate interview, noted that the issue of fairness was a "ticklish topic" in data science, proceeding to share his impression of the challenges associated with implementing overarching ethical discussions in concrete data science projects (Interview 5). Significantly, as he emphasised, the frustration he sometimes experienced did not stem from a lack of interest but rather from a difference in the doxa of data and social sciences regarding data epistemological practices:

... it is impossible to disagree with that, of course; it is insanely important, that much is clear. It's just also difficult to do something about it ... It's just because ... I don't have an answer either ... it's just that modelling; it's a hardcore discipline based on data and what data is. And then there are these softer considerations, which you must also take into account, but which cannot be set in a formula. It's just really difficult. (Interview during fieldwork)

The proposition of a contrast between the "hard" practices of data science and the "soft" considerations of the surrounding journalistic domain generates "science frictions" (Edwards et al. 2011) in the translation of journalistic epistemologies of ethics into data scientific doxa.

Friction should not only be understood as a "clash" of cultures; however, but friction also produces movement, actions, and effects, and it often involves productive tension (Tsing 2011). For example, the in-house data scientist working on the PIN Project explained that working within the domain of journalism shaped his way of thinking about ethics more than some of his previous places of employment. He attributed this to the fact that the field of journalism is conscious about—and critical towards social media platforms. He often encountered journalists expressing that there are "many dangerous things about Al". The data scientist noted that, within the field of journalism, this generates a conviction, for instance, that a "front page must be human made". He compares this experience with his previous work in a large engineering company that developed hearing aids. There, he noted, "It was just about producing the best sound quality and ensuring that these people who were hard of hearing could hear. There was not so much ethics". Thus, the friction means that the journalists interpret and translate ethics to their own domain, but specifically into the notions of fairness and bias related to data and distribution.

However, while frictions emerge between cultural epistemologies in the mundane workflows of data science in news organisations, these tensions also engender new algorithmic gut feelings for ethics. This was, for example, observed in the process of data labelling in the annotation team as one crucial point of translation. Asked how they relate to the concepts of "fairness and bias", the in-house annotators responded that they took these concepts into consideration only to the extent that they wished to work in an "ethically correct" manner, noting how their actions often unfolded in uncertainty and how their gut feeling served to guide them.

... we have put a lot of work into making it as fair as possible when we tag. We try to see it from the outside. We have one example where there was something with a crafts fair. Or was it a crafts shop? Yes, a clothing shop for craftspeople. And here we thought automatically that we had a category called "men's fashion", because we thought that craftspeople are men. There, we were a bit biased. But then we also thought that, of course, we should also tag it as women's fashion. I even thought that there was a skirt in the photo or something. Like crafts women's clothes, or whatever you call it. So, in this way, we actively care about tagging without bias. There, we even thought our actions were a bit controversial. I never know if we choose to do it like this because we are women. But maybe it was also a bit wrong to do it, because if people think about crafts people, they usually think about men. So, in a way, it was a bit incorrect that we also tagged it as women. If we are to represent the general thought, that is. If everyone thinks about craftspeople as men, shouldn't that be how we tag them? (Interview annotator, fieldwork)

Later in the interview, the annotators mentioned several more examples and their predictive implications. For instance, they noticed that the model they were working with did not tag items with "soccer" when they concerned women's soccer, whereas it could accurately apply the tag to male soccer stories. They couldn't intervene in this case because it was "out of their hands", but the situation alerted them to the potential biases in the system and their ethical implications, leading them to more actively correct the model when it reproduced this pattern. Thus, the imperative to rely on ethical practices as a part of Al-driven data work because they were a part of a news organisation was then translated into a doxa of avoiding bias in the categorisations and in the data. The ethical guidelines related to the epistemic practice of producing news, concerning sources, objectivity, and truthfulness of facts, were not observed to play any role in the discussions of ethics related to ML in a news organisation or in the day-to-day data work.

