

OPINIO JURIS

in Comparatione

Studies in Comparative and National Law

Vol. 1, n. 1/2018

Automated journalism and freedom of information:
Ethical and juridical problems related to AI in the press field

Matteo Monti

Automated journalism and freedom of information: ethical and juridical problems related to AI in the press field

Matteo Monti*

ABSTRACT

Journalism and the Press have always been deeply influenced by technological changes, and so they are in the digital world: from the competition of new media and the challenges of the Web 2.0 to the creation of a new way to produce news, i.e. automated journalism. Between the different notions of the use of AI in the Press field (automated journalism, robot journalism, News-Writing Bots, algorithmic journalism) in this paper the wording “automated journalism” is preferred as long as it seems to describe in a better way the practice of this type of journalism and it seems more used by the scholars who have studied this topic. Automated journalism is the use of AI, i.e. software or algorithms, in order to automatically generate news stories without any contribution of human beings, apart from that of programmers who (eventually) have developed the algorithm.

This paper aims to analyse the ethical and juridical problems of automated journalism, in particular, looking at the freedom of information and focusing on the issue of liability and responsibility. From a legal point of view, the analysis shall embrace and share the European concept of the freedom of information and media regulation, focusing in particular on the Italian legal system. Indeed in the range of European legal systems, the Italian system has more broadly developed the idea of freedom of information, and it has multiple approaches to the topic, which are partially explored here.

* PhD. Candidate, Scuola Superiore Sant'Anna.

The first paragraph of the paper shall explore the field of the media outputs in which automated journalism – as currently developed – could produce innovations and could be implemented. The utilization of the Italian model serves to understand how the pieces of automated journalism could be framed from a legal point of view.

The second paragraph shall analyse the legal and ethical problem of automated journalism by looking at the problems of liability and data use. As a consequence, a first section shall be dedicated to the issue of liability and another one to that of data utilization.

In the final remarks, some solutions and guidelines shall be proposed looking at the problems highlighted in the paper.

KEY WORDS

Automated Journalism – Freedom Of Information – Artificial Intelligence – Algorithm – Press- Journalism

Table of contents

1. Introduction: What is Automated Journalism?
2. Reasonable and current use of AI in the Press Field: The framework of Automated Journalism
3. Legal problems and ethical issues related to Automated Journalism
 - 3.1. Is Automated Journalism protected as free speech, and who is liable for it?
 - 3.2. Problems and ethical issues linked to Automated Journalism
4. Final Remarks

1. Introduction: What is Automated Journalism?

Journalism and the Press have always been deeply influenced by technological changes¹. Nowadays, they compete against new forms of media in the digital world² and face the challenges of Web 2.0 developments, including *automated journalism*. Although various terms are used to describe the use of AI in the Press field (e.g., automated journalism, robot journalism, news-writing bots, and algorithmic journalism), in this paper, the term

¹ R. Parry, *The Ascent of Media: from Gilgamesh to Google via Gutenberg*, London-Boston, 2011.

² Cfr. H. Örnebring, R. Ferrer Conill, *Outsourcing newswork*, in T. Witschge, C.W. Anderson, D. Domingo, A. Hermida (eds.), *The Sage Handbook of digital journalism*, London, 2016; M. Powers, *In Forms That are Familiar and Yet-to-be Invented: American Journalism and the Discourse of Technologically Specific Work*, in *Journal of Communication Inquiry*, 36, 1, 2012.

‘automated journalism’ is used because it seems most descriptive of the practice of this type of journalism and is often used in the research literature on this topic³.

Automated journalism is a component of *post-industrial journalism*⁴, which is a term used to describe the technological challenges affecting journalism, and involves the use of AI (e.g., software or algorithms) to automatically generate news stories without any input from humans, except for the programmer(s) who developed the algorithm. An AI algorithm independently collects and analyses data and then writes a news article. Automated journalism is based on *natural language generation* (NLG) technology, which permits, generally, the creation of text-based journalism from a dataset of digitally structured data: “Early examples of the use of NLG technology to automate journalism are mostly confined to relatively short texts in limited domains, but are nonetheless impressive in terms of both quality and quantity. The text produced is generally indistinguishable from a text written by human writers and the number of text documents generated substantially exceeds what is possible from manual editorial processes”⁵. Automated journalism operates by either independently writing and publishing news articles without input from a journalist or by ‘cooperating’ with a journalist who can be deputized to supervise the process or provide input to improve the article⁶.

AI, in the contemporary sense of the term, was first used in a newsroom by *The New York Times (NYT)* in a project named ‘Editor’, which involved applying tags to traditionally-written news articles. Another early but more sophisticated use of AI in news writing was *The Washington Post (WP)*⁷ utilizing Heliograf software to cover the 2016 Olympic Games in Rio; the software collected data related to the events schedule, results, and medal tallies. Since then, *WP* has begun to cover financial news and local sports events via automated journalism⁸. According to journalist Joe Kadhane⁹, *WP* incorporated AI to cover simple local stories because it reduces costs and could broaden its audience and increase its market share.

