AI transparency in journalism: labels for a hybrid era

By Katharina Schell

Contents

Introduction

The deployment of AI in media production has prompted a debate on the appropriate labelling of AI use in journalistic processes. While the concept of transparency is gaining traction, many practical issues remain unresolved.

Furthermore, regulatory initiatives have not yielded the clear guidelines many stakeholders had anticipated. Key questions remain unanswered:

- What factors should be considered when determining whether to label the involvement of artificial intelligence (AI) in journalistic production?
- What characteristics should the labels possess?
- What information should they convey, and
- How can the labelling process be integrated into the workflow of media production and distribution?

The aim of this project is to propose a practical framework and accompanying tools that help publishers, editorial teams and journalists make well-informed and plainly explainable decisions on whether, when and how to disclose the use of AI in their journalistic products. Contingent on these decisions, the project also sets out a range of formal options for the labelling itself.

These approaches should be as sustainable as possible. In other words: based on current use cases, the proposed solutions should keep pace with technological developments, at least for some time to come.

The discourse surrounding the implementation of artificial intelligence in journalism is unduly preoccupied with technological aspects; there is a disproportionate focus on AI, with insufficient attention directed towards the fundamental aspects of journalism. It is imperative to address this imbalance.

As my primary working concept, I use the factor of autonomy in the human-machine relationship. The thinking is as follows: the extent to which human editorial autonomy has been transferred to the machine is a crucial factor in determining whether or not to disclose the AI's contribution.

However, the integration of autonomous machine decisions into the journalistic workflow adds a new dimension to the concept of journalistic agency.

Consequently, I propose an extension of the concept of journalistic autonomy to include the aspect of "authorial autonomy". This involves describing the journalist as an author whose agency is constrained by a new actor in his or her own domain.

A note on terminology

■ What do I define as *journalism* in the context of this project?

Framing the journalist as "author" already suggests I am primarily concerned with the field of text-based news journalism, which includes readable texts (online or in print) and – to the extent to which scripted text is involved – some audio content (radio, podcasts, online).

I have chosen text-based news for a number of reasons:

1. In text journalism, the question of what is considered "AI-generated" is particularly difficult to answer.[1] Current regulatory efforts, such as the [EU AI Act](), allow a considerable degree of interpretation with regard to AI-generated or -assisted journalistic texts (in contrast to the case of visual media genres).[2] This is also reflected in the correspondingly vague wording used in various guidelines of well-known media outlets (for an overview, see [here]()).

2. The journalistic text remains the exemplar of the journalistic genre, demonstrating the typical workflow from information gathering to content creation and distribution.

As a final distillation of my definition of text-based journalism, this project considers hard news – the segment of the broad range of journalistic expression that can claim to have the most direct democratic impact. Concerns that AI in journalism could have a negative impact on the information society need to be taken particularly seriously in this sector.

■ What do I define as *Artificial Intelligence* in the context of this project?

It is necessary to narrow down the catch-all AI buzzword. In current media discourse, the term *Artificial Intelligence* is equated almost exclusively – and implicitly and incorrectly – with Generative Artificial Intelligence.[3] However, the ways in which AI is used in journalism are far more diverse.

For the purposes of this paper, I define AI as a system that makes autonomous decisions in a delimited, intellectually defined and journalistic field and thus generates or modifies a journalistic text or participates in its generation or modification.

This definition also covers the use of AI in the distribution of the resulting content product, insofar as it comes to specific publication decisions about specific journalistic works – that is, when it comes to a system's editorial decision about whether or not to publish a particular story. However, I am *not* addressing further algorithmic curation that might occur in a content distribution systems of a platform (such as a media company's website).

Conditio sine qua non!
In all of the considerations in this text, I am working from two premises:
1. Editors and media companies do not publish content for which they have used AI without thorough editorial review.
2. Deepfakes, in the sense of regulatory requirements, must be labelled in any case. These are therefore not considered part of the discussion.

The flaw in the transparency hypothesis

The debate – one might even say *moral panic* – about fake news since the mid-2010s means "transparency" is a term we are all familiar with. Hardly any medium with quality standards has avoided taking up the cause of transparency. And so what if AI comes into play? We remain transparent, we adapt this principle in a contemporary way, and that's it – right? Unfortunately, not.

For a start, research suggests we have failed to correctly interrogate the seemingly promising principle of "transparency".

"For all its popularity and its potential importance in addressing journalistic and public concerns, discussion of transparency has suffered from a lack of clarity in its definition," find Drs Kyle Heim & Stephanie Craft.[4] Meanwhile Finnish journalist Jussi Latvala reports: "I have found no agreed-upon set of practices at the core of the journalistic transparency ideal."[5] And now a new player has come along to rudely illustrate as much.

A lot has been written about the contemporary interpretation of transparency in the AI age, and the normative consensus is: "Transparency is […] a deeply held normative value that aims to guide the behaviour and conduct of journalists."[6] Yet no one seems to know how to implement the ideal in practice.

Latvala didn't have AI in mind when he challenged the dogma of journalistic transparency in 2023. He was questioning the mantra-like belief that transparency in journalism would increase user trust.

This same belief underlies much of the public reflection on AI transparency in journalism. In their guidelines, *AI Transparency in Journalism*, Swedish media publishers state: "Transparency is a means of creating trust."[7] Their document exemplifies a boom in memoranda of understanding from publishers and newsrooms keen to demonstrate their responsible and transparent use of AI.

Hardly any of these guidelines can do without the key concept of transparency, which is incidentally a symptom of the isomorphism of these texts, as noted in several recent works (more on this here).[8] At least implicitly, all the guidelines for artificial intelligence express an *a priori* distrust – trust withheld in advance of available evidence of untrustworthiness. RSF guidelines, for example, warn that AI

could become "a structural challenge to the right to information [in the sense of] the freedom to seek, receive and access reliable information."[9]

Diagnosis: Advance mistrust

The media's advance mistrust in AI is based on its fear of deception (more on this in detail on). This is understandable, given the trajectory of "transparency" as a household remedy for digital disinformation. Now AI is being added to the threat scenario as another risk factor. It is argued that the lack of transparency in AI systems could undermine journalistic accountability and make it difficult for audiences to assess the credibility of AI-generated content.

"Audiences may feel deceived if they assume an AI-generated piece was written by a human. This erodes consumer trust over time."[10] These are lines from a hyperbolic marketing blog, but they sum up the general suspicion in a conveniently crisp way: AI is something to be mistrusted. If the medium I trust uses AI without telling me, I will lose my trust.

The source of logical tension here is the ambivalent nature of the mistrust: on the one hand, the technology – the much-vaunted black box – must be taken as universally not trusted. On the other hand, the hypothesis is that the media company that uses this untransparent technology loses trust unless it does so transparently. And so the transparency hypothesis is carried over unquestioningly into the AI age.