These three examples show that algorithmic gut feelings vary across different data science components and rarely express a clear distillation of pure data science concerns or a frictionless merging of journalistic and data science doxa. More accurately, they emerge from encounters between the journalistic institution with its epistemic practices and value and data scientists working within it, who, on the one hand, work within the "domain-independent" doxa of data science; nevertheless, it also emphasises how the journalistic domain affects their own epistemological practices.



#### Articulating Relevance as Accuracy in Relation to User Needs

As journalism becomes increasingly data-driven (Christin 2020; Petre 2021), debates emerge on how to ensure normative journalistic criteria of "relevance" and impact in algorithmically driven processes. As a part of journalistic gut feeling, relevance has been identified as a news value, defined as "Stories about groups or nations perceived to be influential with, or culturally or historically familiar to, the audience" (Harcup and O'Neil 2017, 1482). When data scientists working on the PIN Project articulated relevance, they predominantly adhered to concepts prominent in data science, such as "accuracy" and "probability". Getting a model to work requires a great deal of data work, including labelling journalistic content for topic classification model training. This process provides a window into how the "journalistic" is translated into Al models.

One example was an exchange concerning the use of AI models to produce a taxonomic system that could facilitate the topic classification processes needed to recommend content to different users. The baseline used for building this topic classification for the PIN Project was the "Content Taxonomy" developed by the Interactive Advertisement Bureau (IAB), a standard for the advertisement industry that aims to make content classification consistent across the media industries. On the Bureau's website, this taxonomy is described as a service to publishers providing them with "a consistent and easy way to organise their website content", and differentiate easily between, for example, "sports" versus "news" versus "wellness" material". At several meetings, this AIB baseline was discussed in relation to its place in the PIN Project, specifically how to finetune the taxonomy developed for marketing purposes to editorial categories. In the end, the project team decided to combine the IAB with a taxonomy developed by IPCT, an organisation representing news media organisations specifically, and added their own original categories to this. Discussing this merger, the in-house data scientist noted the need to balance the specificity of the media organisation with generalisability, so that what they developed could also be used by others. This comment triggered an exchange between project members on whether the taxonomy should be drawn from the article database "as is", or whether the database should be adjusted by removing categories of articles that were not deemed important for advertising. A discussion between a developer and the annotators on a category labelled "beekeepers" illustrates this "translating friction". Following daily discussion on different categories of stories, they jointly decided to remove this category along with many others because of their "low resolution", that is, its specificity that most likely would not have editorial relevance. The developer also explained that they had consulted with marketing before doing so. The project manager and head of innovation at EB then asked whether the removals of these "small" categories were based on editorial or marketing concerns, to which the data scientist responded that they were primarily out of pragmatic concerns. This exchange shows that an algorithmic gut feeling can emerge out of the productive friction between several epistemologies, including marketing ("Is this specific category important for subscribers?"), editorial goals, and data scientific workflows ("How can we build the most efficient model?").

In addition to the frictions around marketing taxonomies, frictions emerged around how to define "relevance" as either in relation to "user interest" or to more journalistic normative goals of "serving a public". A data scientist described the process of increasing the probability that the models would deliver results that the user finds relevant. This process relied upon gut feelings of "when it does not work", including continual qualitative and quantitative monitoring and, for the data scientist on the team, an assessment of whether the recommendations matched the input offered. He described this sense as being triggered "in relation to the goals one wishes to achieve", explaining that "the goal is to develop a model that can find articles, that is similar to this article, which we can recommend after this article. And if it is totally unrelated, then one could say, okay, this is not how we want it to behave". This example illustrates how data scientists develop a sense of how to "make the model behave well" but according to user needs rather than journalistic relevance criteria as understood by Harcup and O'Neil (2017, 1482). The data scientists rarely specified what this "criteria of relevance" was; it simply became a "feeling for the game". When elaborated upon, the criteria referenced specific groups of EB users (e.g., "many of our readers are men") or in relation to questions such as "How can we convert users to subscribers?".

The attempts to translate and understand the journalistic doxa were complicated by the fact that the data science team was located on a floor different from the editorial part of the organisation within the EB building. In an interview with the two in-house annotators, they explained that this spatial separation sometimes challenged the project development and caused friction:

Annotator 1: We had an example in a project in which we had a category called news. We then had some subcategories called local news and international or something like that ... We had to talk about how it would actually be nice to know when a journalist or what a journalist would define as news.

