Epistemological Foundations for Ethical AI in Journalism

Laurence Dierickx, Carl-Gustav Lindén
University of Bergen, June 2023

10th IEEE Swiss Conference on Data Science

AI-systems in journalism

Used in 50% of newsrooms worldwide

Survey World Association of News Publishers, WAN-IFRA, May 2023

News detection	News discoveries	Social media analysis	Audio visual search engines
News verification	Automated fact-checking	Data and content analysis	Text summarisation
Audio transcription	Machine translation	Text-to-speech applications	Data-to-text generation
Multimedia creation	News recommenders	News personalisation	...

How to blend AI systems with journalism ethics?
How to ethically use AI systems?

Ethics in journalism and AI

Respecting the truth
A matter of social responsibility

Practices in journalism and AI

Accuracy, Objectivity, Transparency
A matter of epistemology

Accuracy, an indicator for quality and accountability

Ethical standards shared in all Codes of Ethics in Europe

Broader scope: truth, objectivity, fairness, transparency, credibility. Refers to each stage of the process.

Presenting facts as they are (subjective factors).
In practice: verifying facts, ensuring sources' reliability.

Not quantifiable, assessment implies human judgement

In DDJ: data quality dimension (abnormal values, duplicates, consistency, level of comprehensibility, etc.). Prioritizing the quality of input data collection & curation

Accuracy in Data and Computer Science

Grounded in ethics: system's fairness & accountability

A measure to assess model performances (not the only one), consider statistical patterns learned, evaluating accuracy through the prism of the ground truth makes predictions pointless

Explainability methods used to assess the accuracy or plausibility of ML predictions without requiring the ground truth: model-agnostic techniques, counterfactual explanations,...

Accuracy alone cannot guarantee the trustworthiness of the model or its correct and complete description

ACCURACY
Challenges in ML

Using data sets collected from Wikipedia (not free from political bias, unknown users' expertise)

Classifiers: unbalanced or irregularities in classification. E.g., in automated fact-checking: "True", "False", "Partially true", "Contradiction", "Compatible", ...

Dataset labelling methods and crowdsourcing (expertise, human judgements)

Accuracy: Exacerbated challenges with GAI

Sources gathered on the web, including Wikipedia, user-generated content, political and religious-oriented pages
(Source: Washington Post)

Lack of accuracy in the outputs and so-called hallucinations (errors, glitches, junks)

False and harmful statements about persons

Reassessing the myth of objectivity

Constitutive of the professional self-perception, occupational norm, a part of journalism culture

Impossible objectivity (human biased), but explicitly mentioned in several codes of ethics as a synonym for impartiality or a lever for trust

In DDJ, no doubts about the “objectivity” to work with numbers (obscure the subjectivity of the analysis and the possibility of using biased data)

In fact-checking, objectivity through transparency (“Opening the black box of journalism”)

Objectivity in Data & Computer Science

Aura of objectivity of algorithmic processes, opposed to journalist' subjectivity as considered mechanically reliable, precise, credible

Computer code fulfils an editorial function: results from human decisions, implies human judgements and choices

Incomplete or biased datasets cannot be considered objective (need for assessing data quality), but some algorithms can lead to false positive and negative, or amplify bias

Even readers considered AI-generated content more neutral or objective (rule-based systems)

Objectivity: Exacerbated challenges with GAI

Biases in the data collection (sources)

Biases in the annotation process (questions the expertise of poorly paid workers, socio-cultural background)

Sources influence the model and the values encompassed in the model

Oriented answers reported: unhinged and erroneous responses or made-up outputs.

Challenges about service accuracy and reliability

The limits of transparency

"Black box" of journalism: editorial choices rarely explained

Transparency: substitute for truth and fairness, the new objectivity

Infuses DDJ and fact-checking practices (exposing sources and methods)

Ethical standards in the US, in EU: Belgium

A lever for trust and accountability (ideal?)

All cannot be explained (editorial guidelines, meetings, …)

Disclosure of the nature of the author (automation)

Transparency in Data & Computer Science

Computer code and algorithmic processes not understandable by all + proprietary protected + restricted for complex models

Not necessary to assess reliability, no guarantee of the trustworthiness of the results

Explainability as an alternative to making more sense to the user (decision-making process, technical process)

AI reasoning to consider social values and moral/ethics to understand how a given system reasons

Informing how algorithms are designed and implemented does not require transparency

Interpretability: a passive feature that allows users to understand the model and make sense of it

Transparency: Exacerbated challenges with GAI

Machine learning data often lacks transparency, explainability, or interpretability (need for documentation, critical for systems relying on vast amounts of data)

Large language models: no transparency on their data sources, labelling and training processes

Opacity challenges the trustability of the outcomes
Does opacity alone can explain the lack of trust in news content?

Fairness, a means for accountability

In journalism, honesty, privacy, professional distance, truth, and impartiality: providing well-balanced information to avoid harmful bias and stereotypes. A path to objectivity, if not equated with objectivity

In ML, absence of bias and equal treatment. Biases can be related to the training data or the implicit human values of people involved in the programming/labelling.
Can be measured through accuracy, recall and precision or human evaluation (explaining the how and the why; explaining the training data and the process at work

Consequentialist ethics (related to human decision)

Building interdisciplinarity

- Social approach of technology that considers all the stakeholders
- A common epistemological approach to blend AI with journalistic values (ethics by design)
- Computational + journalistic thinking = dialogues, exchanges (data and AI literacy in journalism)
- Favouring a mutual understanding a lever for building trust
- Deontological lenses are the limits because ethics is first a matter of practice (beyond theory)

Thank you!