What do we know about user mistrust of AI?

The extant research provides no compelling evidence that users exhibit a generalised mistrust of AI in journalistic production. In past years, there has been a [notable focus](#) on algorithmic recommendation systems and the distinctiveness of automated journalism.[11] [Altay & Girardi (2024)](#) find general "AI aversion" is "well-documented in the literature", yet it remains unclear whether this aversion manifests to a similar, greater, or lesser extent in relation to journalism.[12]

More recent work has focused on general attitudes to AI in journalism, looking at specific use cases and providing initial insights into user-focused transparency strategies. The 2024 edition of the Reuters Institute for the Study of Journalism's [Digital News Report](#) (DNR) looked at attitudes, specifically as a type of emotional state, in 28 countries and found [evidence of AI aversion](#): "Our qualitative research shows that people's starting point is generally one of resistance, suspicion, and fear."[13] Around [half of respondents](#) (US: 52%, Europe: 47%) are uncomfortable with the idea of news being primarily produced by AI. However, the picture is different when it comes to assisting AI, where most of the work is still done by journalists: 42% in the US and 33% in Europe are quite comfortable with this.[14]

The DNR team's further research also shows that discomfort correlates with other factors. The higher the level of "AI awareness", the "relatively more comfortable" users are with AI applications in journalism.[15] Yet awareness of AI is still low, with 49% of respondents (across all markets surveyed) saying they have heard little or nothing about AI. At the same time, socio-demographic differences are apparent (younger, male respondents tend to be more familiar with AI).[16]

Users' attitudes also vary according to subject matter: the use of AI is less welcome in hard news, such as political, crime or local reporting, than in entertainment journalism or sports.[17] The latter in particular may be due to the fact that in the first wave of news automation, sports reporting became known as a prime use case for rule-based generation.

Finally, there is evidence of use case-specific acceptance based on where in the journalistic workflow the AI was used: "Depending on whether it is used behind the scenes, to deliver news in new ways, or to generate entirely new content."[18]

Bear in mind that findings about the state of inclination are largely based on the status quo of media production and reception. Empirical studies can hardly take into account the prospect that new technologies have the potential to significantly change media consumption itself, i.e. the way users access news content.

It is probably not too far-fetched to assume that the way in which information is conveyed, not least in the form of dialogue, will have a disruptive effect. Whether conversational agents lead to a higher perceived credibility of information has already been investigated (in the study mentioned, the answer is "yes").[19] The "preference for dialogue with the machine" could lead to a shift in the way some users interact with news content, using AI not just to read the news, but to "process, interpret and personalise it in ways that meet their individual needs".[20]

A recent study suggests that users are increasingly taking a pragmatic approach to the idea of AI in journalism: "Much of the public believes that journalists are currently using generative AI to perform certain tasks[...]", such as proof-reading, headlines and "writing the text of an article".[21] In this context, a wide-ranging study of possible discrepancies between perceived use and editorial reality would be desirable. For now, it can only be assumed that users – based on this online survey in six countries – tend to overestimate the use of AI, especially generative AI, in media production, which the authors rightly and aptly describe as a thoroughly cynical assessment of journalism.

This makes the question of users' intentions all the more interesting when confronted with content that is indeed made using AI and labelled as such. "They want a choice and to be able to exercise a certain level of agency. [...] Being able to filter news that has been written by an AI or a human," according to Piasecki et al.[22]

So users are making up their minds concerning what to think about AI in journalism, which means that now could be the perfect time for a well thought-out disclosure strategy: "Participants see AI technology as an important emerging technology that will dramatically impact their lives in the future; participants had positive and negative feelings about that fact, but many expressed a willingness to change their opinion about the risks and benefits of AI as the technology evolves."[23]

Transparency rules: many words, some gaps

With advance mistrust and an audience unclear on what they think about AI, the starting point for drafting specific media guidelines is rather unfavourable.

There is a fundamental tension for media publishers concerned about the disclosure of AI: "Much of the prevailing discourse surrounding the labeling and detection of AI-generated media is grounded in fears that such content could mislead or otherwise deceive members of the public."[24] In a journalistic context, intent to deceive is an accusation that comes close to libel. "It is important to distinguish between the use of AI in editorial media compared to other types of digital platforms," the Swedish publishers also stress. [25]

In addition, the term *media* is not clearly defined: regulators, industry, platforms, publishers and journalists all use it in different contexts. The fact that key regulatory documents such as the EU AI Act do not make any clear statements about media production does not make things any easier.

AI policies and guidelines from media companies
Recently, several insightful comparative studies of AI guidelines have been published. They agree that most of these texts are very similar – too similar, according to some.[26] Here, I will examine the disclosure provisions of 13 directives, focusing in particular on their attempts to define a threshold above which the use of AI in texts must be made transparent. The approach of the texts can be roughly divided into three categories:

On the one hand, there are policies that strictly reject the use of generative AI, such as those found at *WIRED* and to some extent at the *Guardian* (although there is already a restriction and approval process in place for "exceptional use").[27,28] It can be assumed that sooner or later the guidelines in this category will be made more permissive, as has already happened in some media organisations. My employer, the Austria Presse Agentur (APA) wrote in its 2023 policy that it "does not use any text modules from generative AI in text messages", but will adjust this in the 2024 update (see below).[29]

In the second category, I include policies with generally worded transparency intentions, such as the BBC's assurance that it will ensure that users "can understand why we use [AI], how it works and how it affects them".[30] Or Reuters, who, "[…] will implement practices intended to make the use of AI in our products

and services explainable".[31] The Algemeen Nederlands Persbureay (ANP) leaves the decision to disclose to the editorial team if they "deem it appropriate".[32]

And, finally, there are texts that try to specify how strong the AI factor has to be in the end product in order for labelling to be required. For *Der Spiegel* it is "a significant role" that AI plays, while for German public broadcaster Bayerischer Rundfunk, transparency is required for "largely automated content".[33,34] "A considerable impact on journalism" is the transparency trigger for Swedish daily *Aftonbladet*.[35] The Swiss media company Ringier Group frames disclosure from the opposite view: "Labelling is not required in cases where an AI tool is used only as an aid".[36] Deutsche Presse-Agentur (dpa) limits the labelling requirement to cases "where content is generated exclusively by AI".[37] The Associated Press (AP) names specific use cases for "experimentation" with AI (translation, news summaries, headlines), but considers only machine translation to be subject to labelling requirements.[38] In its updated guidelines, which have not yet been published, APA cites the "degree of automation" as the crucial consideration.[39] The earlier-mentioned Swedish paper from the *Nordic AI Journalism* initiative does not define "significant journalistic impact", but leaves the decision to the media company, where it should take place "in the editorial process".[40]

Almost all the policies I looked at emphasise that no content modified or generated by AI should be published without editorial review. Another common feature is the absence of the audience as a target of transparency efforts. If the user is lost from view in attempts to extend the concept of transparency to AI, these efforts end up in a teleological vacuum – lacking purposeful direction.