Annotator 2: Yes, very much.

Annotator 1: But we just never did. Maybe because we are on another floor. It is just a bit easier to run over to the other table instead of up and down the stairs.

Annotator 2: Yes, because that was the problem. What is a news story? You could just say that all articles in a news organisation are news. But that isn't ... that doesn't say much when you have to tag them all as news. (Interview, fieldwork)

This exchange illustrates our observation that the categorising of news is highly relational, as previous studies have also shown (Møller Hartley, 2011, 2013). Further, it shows how the staff in the developing department sought an understanding of "the journalistic" doxa in this case what could be defined as "news". The annotators at EB ultimately defined news as an article that "presents a turn in the developments of a case". However, they also noted that each categorisation presents new challenges: "That is the challenge, because we constantly have to think about the temporal frame or the context of the publication of the article we are tagging. Because we might as well work with an article that is 5 years old. We have to be careful there. Because what was news back then? That can be a bit difficult". As the following section also shows, discussions of different journalistic categories were important for the accuracy of the models they built for recommender systems, making sure that users were presented with a personalised "right mix" on the front page.



#### Articulating "The Right Mix" of Topics for Content Exposure

Even pre-dating Al-driven distribution of content with the use of recommender systems, the journalistic front page of a site is prioritised according to what is referred to as "a right mix" reflecting the position and profile of the specific publisher (Kammer 2013; Møller Hartley 2013). A public service provider's front page will thus have a different mix than a tabloid newspaper, the latter typically featuring more stories on entertainment and gossip (Møller Hartley 2012). The New York Times features more international and national news stories and depends more heavily on its own staff for both stories and images than other metropolitan newspapers (Kim and Chung 2017). Such a mix will seek to balance between market concerns measured by clicks and values of providing content serving the public (Møller Hartley 2013). Ekstrøm et al. linked the epistemological truth-seeking element to methods of doing journalism and the presentation of that content; however, we observed that the process of ensuring the curation of a certain mix is also shaped by Al-driven models of news distribution. In our observations, the algorithmic gut feeling manifested as a form of "organisational knowing" of the journalistic doxa that was translated into the domain of data science through measurable categories, which were balanced on the page as a mix of, for example, maximum x percent sport and maximum x percent crime stories.

The profile of tabloids, including EB, has historically been shaped by a focus on content related to crime, sport, gossip, and what could be labelled "sex and relationship" content. This content is not erotic per se but focuses on (famous) people's sex lives, nudity, and gossip, illustrating its highly commercial media model. In recent years, EB has sought to reorient focus from nudity to news and lifestyle, which we observed translated into modelling and ML practices. Here, data workers translated the values of "the good EB mix" into the evaluation of the model, that is, an evaluation of how well the model was performing and how accurate the model would be according to the numeric goals of a certain mix. However, in contrast with the journalistic balancing of, for example, commercial and public service-oriented stories as a part of the "right" mix, the "right mix" from the data science perspective is evaluated as topics consumed by the user. When the models were tested, the results were often evaluated by the the project manager, who also functioned as a liaison between the editorial goals and the goals of the models they built. In the meetings, they discussed the risks of a recommender model giving the reader "too much crime" or "too much erotic" content and reflecting on how previous user patterns shaped the modelling outputs. One proposed solution to "rebalancing" the output was to change user profiles, adding users with more diverse consumption patterns to the user dataset. Another was to exclude the consumption patterns related to the traditional "Page 9" content (the Danish equivalent of Page 3, the British newspaper convention of publishing a large image of a topless female model). At times, the developers reflected on the frictions between the domain of editorial strategies and the doxic values of journalism from the doxa of data science. As one data scientist working on the recommender models noted:

We have been told that our recommender models must not recommend "sex and relationship" content, even though there could well be a good reason for it. I mean, people would

be crazy for it. So it is—I don't know if I would call it an ethical consideration or what I would call it. This is some kind of consideration. I actually think it's quite interesting in a way. Because, in a way, it's quite inconsistent. I kind of think that if we publish content like that, page 9 girls, sex and relationship and content like that, then I think that it must be something that we would editorially stand for. And then I think it's a bit curious that we don't allow our models to recommend it. If you say A, you must also say B. And so, I don't think they have done that there. Although I do understand that if we had a model that constantly 100% recommended Page 9 girls, then we might well have to rethink it. (Interview, fieldwork)

The quote shows how the developers sought to align computational evaluations with editorial strategies, with the project manager often working as a liaison and knowledge broker between the editorial mission and the data science doxa. It also illustrates the difficulties that the data workers faced in perceiving what the editorial goals were, and even when they knew them, they did not always translate directly into the evaluations of the ML models they built.

#### **Discussion: Reconfiguring Journalistic Epistemologies?**

The case analysis showed that the epistemic practices of the data science components oscillate between data science and journalism, producing an algorithmic gut feeling that was data driven and shaped by data science epistemic practices but also differed from the traditional journalistic gut feeling, as illustrated in Table 1.

Curating the front page is increasingly influenced by algorithmic systems and the categorisations of what "news is", and as a consequence, "the journalistic" and editorial goals are increasingly defined by departments other than the editorial department in news organisations. The question is: How do these systems and their underlying epistemologies impact the epistemic practices of journalists, editors, and managers? Such questions gain in relevance as individual journalists and editors become embedded in increasingly complex—and, to many, seemingly opaque—systems, reaching far beyond news organisations. Studies have shown emerging engagement capital (Lindblom, Lindell, and Gidlund 2022), which suggests an increasing focus on measurable news judgement and categorisations. Furthermore, research has shown that digital journalism involves a larger and more heterogeneous set of social actors, technological actants, and audiences than ever before, who will also have a say in translating epistemic practices, as we have seen in this article. Chew and Tandoc (2022) argue that a new form of habitus develops in tech start-ups as a result of new entrants in the field merging with the traditional journalistic habitus, changing and driving the journalistic field in specific directions. As we demonstrated in this study, some of these actors are data scientists who influence journalistic epistemic practices by translating them in specific ways into the distribution and automation of journalistic content. An influence that should not be underestimated. However, we would be cautious of suggesting that this emerges new forms of capital or habitus, as the doxic values of the algorithmic gut feeling seem to suggest a pull towards an economic pole already (and always) present in the field of journalism.

By focusing on the development of ML systems and how data scientists perceive and articulate "the journalistic", we see how the notion of an algorithmic gut feeling captures these oscillations and the productive frictions they give rise to and that the data scientists navigate the two doxa's and attempt to merge the journalistic doxa into their own epistemic practices. The literature in this area often treats the fields of journalism and technical systems as separate (Petre 2021, Christin 2020) but nevertheless shows that algorithmic judgement plays an increasing role in journalistic practices and that the data-driven work to some extent reconfigures news values and news judgement (Carlson 2015; Schjøtt Hansen and Møller Hartley 2021).

Our analysis adds to this body of work, illustrating that data science projects within news organisations have also begun to handle traditional journalistic tasks such as categorising according to genre and topics and deciding on timeliness of stories (including the practice of "routinising of the unexpected", Tuchman 1973). Viewing these processes of "articulating the journalistic" in data work via the lens of doxa, we have shown how data workers interpret the journalistic translating it into their own domain doxa. The data workers we observed devoted many meetings to ensuring that the team was able to identify the characteristics of "journalism" and encode such characteristics in the tools they developed. Both domains evaluated what the "good mix" was according to normative doxic values, but the data workers developed the models to enhance news consumption and struggled to incorporate softer and less numerical values when using their gut feeling to evaluate what the models did.

How each news organisation judges different news stories in its taxonomies is, of course, specific for each news organisation, yet we also believe that our observations about the productive frictions and exchanges shaping algorithmic gut feelings are relevant to other data science projects in news organisations. It is worth emphasising that EB is one of the most commercial news organisations in Denmark, and yet we saw strong ideals of "serving the public" in the everyday building and adapting the data science domain to the journalistic domain. This indicates that the frictions we observed are likely to be even more pronounced in less commercial news organisations. Differences between news organisations and the impact of these frictions on journalistic epistemic practices are important areas of study for future research.