The EU AI Act
The European Union is taking a different approach in its regulatory efforts, known as the *EU AI Act*. It establishes a framework for artificial intelligence (AI) systems, categorising them by risk level. It also emphasises transparency throughout the AI system lifecycle, encompassing the design, development, deployment, and post-market monitoring phases. Article 50, which is relevant to media professionals, explicitly takes the user perspective and focuses on preventing deception: when natural persons interact with AI systems, they must be informed, the Article states.[41]

Clause 4 of Article 50 establishes disclosure requirements for the distribution of AI-generated content. It seems simple enough: content that is entirely AI-generated,

such as images, videos or audio files, is considered to be a deep fake and must be labelled as such.

However, the second paragraph of Clause 4 is tricky from a media point of view: "Deployers of an AI system that generates or manipulates text which is published with the purpose of informing the public on matters of public interest shall disclose that the text has been artificially generated or manipulated. This obligation shall not apply [...] where the AI-generated content has undergone a process of human review or editorial control and where a natural or legal person holds editorial responsibility for the publication of the content."

The first analyses of how much room for interpretation these wordings leave for disclosure in editorial practice are available.[42] The short answer is: a great deal. On closer inspection, it is not clear how "direct interaction" is to be defined in the context of media consumption, or what text manipulation can mean in the context of media production. Moreover, would any reputable medium knowingly disseminate deep fakes?

A superficial reading suggests that the internal rules that media companies have already adopted are far stricter than the requirements of the EU regulation, in that they commit to editorial control and labelling of AI content. In my view, however, it is fair to question whether the AI Act can achieve the intention of enabling users to consume information in an informed way.[43]

Autonomy, agency and AI

Journalism sees itself as independent and strives for the greatest possible autonomy. This "invocation" may have what [Örnebring & Karlsson](#) describe as a "ritualistic quality", but it determines the idea of what role journalism fulfils and how it acquires the power to act.[44]

Their notion of [journalistic autonomy](#) as "the idea that journalism as a societal institution, as well as individual journalists in their workplace (the newsroom), should be free from undue influence from other societal institutions and actors" is challenged, however, by "questions of control, dependence and autonomy in the light of the 'AI goldrush' in the news".[45,46]

Is semi-autonomous technology yet another player entering the editorial arena and restricting journalists' freedom of action? Yes, [according to some](#): "AI introduces a non-human actor into the newsroom that simulates acts of journalism usually associated with a human, such as news writing or content moderation."[47]

The rich body of research on the tense relationship between journalistic autonomy, agency and independence can only be implicitly addressed here, so I will fall back on [Felix Simon](#)'s apt and compact definition: "a journalist's capacity to act (agency) in a self-directing manner (autonomy)".[48] I focus on the power of action in editorial decisions and relate it to key concepts in editorial guidelines such as *editorial responsibility* and *significant journalistic impact*.

Decision-making authority

Journalists consider the "level of editorial control" to be the [key indicator](#) of their autonomy.[49] I would add "the authority to make and to exercise decisions" to this definition. But, if journalists these days "have *episodic* autonomy over what they publish", the room for action seems limited.[50] And the potential of AI systems to limit "journalists' discretionary decision-making ability and therefore their agency and autonomy" is obvious.[51]

Journalism, especially in a partially formalised environment such as news journalism, has long been "[vulnerable to automation](#)", has even "had strong [elements of 'robotness'](#) for a long time".[52,53] The *[Churnalism](#)* Møller cites ("a form of routinised information packaging that lacks original ideas and instead recycles news content from other sources") is an example.[54]

What distinguishes the current paradigm shift from the history of constant digital change in media is that technology not only potentially *changes* editorial decisions, but also *makes* them. And that an *external* influence is being *internalised* by newsrooms. Not only is technology "frequently [developed or at least co-developed](#) inhouse by powerful actors within the institutional framework", but journalists themselves are often involved in implementing it in editorial workflows, sometimes in a leading role.[55] They exert "procedural influences"– and these are "generally seen as the most important (and therefore the most important factors limiting journalistic autonomy)".[56]

It is significant that the technology is openly recognised as a new player in the newsroom. However, the way we name it – for instance, calling it an "assistant" (APA's in-house GenAI application is called "TextAssistant") – speaks for itself: the player has been brought on board with journalistic sanction, but ranks below editors and reporters in the hierarchy. The new "assistants" have a "second-order" autonomy.[57]   In this way of thinking, the authority remains with the human journalist, but certain decisions are taken by the *assistant*.

But what crucial decisions are made in the prototypical day-to-day work of a newsroom? "At the level of daily news work, autonomy is understood as the extent to which journalists are free to decide story angles, what sources to use, and what narrative frames to employ."[58] A model process:

- Which story is selected/researched?
- What information and sources are used, which are discarded?
- What is the tonality, angle, format and audience of the story?
- When, how and where will the story be published?

The autonomy of authorship
What's the story? This question is asked and answered thousands of times every day in newsrooms around the world. Humans like and need to consume information in the form of narratives. Journalists were satisfying this need long before *storytelling* became a buzzword.

But there remains a curious gap in considerations of journalistic autonomy and agency – and not only in the context of AI: it is the framing of the journalist as a member of a *writing guild*. To put it simply: as an author. This is rarely considered in

journalism research or, if it is, then it is done so from the point of view of the recipient: a popular research question explores if users can distinguish between a text written by a human or machine.

The hype around generative AI as a potential threat to journalists is aimed precisely at what we will call the authorial authority and autonomy of journalists. Again and again – and unfortunately, one might say – articles are published by journalists suggesting that generative AI writes almost as well or better than a human.

My favourite example of this is an early piece of GPT reporting from the *Guardian*: "A Robot Wrote This Entire Article. Are You Scared Yet, Human?"[59] This article is from 2020, but it set the tone for public, audience-focused discourse on AI and journalism. More interestingly, by suggesting that the entire text was "written" by "a robot", it ignored and distorted the journalistic achievement of the author (in fact, the byline named "GPT-3" as the author). Someone – some human – at the *Guardian* triggered, edited, shortened and assembled several of the model's texts. Curation, editing and publication decisions: these are the core tasks of an authorial regulatory power in journalistic narrative, and even in this *scary* example, humans took on these tasks.[†]

So what is the difference between writing and telling a story? This is a question that literary scholars have been asking for centuries. Perhaps we can learn a thing or two from them. After all, as Ørmen & Gregersen say: "Journalists, like all other authors of narratives, organise their narratives by manipulating the sequence of events from a straightforward chronological sequence to a more effective one."[60]

The agency of authors is "deliberate", one that pursues certain intentions and can be held accountable for its work.[61] For Van Woudenberg, this argues against attributing characteristics of authorship to ChatGPT. But let's take this definition further in a journalistic context – for which it is perfectly suited – and assume that the journalistic author delegates parts of their intentions to the machine, which then produces on behalf of the author: the authorial autonomy would be shared, the intention and responsibility would remain with the human, but the work itself would be the task of the AI.

---

† It should be noted that the Guardian emphasised that the process "was no different to editing a human op-ed" – but that doesn't change the fact that editorial and authorial decisions produced the end result and that there was no reason to fear either way.

"Distributed authorship" is the term [Hannes Bajohr](#) uses to describe the interaction between humans and machines in an author network.[62] Rather than viewing authorship as a solely human endeavour, this model recognises the agency of non-human actors. This involves a loss of control on the part of the author. The greater the loss, Bajohr says, the greater the role of editorial choice, and: "It is therefore plausible to think of authorship as editorship."[63] But this, in turn, disturbs an understanding of autonomy based on "the extent to which journalists are free to decide on the stories they cover or edit, as well *as the selection of story angles, sources, and narrative frames"*.[64,‡]

The journalist as author must relinquish power. Coeckelbergh and Gunkel describe a "[joint-agency](#)": "Instead of regarding humans as absolute and 'authoritarian' authors, we propose [...] to regard humans, language, and technology as co-authors in the processes and performances of these generative models like ChatGPT."[65]

Yes, the time has come to dig out Roland Barthes' dictum on the *death of the author* and to remember Michel Foucault, who predicted the *obsolescence of the author*. Coeckelbergh and Gunkel continue:

> *"Though Barthes and Foucault did not address themselves to LLM and generative AI, their work on authorship expertly anticipates our current situation: One day there will be "no longer an identifiable human author at all".*

Ørmen and Gregersen suggest it is already the case:[66]

> *"The much-discussed 'death of the author' in literary studies is altogether real in certain parts of journalism, where robots are now responsible for writing up certain types of journalistic pieces".*

This is getting a bit depressing. To be clear, the *death of the author* in journalism need not mean the end of journalism, nor the demise of journalists as machine

---

‡ Emphasis author's own.

editors. The concept of co-authorship in journalism allows us to glimpse the potential scale of the disruption that the AI era may bring to journalism and, by extension, its social function.

"Retooling" is only the beginning. [67] The transformation will no longer be limited to business models for journalism, distribution channels for journalism, formats or platforms. What is being reshaped, if you will, is the "essence of journalism" and its future impact on our information systems.[68]

We are on the cusp of an age of hybrid journalism in which the actors may be human or machine.[69] We have not yet adapted our notion of journalistic autonomy.

Another note on terminology

In *hybrid journalism*, both humans and machines have agency. Their balance of power in the journalistic process can be described in terms of two areas of autonomy: editorial decision and authorial autonomy.

Disclosure decisions in the agency matrix

I recommend using the balance of power in the hybrid journalistic process as a decision criterion for or against the disclosure of AI in this process.

Here, a case-by-case approach is useful. As an application example, I illustrate below how such a framework can be used.

Table 1: Impact matrix – weighting editorial and authorial autonomy in a human-machine-hybrid news production environment

This matrix outlines the tasks involved in hybrid news production, categorising them by the role of human, machine, or hybrid decision-making and their respective impacts on editorial and authorial control.

Stages in news production	Tasks		Editorial decisions made by...	Impact on authorial autonomy
News gathering	Data exploration	Data mining and aggregation	Machine	None
		Data extraction	Hybrid	None
		Data analysis	Hybrid	Low
	Fact-checking	Claim detection	Machine	Low
		Cross-referencing	Human	Low
	News and trend detection	Feed monitoring	Hybrid	Low
		Content categorization and filtering	Machine	Mixed
	Speech-to-Text		Human	None
	Inspiration, brainstorming, ideation		Human	Mixed
Content Creation	Story automation (rule based, data-to-text)		Human	Mixed
	Story generation (GenAI)		Machine	High
	Content reformatting	Article shortening	Hybrid	Mixed
		Story summarising/abstracting	Hybrid	Mixed
		Social media posts, newsletters	Hybrid	Mixed
		Headlines	Hybrid	Mixed

Table: Continues on page 18 · Created with Datawrapper

Table 1: Impact matrix (continued)

Stages in news production	Tasks		Editorial decisions made by...	Impact on authorial autonomy
Content Creation	Proof-reading		Human	None
	Translating		Hybrid	None
	Speech-to-Text		Human	None
	SEO writing		Hybrid	Mixed
	Content tagging	Entities	Machine	None
		Topics	Machine	None
Content distribution	Automated content distribution		Machine	Mixed
	Content clustering	Recommendation and personalisation	Machine	Mixed
		Topic clustering	Machine	Low
	Semantic search		Machine	None
	Conversational interfaces		Machine	High
	Content metrics, audience analytics		Machine	Mixed

Table: Katharina Schell • Created with Datawrapper

This matrix outlines the tasks involved in hybrid news production, categorising them by the role of human, machine, or hybrid decision-making and their respective impacts on editorial and authorial control.

There are two important things that I would like to emphasise:

- The list of use cases is not exhaustive. It is intended to demonstrate how the system can be used to make a decision for or against labelling AI using the [blueprint](#) we will discuss below.

- The classification is not risk based. However, it is used to identify tasks with a high impact on autonomy, which may indicate a higher need for scrutiny to maintain ethical standards, transparency and journalistic integrity.

We need to think beyond labels

Let us assume, on the basis of what we have said so far, that disclosure of AI involvement in journalism is desirable in principle, even if it is not entirely clear how to account for the "varying levels of algorithmic intervention".[70] Let us further assume that we have defined the AI use cases that are worthy of disclosure in accordance with the impact matrix. And let us assume that this labelling is intended to empower the user to act, and that journalists do not intend to deceive their users. What should these labels look like?

The most obvious solutions are dominated by initiatives in the visual field, just as the term synthetic content is primarily associated with these genres.[71] Labelling a visual element with a visual label seems to be a simpler undertaking and more technologically feasible – as demonstrated by the Coalition for Content Provenance and Authenticity (C2PA).[72] The main motivation of these initiatives is to mitigate deception and manipulation on the one hand, and to address intellectual property issues on the other.