## **Concluding Remarks**

Our observations and interviews point to data scientific and journalistic epistemologies as guided by different doxa, whereas our time spent at EB showed how the domains shared border zones, with each respective doxa becoming reconfigured through mutual exposure and friction. One example is the ethical aspect of journalistic epistemology; here, it is clear that even if the developers operate within the domain of data science, they are also exposed to journalistic epistemologies of ethics. Although these social science-informed epistemologies might appear "fluffy" to the data scientists and, thus, difficult to integrate into an everyday computational workflow, they nevertheless also infused and guided the project we observed, for instance, through the annotation and testing practices. Conversely, the doxa of data science also translates back into the domain of journalism, for instance, in the ongoing discussions on how to develop classification and recommendation systems and how to include editorial concerns while excluding certain material (Page 9) within ML models used to recommend and personalise content.

These exposures also expand to the categorising of news. As the ritualising of the unexpected and the norms of objectivity (Tuchman 1973) move into algorithmic systems, they are also shaped by the algorithmic gut feeling in data work. Thus, we show that accuracy, traditionally associated with a very particular meaning in journalism linked to sources and facts, is perceived in very different ways in the data science domain. As linked to the data and the models. A limitation of this study is that it has not dealt with how these systems were received in the newsroom and how the implementation of the systems would create further frictions beyond the development stage. How these semantic differences translate back into newsrooms remains to be seen, and future research should investigate how journalists and editors engage with the working algorithms, for example by gaming the system, and the subsequent re-translation of these gaming strategies back into Al-driven systems.

Studies of the interactions between journalistic and computational components in Al-driven systems have become relevant, especially considering the conditions under which many Al and ML projects unfold in news organisations. One example is the spatial layout of these projects. As noted, at EB, the IT department resides on a floor different from the editorial one, which makes in situ exchanges less likely to happen between computational and journalistic fields. Another example is how professional path dependencies shape present ML projects. In our case, the marketing department was more involved in the development phase than the editorial department. This organisational pattern seemed to be shaped by a more immediate epistemic alignment between marketing and data science work, for instance, through taxonomies and segment analysis. However, we also never observed any journalists actively seeking out the development floor to ask the data scientists or annotators questions about their work. How these special issues matter for how such projects unfold should also be a theme of future research.

There are arguably many good reasons for modularising processes in complex development projects: one is to ensure that a certain degree of knowledge is present to make relevant decisions, and another is to ensure a smoother workflow. There are also good reasons for journalists choosing not to engage actively with the development of infrastructural backend systems and Al. However, limiting exposure also runs the risk of making journalists dependent on distribution systems, which they possess very little knowledge about, a risk that is likely greater in news organisations developing ML and AI projects without the research dimension involved. Our observations centre on a project funded partly by research funds and run by a project manager with a joint affiliation between research and management and expertise in both data and social sciences. The concrete project is further rooted in a media system characterised by a strong focus on public service media ideals. This allowed the staff to expand the border zones of mutual exposure, facilitating economic and epistemological wiggle room beyond the everyday labour of "making stuff work". However, this raises the question of how such projects will develop in news organisations without such a research dimension and the public service ideals involved? It is likely to cause less friction, but it also risks not taking the journalistic doxa into account in the development process. These questions are important for further examination in future research.

For journalism, the stakes are high. Many news organisations struggle with economic difficulties, dropping advertising revenues and increasing the interdependencies of

platforms and the providers of tech solutions. To some of these organisations, ML used in, for example, recommender systems or targeted advertising has come to signify an easy solution to many of these problems. This article highlights the enormous epistemic and manual labour involved in these "solutions", labour that also generate effects inside news organisations. On the one hand, we show how data scientists become involved and invested in "the good of journalism", in the values of the specific news organisation, and developing a feel for the journalistic doxa in the process. On the other hand, we also show that such investments and involvements are anything but "smooth" and that the frictions they generate require a high level of expertise, time, and organisational adaptation to become productive, not only for the everyday workflow and the epistemic value of journalism but also for the role of journalism in democratic societies.