Caution: side effects of labelling

First, a word of caution: "labels could backfire" and several studies have demonstrated this effect.[73] Altay & Gilardi found: "Labelling headlines as AI-generated reduced the perceived accuracy of the headlines and participants' intention to share them."[74] And Longoni et al. observed user aversion to AI indicated "potentially detrimental consequence of disclosing use of generative AI, which may further exacerbate the already declining public trust in news outlets".[75]

We face a "paradox of AI disclosure": Users want to know how journalism is using AI, but if the industry fulfils this desire, it will face potential reputational costs. "Rather than being rewarded for transparency, news organisations that disclose their use of these tools are perceived as less trustworthy and may therefore have fewer incentives to be so forthcoming," according to Toff & Simon.[76]

Hasty labelling strategies can distort the view of the whole news environment. Take fake news labelling: when content was labelled as *fake*, unlabelled content received higher trust ratings. This *implied truth effect* could become an "implied authenticity effect" in the context of AI.[77]

The most unfavourable outcome would be a news environment in which journalistic articles are correctly labelled in accordance with rigorous standards, yet are perceived as being of lesser value by the audience than generic, non-journalistic and unlabelled AI content. One conclusion of academic research: "If the content is not harmful, the benefits of labelling it as AI-generated are debatable, especially given the negative effects of these labels."[78]

But we won't be able to satisfy users' need for information in this way. So what communication measures can we take to empower users to make informed choices about consuming content? And at the same time not lose their trust in journalistic content?

Procedural insights

Unsurprisingly, users do not find labels such as "manipulated", "deepfake" or "not real" particularly trustworthy, so the goal would be to describe in a more nuanced way exactly what the AI has done.[79] It is worth exploring the concept of *process-oriented* labelling. This is not limited to a *description* of the content, but makes it clear how the piece of journalism was created and at which points in the *process* the AI took over which tasks. This allows for more "neutral" tagging.[80]

However, this could end up being a lot of text if we assume that multiple AI applications may be involved in the production of an article – or even a team of AI agents running automated workflows.[81] Yet the explanation should be written in a way that is easy for users to understand: "Disclosure is meaningless if readers are unsure of how to interpret the labels."[82]

The Nordic publishers suggest the wording *created with the support of,* "to signal AI's actual impact on the content and remind the media consumer that there is an editorial process (and staff) behind the content".[83] The proposition's weakness is demonstrated by the example of labelling an AI-generated illustration. This use case in particular is so heavily influenced by AI that it can hardly be considered mere 'support'. Wouldn't users feel even more misled?

Whichever way you look at it: "Deciding which 'flavour' of transparency to provide is one of the main challenges."[84] This is echoed in policy terms, too: various AI guidelines occasionally include specific communication measures such as bylines, text boxes, endnotes or register entries – but the majority of governance texts refrain from giving specifics.

Fixation on labels does not seem very effective. There are cases where a simple visual signal or a short explanatory text may suffice. But labels cannot cope with the increasing hybridisation of editorial processes. To quote Wittenberg: "Rather than treat labelling as a one-size-fits-all solution, both scholars and practitioners should carefully consider the objectives that labelling is intended to accomplish and engineer solutions accordingly."[85]

The more complex the 'yes or no' decision becomes, the stronger the case for sustainable systems of transparency "that empower audiences to label themselves and to evaluate content manipulation, accuracy, and context". [86]

Disclosure strategies: a blueprint for procedural insight

Here, I outline a labelling concept for text-based journalism that explains the journalistic process and its steps in a transparent and comprehensible way and discloses whether AI systems were involved.

On the one hand, this avoids a stigmatising effect (in the eyes of the user) and, on the other hand, focuses on the journalistic performance and the editor as an active user of the technology. It is therefore crucial that the system is rolled out across the media company's production – not just where AI is involved.

The cornerstones are what I call *stages* of procedural disclosure: direct, indirect, specific and general. The stages can be further grouped as formal labelling and content transparency.
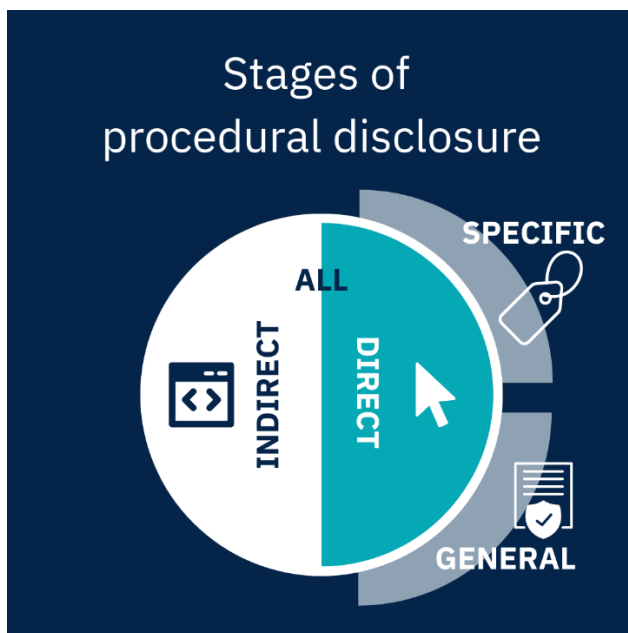
Formal labelling

■ *Direct labelling:* user orientated. This information is visible to the user; from a media perspective, direct labelling is primarily relevant for distribution on the company's own channels (such as the website).

■ *Indirect labelling*: machine-readable metadata tagging. Essential for distribution on other channels, metrics & analytics, archiving, and academic research

Content transparency

■ *General transparency*: basic information. Low-threshold access to the editorial principles of the newsroom, explained in a clear and understandable way

■ *Specific transparency:* information about the specific journalistic content in question. Description of how this piece of journalism came about (sources, authors, use of AI-tools, versions…); can call on the metadata added as part of indirect labelling

In terms of architecture, the concept is inspired by the UX recommendations of the Coalition for Content Provenance and Authenticity ([C2PA](#)), which provides several levels of information according to the *click to inspect* principle.[87]
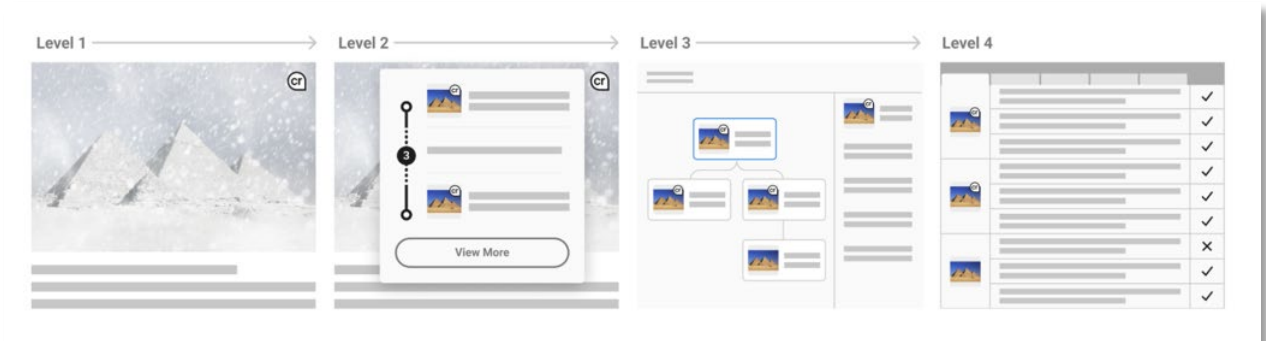

Fig. 1: Levels of Disclosure: click to inspect (Source: C2PA)

Admittedly, this approach is quite demanding on users, who have to actively retrieve the information.
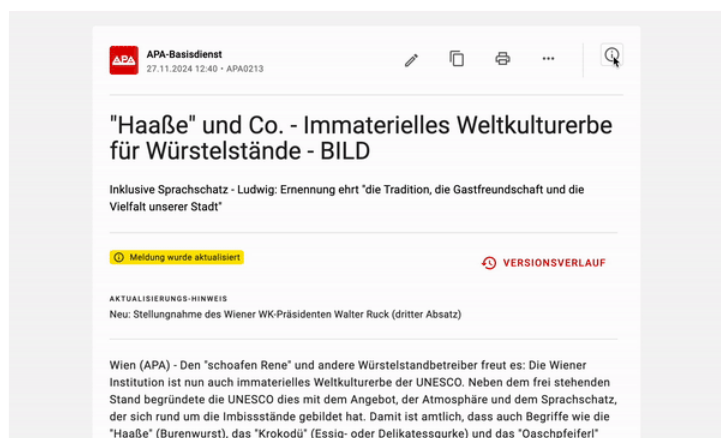

Fig. 3: Show/hide metadata: click-to-inspect in text journalism

In addition, such a system involves implementation efforts for media companies. However, the industry already has experience in providing descriptive or contextual information alongside its journalistic content.

The proposed approach would allow these strategies to be formalised. The *click to inspect* principle for accessing additional information has long been implemented in the media sector. One example is the APA news agency's metadata box (see above).

Conclusion

AI applications have taken their place in journalistic media production. But when it comes to how we communicate this, we talk too much about technology and too little about journalism. We are allowing semi-autonomous systems into our newsrooms. We let a new player in as an 'assistant' or even as a 'colleague' - but we don't think about the impact on our freedom of action.

When we outsource decisions to semi-autonomous systems, we need to be aware of the impact their decisions can have on our journalism – how it is created and how it is published. And when we use semi-autonomous systems to help us write our articles, we need to be aware that we are making them co-authors.

I have identified two crucial factors in the question of what we should disclose: Who has the editorial decision-making power? And who is the 'author' of the final journalistic product?

The framework I have developed can help media organisations and newsrooms classify the use cases in their organisations and decide for or against a particular disclosure.

So far, so good. But how do we disclose?

Disclosure strategies cannot be an end in themselves. They need to be uncompromisingly user-centric. I conclude that the seemingly straightforward approach of simple labels is problematic for several reasons.

First, we label AI participation in journalism as something exceptional – but in some use cases it is already the rule, and this will increase. At what point do such labels become so obsolete that we do away with them altogether?

Secondly, if we only label 'AI', we obscure the view of the journalistic process. Highlighting only AI would marginalise journalism as an afterthought. But we stand by our journalism – whether AI was involved or not! We should convince users of that.

Third, as long as AI is still seen as an 'exception to the norm', a simple AI label could have the opposite effect: users would trust the content less – even though we have worked in a journalistic and ethical way.

Fourth, for users with little knowledge of AI technology, or for users who are already AI-literate, such labels do not provide any added value.

In our business, transparency increasingly means explainability. I have outlined the design of a labelling system that transparently explains how we do journalism and, of course, provides information about where AI is used. The goal is for users to be able to actively learn about how we do journalism, at different levels of granularity.

It will take a lot of work to develop and implement this system. But we are facing a hybrid disruption of journalism, and there are no easy answers.

Let's stop labelling AI. Let's label journalism.

Acknowledgements

References

1 Ziv Epstein et al., 'What Label Should Be Applied to Content Produced by Generative AI?' (PsyArXiv, 28 July 2023), 2, doi:10.31234/osf.io/v4mfz.

2 Stanislaw Piasecki et al., 'AI-Generated Journalism: Do the Transparency Provisions in the AI Act Give News Readers What They Hope For?', *Internet Policy Review* 13, no. 4 (23 October 2024), Available at: https://policyreview.info/articles/analysis/ai-generated-journalism-transparency-provisions (Accessed: 30 October 2024).

3 Katharina Schell, 'KI, Medien, Journalismus: Wanderung am Autonomie-Grat', in *Jahrbuch für politische Beratung 2023/2024*, ed. Thomas Walter Köhler and Christian Mertens, edition mezzogiorno (PROverbis, 2024), 320.

4 Kyle Heim and Stephanie Craft, 'Transparency in Journalism', in *The Routledge Handbook of Mass Media Ethics*, ed. Lee Wilkins and Clifford G. Christians (Oxford: Taylor & Francis, 2020), 308.

5 Jussi Latvala, 'The Mirage in the Trust Desert: Challenging Journalism's Transparency Infatuation', 2023, 12, Available at: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2023-08/RIJS%20Fellows%20Paper_Jussi%20Latvala_TT23_FINAL.pdf.

6 Jessica Zier, '"This Article Is AI-Generated": AI Disclosure and Labeling for News Content', Medium, 1 August 2024, Available at: https://generative-ai-newsroom.com/this-article-is-ai-generated-ai-disclosure-and-labeling-for-news-content-73d6e6cb6d50 (Accessed: 21 October 2024).

7 Anne Lagercrantz et al., 'AI Transparency in Journalism', 2024, 4, Available at: https://www.nordicaijournalism.com/ai-transparency.

8 M. F. de-Lima-Santos, W. N. Yeung, and T. Dodds, 'Guiding the Way: A Comprehensive Examination of AI Guidelines in Global Media' (arXiv, 7 May 2024), doi:10.48550/arXiv.2405.04706; Kim Björn Becker, Felix M. Simon, and Christopher Crum, 'Policies in Parallel? A Comparative Study of Journalistic AI Policies in 52 Global News Organisations' (SocArXiv, 7 September 2023), doi:10.31235/osf.io/c4af9.

9 'Paris Charter on AI in Journalism', Available at: https://rsf.org/sites/default/files/medias/file/2023/11/Paris%20charter%20on%20AI%20in%20Journalism.pdf (Accessed: 18 November 2024).

10 'Transparency in AI-Generated Content: Disclosure and Labeling Practices', Upvoty, Available at: https://www.upvoty.com/blog/transparency-in-ai-generated-content-disclosure-and-labeling-practices (Accessed: 21 October 2024).