#### **Acknowledgements**

We express our sincere gratitude to the staff at the PIN project and EB for their generosity in taking their time to share insights, perspectives and thoughts as well as patiently explaining their practices. Their contributions have greatly enriched the scope and depth of our research. We especially wish to thank Kasper Lindskow for allowing us into the 'backrooms' of data science in such a hospitable and engaged manner. Additionally, we extend our heartfelt appreciation to the participants of the AI, Media & Democracy Lab Workshop, which included members of the AI, Media & Democracy Lab (AIMD), Center for Tracking and Society (CTS), and associates of the Platform Intelligence in News project (PIN). Their engagement and collaborative spirit have significantly enhanced our understanding of the opportunities and challenges of machine learning in news. A special thanks goes to organiser of the workshop, Arní Mar Einarsson, for extending the invitation to present segments of this article at the workshop, providing us with an invaluable platform to disseminate our research findings. Also thanks to the Journalism and Democracy Research Group at Roskilde University for feedback on an earlier draft. Moreover, we would like to acknowledge the reviewers for their constructive and thoughtful feedback and the editors of this special issue for their guidance and support throughout the publication process. And finally, we extend our gratitude to The Velux Foundation, The Independent Research Fund Denmark and Innovation Fund Denmark for allowing this research to come into fruition.

#### **Disclosure Statement**

No potential conflict of interest was reported by the author(s).

#### **Funding**

The research has been supported by The Velux Foundation, The Independent Research Fund Denmark and Innovation Fund Denmark.

#### References

Bodó, B. 2019. "Selling News to Audiences - A Qualitative Inquiry into the Emerging Logics of Algorithmic News Personalization in European Quality News Media." Digital Journalism 7 (8): 1054–1075. https://doi.org/10.1080/21670811.2019.1624185

Bourdieu, P. 1998. Practical Reason: On the Theory of Action. Cambridge: Polity Press.

Bucher, T. 2017. "Machines Don't Have Instincts': Articulating the Computational in Journalism." New Media & Society 19 (6): 918-933. https://doi.org/10.1177/1461444815624182