11 Benjamin Toff and Felix M. Simon, '"Or They Could Just Not Use It?": The Paradox of AI Disclosure for Audience Trust in News' (SocArXiv, 2 December 2023), 4, doi:10.31235/osf.io/mdvak.

[12] Sacha Altay and Fabrizio Gilardi, 'People Are Skeptical of Headlines Labeled as AI-Generated, Even If True or Human-Made, Because They Assume Full AI Automation', *PNAS Nexus* 3, no. 10 (1 October 2024): 9, doi:10.1093/pnasnexus/pgae403.

[13] Amy Ross Arguedas, 'Public Attitudes towards the Use of AI in Journalism', Available at: https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2024/public-attitudes-towards-use-ai-and-journalism (Accessed: 21 October 2024).

[14] Nic Newman et al., 'Reuters Institute Digital News Report 2024', 2024, 20, doi:10.60625/RISJ-VY6N-4V57.

[15] Ross Arguedas, See footnote 13

[16] Ross Arguedas, See footnote 13

[17] Newman et al., 'Reuters Institute Digital News Report 2024', See footnote 14.

[18] Ross Arguedas, See footnote 13

[19] Christine Anderl et al., 'Conversational Presentation Mode Increases Credibility Judgements during Information Search with ChatGPT', *Scientific Reports* 14, no. 1 (25 July 2024): 5, doi:10.1038/s41598-024-67829-6.

[20] Nick Hagar, 'Do People Want to Get Their News from Chatbots?', Medium, 12 September 2024, Available at: https://generative-ai-newsroom.com/do-people-want-to-get-their-news-from-chatbots-8a806495b65d (Accessed: 12 November 2024).

[21] Richard Fletcher and Rasmus Kleis Nielsen, 'What Does the Public in Six Countries Think of Generative AI in News?' (Reuters Institute for the Study of Journalism, 2024), 26, doi:10.60625/RISJ-4ZB8-CG87.

[22] Piasecki et al., 'AI-Generated Journalism', See footnote 2.

[23] 'Focus Group Insights #2: Perceptions of Artificial Intelligence Use in News and Journalism', Center for News, Technology & Innovation, 28 October 2024, 2, Available at: https://innovating.news/article/focus-group-insights-2-perceptions-of-artificial-intelligence-use-in-news-and-journalism/ (Accessed: 26 November 2024).

[24] Chloe Wittenberg et al., 'Labeling AI-Generated Content: Promises, Perils, and Future Directions', *An MIT Exploration of Generative AI*, 27 March 2024, 4, doi:10.21428/e4baedd9.0319e3a6.

[25] Lagercrantz et al., 'AI Transparency in Journalism', 4.

[26] Hannes Cools, 'Towards Guidelines for Guidelines on the Use of Generative AI in Newsrooms', Medium, 10 July 2023, Available at: https://generative-ai-newsroom.com/towards-guidelines-for-guidelines-on-the-use-of-generative-ai-in-newsrooms-55b0c2c1d960 (Accessed: 19 November 2024); de-Lima-Santos, Yeung, and Dodds, 'Guiding the Way'; 'AI Newsroom Guidelines Look Very Similar, Says a Researcher Who Studied Them. He Thinks This Is Bad News', Available at: https://reutersinstitute.politics.ox.ac.uk/news/ai-newsroom-guidelines-look-very-

similar-says-researcher-who-studied-them-he-thinks-bad-news (Accessed: 18 November 2024); Becker, Simon, and Crum, 'Policies in Parallel?'

[27] 'How WIRED Will Use Generative AI Tools', WIRED, Available at: https://www.wired.com/about/generative-ai-policy/ (Accessed: 15 November 2024).

[28] Guardian News & Media, 'Editorial Code of Practice and Guidance', Available at: https://uploads.guim.co.uk/2023/07/27/GNM_editorial_code_of_practice_and_guidance_2023.pdf (Accessed: 15 November 2024).

[29] 'Leitlinie zum Umgang mit Künstlicher Intelligenz' (APA - Austria Presse Agentur, 2023).

[30] 'BBC AI Principles', 2024, Available at: https://www.bbc.co.uk/supplying/working-with-us/ai-principles/.

[31] 'An Introduction to Artificial Intelligence at Thomson Reuters', 2024, Available at: https://www.thomsonreuters.com/en/artificial-intelligence/introduction-to-artificial-intelligence-at-thomson-reuters.html (Accessed: 15 November 2024).

[32] 'Leidraad: zo gaat de ANP-redactie om met AI', Google Docs, 2024, Available at: https://drive.google.com/file/u/0/d/1-3sAJtkOJrdIGw-gZFqYDEGQQNw13e0U/view?usp=sharing&pli=1&usp=embed_facebook (Accessed: 15 November 2024).

[33] DEV SPIEGEL, 'Künstliche Intelligenz und der Journalismus: wie wir beim SPIEGEL darüber denken', *Medium* (blog), 12 June 2023, Available at: https://devspiegel.medium.com/k%C3%BCnstliche-intelligenz-und-der-journalismus-wie-wir-beim-spiegel-dar%C3%BCber-denken-c83ee5c68965 (Accessed: 15 November 2024).

[34] 'Umgang mit Künstlicher Intelligenz: Unsere KI-Richtlinien im Bayerischen Rundfunk', 22 July 2024, Available at: https://www.br.de/extra/ai-automation-lab/ki-ethik-100.html (Accessed: 15 November 2024).

[35] 'Aftonbladets AI-policy – så förhåller vi oss till den nya tekniken', 5 April 2023, Available at: https://www.aftonbladet.se/a/76ydy9 (Accessed: 15 November 2024).

[36] 'Ringier Leverages AI Benefits While Minimising Its Risks with New Guidelines', International News Media Association (INMA), Available at: https://www.inma.org/blogs/ideas/post.cfm/ringier-leverages-ai-benefits-while-minimising-its-risks-with-new-guidelines (Accessed: 15 November 2024).

[37] Sebastian Raabe, 'Fünf dpa-Guidelines für Künstliche Intelligenz', *dpa innovation* (blog), 3 April 2023, Available at: https://innovation.dpa.com/2023/04/03/kuenstliche-intelligenz-fuenf-guidelines-der-dpa/ (Accessed: 15 November 2024).

[38] 'AP Definitive Source: Updates to Generative AI Standards', 2024, Available at: https://blog.ap.org/updates-to-generative-ai-standards (Accessed: 15 November 2024).

[39] Internal document APA

[40] Lagercrantz et al., 'AI Transparency in Journalism', 6.

41 'EU AI Act' (2024), Available at: http://data.europa.eu/eli/reg/2024/1689/oj/eng (Accessed: 30 October 2024).

42 Piasecki et al., 'AI-Generated Journalism', see footnote 2.

43 Piasecki et al., 16.

44 Henrik Örnebring and Michael Karlsson, *Journalistic Autonomy: The Genealogy of a Concept* (University of Missouri Press, 2022), 5, Available at: http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=29117511 (Accessed: 1 November 2024).

45 Henrik Örnebring and Michael Karlsson, 'Journalistic Autonomy', in *Oxford Research Encyclopedia of Communication*, ed. Test, 2019, 1, doi:10.1093/acrefore/9780190228613.013.829.

46 Felix M. Simon, 'Uneasy Bedfellows: AI in the News, Platform Companies and the Issue of Journalistic Autonomy', *Digital Journalism* 10, no. 10 (26 November 2022): 1833, doi:10.1080/21670811.2022.2063150.

47 Alfred Hermida, 'From Automata to Algorithms: A Jobs-to-Be-Done Approach to AI in Journalism', *Estudios Sobre El Mensaje Periodístico* 30, no. 3 (26 September 2024): 672, doi:10.5209/emp.97746.

48 Felix M. Simon, 'AI and Agency in Journalism', in *Unprinted. Publication Beyond the Press*, ed. Daria Kohler and Daniel Wakelin (Cambridge University Press, 2025).

49 Helle Sjøvaag, 'Journalistic Autonomy: Between Structure, Agency and Institution', *Nordicom Review* 34, no. s1 (1 December 2013): 164, doi:10.2478/nor-2013-0111.

50 Örnebring and Karlsson, 'Journalistic Autonomy', 2019, 232. Emphasis is author's own.

51 Simon, 'AI and Agency in Journalism', 7.

52 Lynge Asbjørn Møller, Morten Skovsgaard, and Claes de Vreese, 'Reinforce, Readjust, Reclaim: How Artificial Intelligence Impacts Journalism's Professional Claim', *Journalism*, 30 July 2024, 6, doi:10.1177/14648849241269300.

53 Örnebring and Karlsson, *Journalistic Autonomy*, 2022, 265.

54 Møller, Skovsgaard, and de Vreese, 'Reinforce, Readjust, Reclaim', 8. See footnote 52.

55 Örnebring and Karlsson, *Journalistic Autonomy*, 2022, 260.

56 Örnebring and Karlsson, 'Journalistic Autonomy', 2019, 10.

57 Örnebring and Karlsson, *Journalistic Autonomy*, 2022, 258.

58 Sjøvaag, 'Journalistic Autonomy', 158.

59 'A Robot Wrote This Entire Article. Are You Scared yet, Human?', *The Guardian*, 8 September 2020, sec. Opinion, Available at: https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3 (Accessed: 25 November 2024).

60 Jacob Ørmen and Andreas Gregersen, 'News as Narratives' (Oxford University Press, 25 February 2019), 8, doi:10.1093/acrefore/9780190228613.013.908.

[61] René Van Woudenberg, Chris Ranalli, and Daniel Bracker, 'Authorship and ChatGPT: A Conservative View', *Philosophy & Technology* 37, no. 1 (March 2024): 3, doi:10.1007/s13347-024-00715-1.

[62] Hannes Bajohr, 'Writing at a Distance: Notes on Authorship and Artificial Intelligence', *German Studies Review* 47, no. 2 (May 2024): 325, doi:10.1353/gsr.2024.a927862.

[63] Bajohr, 326.

[64] Zvi Reich and Thomas Hanitzsch, 'Determinants of Journalists' Professional Autonomy: Individual and National Level Factors Matter More Than Organizational Ones', *Mass Communication and Society* 16, no. 1 (1 January 2013): 135, doi:10.1080/15205436.2012.669002.

[65] Mark Coeckelbergh and David J. Gunkel, 'ChatGPT: Deconstructing the Debate and Moving It Forward', *AI & SOCIETY* 39, no. 5 (1 October 2024): 2223, doi:10.1007/s00146-023-01710-4.

[66] Ørmen and Gregersen, 'News as Narratives', 17.

[67] Felix M. Simon, 'Artificial Intelligence in the News: How AI Retools, Rationalizes, and Reshapes Journalism and the Public Arena', 2024, doi:10.7916/ncm5-3v06.

[68] Møller, Skovsgaard, and de Vreese, 'Reinforce, Readjust, Reclaim', 11.

[69] Hermida, 'From Automata to Algorithms', 673.

[70] Epstein et al., 'What Label Should Be Applied to Content Produced by Generative AI?', 2.

[71] 'PAI's Responsible Practices for Synthetic Media', Partnership on AI - Synthetic Media, Available at: https://syntheticmedia.partnershiponai.org/ (Accessed: 21 October 2024).

[72] 'Overview - C2PA', Available at: https://c2pa.org/ (Accessed: 24 November 2024).

[73] Zier, '"This Article Is AI-Generated"'. See footnote 6.

[74] Altay and Gilardi, 'People Are Skeptical of Headlines Labeled as AI-Generated, Even If True or Human-Made, Because They Assume Full AI Automation', 9.

[75] Chiara Longoni et al., 'News from Generative Artificial Intelligence Is Believed Less' (OSF, 16 February 2021), 14, doi:10.31234/osf.io/wgy9e.

[76] Toff and Simon, '"Or They Could Just Not Use It?', 20.

[77] Wittenberg et al., 'Labeling AI-Generated Content', 9.

[78] Altay and Gilardi, 'People Are Skeptical of Headlines Labeled as AI-Generated, Even If True or Human-Made, Because They Assume Full AI Automation', 10.

[79] Epstein et al., 'What Label Should Be Applied to Content Produced by Generative AI?', 5.

[80] Wittenberg et al., 'Labeling AI-Generated Content', 6.

[81] Joris Veerbeek, 'How Teams of AI Agents Could Provide Valuable Leads For Investigative Data Journalism', Medium, 24 September 2024, Available at: https://generative-ai-newsroom.com/how-teams-of-ai-agents-could-provide-

valuable-leads-for-investigative-data-journalism-ac48ece1fdab (Accessed: 17 November 2024).

[82] Zier, '"This Article Is AI-Generated"'.

[83] Lagercrantz et al., 'AI Transparency in Journalism', 12.

[84] Zier, '"This Article Is AI-Generated"'.

[85] Chloe Wittenberg et al., 'Labeling AI-Generated Media Online' (PsyArXiv, 28 June 2024), 13, doi:10.31234/osf.io/b238p.

[86] Claire Leibowicz, 'Lawmakers Push for AI Labels, But Ensuring Media Accuracy Is No Easy Task', Tech Policy Press, 9 September 2024, Available at: https://techpolicy.press/lawmakers-push-for-ai-labels-but-ensuring-media-accuracy-is-no-easy-task (Accessed: 21 October 2024).

[87] https://c2pa.org/specifications/specifications/1.4/ux/UX_Recommendations.html