- Carlson, M. 2015. "The Robotic Reporter: Automated Journalism and the Redefinition of Labor, Compositional Forms, and Journalistic Authority." Digital Journalism 3 (3): 416-431. https:// doi.org/10.1080/21670811.2014.976412
- Carlson, M. 2017. Journalistic Authority: Legitimating News in the Digital Era. Columbia University Press https://doi.org/10.7312/carl17444
- Carlson, M. 2018. "Confronting Measurable Journalism." Digital Journalism 6 (4): 406-417. https:// doi.org/10.1080/21670811.2018.1445003
- Chew, M., and Jr. E. C. Tandoc, 2022, "Media Startups Are Behaving More like Tech Startups— Iterative, Multi-Skilled and Journalists That "Hustle." Digital Journalism 1-21. https://doi.org/ 10.1080/21670811.2022.2040374
- Christin, A. 2020. Metrics at Work: Journalism and the Contested Meaning of Algorithms. Princeton: Princeton University Press.
- Diakopoulos, N. 2015. "Algorithmic Accountability: Journalistic Investigation of Computational Power Structures." Digital Journalism 3 (3): 398-415. https://doi.org/10.1080/21670811.2014.976411
- Dourish, P., and G. Bell, 2011. Divining A Digital Future: Mess and Mythology in Ubiquitous Computing. MIT Press. https://doi.org/10.7551/mitpress/9780262015554.001.0001
- Edwards, P. N., M. S. Mayernik, A. L. Batcheller, G. C. Bowker, and C. L. Borgman. 2011. "Science Friction: Data, Metadata, and Collaboration." Social Studies of Science 41 (5): 667-690. https:// doi.org/10.1177/0306312711413314
- Ekström, M., A. Ramsälv, and O. Westlund. 2022. "Data-Driven News Work Culture: Reconciling Tensions in Epistemic Values and Practices of News Journalism." Journalism 23 (4): 755–772. https://doi.org/10.1177/14648849211052419
- Ekström, M., and O. Westlund. 2019. "The Dislocation of News Journalism: A Conceptual Framework for the Study of Epistemologies of Digital Journalism." Media and Communication 7 (1): 259–270. https://doi.org/10.17645/mac.v7i1.1763
- Fink, K., and C. W. Anderson. 2015. "Data Journalism in the United States: Beyond the "Usual Suspects." Journalism Studies 16 (4): 467–481. https://doi.org/10.1080/1461670X.2014.939852
- Hansen, K. B., and N. Thylstrup. 2023. "Stack Bricolage and Infrastructural Impermanence in Financial Machine-Learning Modelling." Journal of Cultural Economy 2023: 1-19. https://doi.or g/10.1080/17530350.2023.2229347
- Harcup, T., and D. O'Neill. 2017. "What is News?" Journalism Studies 18 (12): 1470-1488. https:// doi.org/10.1080/1461670X.2016.1150193
- Helberger, N. 2019. "On the Democratic Role of News Recommenders." Digital Journalism 7 (8): 993-1012. https://doi.org/10.1080/21670811.2019.1623700
- Kammer, A. 2013. "The Mediatization of Journalism." MedieKultur 29 (54), 385. https://doi. org/10.7146/mediekultur.v29i54.17385
- Kim, Yung Soo, and Deborah S. Chung. 2017. "Anatomy of Front Pages: Comparison between the New York Times and Other U.S. Major Metropolitan Newspapers." International Journal of Communication 11: 949–966. https://ijoc.org/index.php/ijoc/article/view/5898
- Klausner, M. 2022. "A More-than-Digital Anthropology: Ethnographies of Participation and Administration." Zeitschrift Für Empirische Kulturwissenschaft 118 (1+2): 5-24.
- Lanzeni, D., K. Waltorp, S. Pink, and R. C. Smith, eds. 2022. An Anthropology of Futures and Technologies (1st ed.). Routledge. https://doi.org/10.4324/9781003084471
- Lindblom, T., J. Lindell, and K. Gidlund. 2022. "Digitalizing the Journalistic Field: Journalists' Views on Changes in Journalistic Autonomy, Capital and Habitus." Digital Journalism 1-20. https://doi.org/10.1080/21670811.2022.2062406
- Maares, P., and F. Hanusch. 2020. "Interpretations of the Journalistic Field: A Systematic Analysis of How Journalism Scholarship Appropriates Bourdieusian Thought." Journalism 23 (4): 736-754. https://doi.org/10.1177/1464884920959552
- Møller Hartley, J. M. 2011. "Routinizing Breaking News: Categories and Hierarchies In Danish Online Newsrooms." In D. Domingo and E. Paterson, Making Online News: Newsroom Ethnography in the Second Decade of Internet Journalism. New York: Peter Lang.
- Møller Hartley, J. 2013. "The Online Journalist Between Ideals and Audiences: Towards a (More) Audience-Driven and Source-Detached Journalism?" Journalism Practice 7 (5): 572-587. https:// doi.org/10.1080/17512786.2012.755386



- Møller Hartley, J. 2012. Nyheder På Internettet. Frederiksberg: Djøfs Forlag.
- Möller, J., D. Trilling, N. Helberger, and B. van Es. 2018. "Do Not Blame It on the Algorithm: An Empirical Assessment of Multiple Recommender Systems and Their Impact on Content Diversity." Information, Communication & Society 21 (7): 959-977. https://doi.org/10.1080/136 9118X.2018.1444076
- Møller, L. A. 2022. "Recommended for You: How Newspapers Normalise Algorithmic News Recommendation to Fit Their Gatekeeping Role." Journalism Studies 23 (7): 800-817. https:// doi.org/10.1080/1461670X.2022.2034522
- Parasie, S. 2015. "Data-Driven Revelation?: Epistemological Tensions in Investigative Journalism in the Age of "Big Data." Digital Journalism 3 (3): 364-380. https://doi.org/10.1080/21670811
- Parasie, S., and E. Dagiral. 2013. "Data-Driven Journalism and the Public Good: "Computer-Assisted-Reporters" and "Programmer-Journalists" in Chicago." New Media & Society 15 (6): 853-871. https://doi.org/10.1177/1461444812463345
- Parmiggiani, E., T. Østerlie, and P. G. Almklov. 2022. "In the Backrooms of Data Science." Journal of the Association for Information Systems 23 (1): 139–164. https://doi.org/10.17705/1jais.00718
- Petre, C. 2021. All the News That's Fit to Click: How Metrics Are Transforming the Work of Journalists. Princeton: Princeton University Press.
- Pink, S., H. A. Horst, J. Postill, L. Hjorth, T. Lewis, and J. Tacchi. 2016. Digital Ethnography: Principles and Practice. London: SAGE.
- Pink, S., and J. Morgan. 2013. "Short-Term Ethnography: Intense Routes to Knowing." Symbolic Interaction 36 (3): 351–361. https://doi.org/10.1002/symb.66
- Ribes, D., A. S. Hoffman, S. C. Slota, and G. C. Bowker. 2019. "The Logic of Domains." Social Studies of Science 49 (3): 281-309. https://doi.org/10.1177/0306312719849709
- Schjøtt Hansen, A., and J. M. Hartley. 2023. "Designing What's News: An Ethnography of a Personalization Algorithm and the Data-Driven (Re)Assembling of the News." Digital Journalism 11 (6): 924-942. https://doi.org/10.1080/21670811.2021.1988861
- Schjøtt Hansen, A., and J. Møller Hartley. 2021. "Digital Journalism." Digital Journalism 11 (6): 924-942. https://doi.org/10.1080/21670811.2021.1988861
- Schultz, Ida. 2007. "The Journalistic Gut Feeling." Journalism Practice 1 (2): 190-207. https://doi. org/10.1080/17512780701275507
- Slota, S. C., A. S. Hoffman, D. Ribes, and G. C. Bowker. 2020. "Prospecting (in) the Data Sciences." Big Data & Society 7 (1): 205395172090684. https://doi.org/10.1177/2053951720906849
- Smith, B., and G. Linden. 2017. "Two Decades of Recommender Systems at Amazon.com." IEEE Internet Computing 21 (3): 12–18. https://doi.org/10.1109/MIC.2017.72
- Sørensen, J. K. 2019. "Public Service Media, Diversity and Algorithmic Recommendation: Tensions between Editorial Principles and Algorithms in European PSM Organizations." CEUR Workshop Proceedings 2554: 6–11. https://vbn.aau.dk/en/publications/public-service-media-diversit y-and-algorithmic-recommendation-ten.
- Thylstrup, N. B., K. B. Hansen, M. Flyverbom, and L. Amoore. 2022. "Politics of Data Reuse in Ma-Chine Learning Systems: Theorizing Reuse Entanglements." Big Data & Society 9 (2): 205395172211397. 20539517221139785. https://doi.org/10.1177/20539517221139785
- Tsing, A. L. 2011. Friction: An Ethnography of Global Connection. Princeton: Princeton University Press.
- Tuchman, G. 1973. "Making News by Doing Work: Routinizing the Unexpected." American Journal of Sociology 79 (1): 110-131. http://www.jstor.org/stable/2776714. https://doi.org/10.1086/225510
- van Dijck, J., T. Poell, and M. de Waal. 2018. The Platform Society: Public Values in a Connective World. Oxford: Oxford University Press.
- Vulpius, J. 2022. "The Role of Audience Data and Metrics in Becoming "Audience-Centred." In Futures of Journalism: Technology-Stimulated Evolution in the Audience-News Media Relationship, edited by V. J. E. Manninen, M. K. Niemi, and A. Ridge-Newman, 347–363. Cham: Palgrave Macmillan.
- Zelizer, B. 1992. "CNN, the Gulf War, and Journalistic Practice." Journal of Communication 42 (1): 66-81. https://doi.org/10.1111/j.1460-2466.1992.tb00769.x