Currently, many other new producers, such as *The Associated Press*, *Forbes*, *Los Angeles Times*, and *ProPublica*, use automated journalism¹⁰, which is dependent upon access to

³ K. Dörr, *Mapping the Field of Algorithmic Journalism*, in *Digital Journalism*, 4, 6, 2016.

⁴ C.W. Anderson, E. Bell, C. Shirky. *Post-industrial Journalism: Adapting to the Present*, New York, 2012.

⁵ D.Caswell, K. Dörr, *Automated Journalism 2.0: Eventdriven narratives*, in *Journalism Practice*, 2017, p. 2.

⁶ This last issue could be considered a form of computational journalism. See J.T. Hamilton, F. Turner, *Accountability through algorithm: Developing the field of computational journalism*, Report from the Center for Advanced Study in the Behavioral Sciences, Stanford, 2009, available at <https://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf>.

⁷ WashPost PR Blog, *The Washington Post experiments with automated storytelling to help power 2016 Rio Olympics coverage*, in *washingtonpost.com*, 5 ago 2016.

⁸ J. Kadhane, *What News-Writing Bots Mean for the Future of Journalism*, in *Wired*, 16 feb 2017.

⁹ Id.

¹⁰ A. Graefe, *Guide to Automated Journalism*, in *Columbia University Academic Commons*, 2016.

and the availability of structured data to generate news articles. The general advantages of this method are the speed with which data can be collected and articles can be written, fewer errors in the output, and cost savings¹¹. However, the quality of automated journalism depends on the data it uses and often cannot introduce new issues, and it is currently unable to develop an in-depth critical analysis of the phenomena described. Indeed, technically speaking, the main problem of current AI articles is their low quality in terms of the narrative and critical considerations.

Legally and ethically speaking¹², it is clear that the main point of this revolution is the changes it brings about in media institutions (i.e., media outlets) and, above all, the rules of professional journalism¹³. Media institutions are commonly described as composed of regulative (i.e., rules and regulatory processes), normative (i.e., the link between social values and goals), and cultural-cognitive (i.e., the sharing mechanism) features¹⁴. Some scholars, such as Katzenbach¹⁵ and Napoli¹⁶, claim that media technologies should be analysed through the lens of institutional theory. Accordingly, this framework will be adopted in this essay to analyse the ethical and juridical problems of automated journalism, especially as it relates to freedom of information and the press. From a legal viewpoint, the analysis shall embrace and share the European concept of freedom of information¹⁷ and media regulation while focusing on the Italian legal system. Of all the European legal systems, the Italian system has more broadly developed the idea of freedom of information, and it has multiple approaches to the topic, which are partially explored here¹⁸. As a consequence, the Italian legal system can be considered as a prototype case¹⁹ in relation to the European paradigm of the freedom of information.

In the next section, the paper will explore the field of media outputs in which automated journalism, as currently developed, could produce innovations and how these innovations could be implemented, focusing on the Italian categorization of journalism outputs. Then,

¹¹ See A. Graefe, cit. Cf. S.C. Lewis, O. Westlund, *Big data and journalism: Epistemology, expertise, economics, and ethics*, in *Digital Journalism*, 2015.

¹² See in general S.J.A. Ward, *Global Journalism Ethics*, Montreal, 2010.

¹³ On the deinstitutionalization see: P.M. Napoli, *Navigating producer-consumer convergence: Media policy priorities in the era of user-generated and user-distributed content*, in *Communications & Convergence Review*, 1(1), 2009.

¹⁴ W.R. Scott, *Institutions and organizations: Ideas and interests*, Los Angeles, 3rd ed., 2008, p. 52 and ff.

¹⁵ C. Katzenbach, *Technologies as institutions: Rethinking the role of technology in media governance constellations*, in M. Puppis, M. Just (Eds.), *Trends in communication policy research*, Bristol, 2011.

¹⁶ P.M. Napoli, *Automated Media: An Institutional Theory Perspective on Algorithmic Media Production and Consumption*, in *Communication Theory*, 24 (3), 2014.

¹⁷ See for instance art. 20 of the Spanish Constitution and art. 5 of the German one; Art. 11 of the Charter of Fundamental Rights of the European Union and Art. 10 of European Convention of Human Rights.

¹⁸ For an overview in English: F. Casarosa, E. Brogi, *The Role of Courts in Protecting the Freedom of Expression in Italy*, in E. Psychogiopoulou (ed.), *Media Policies Revisited*, London, 2014, 101. For an analysis of the implication of the Italian constitutional principles on the Internet platforms: M. Monti, *Perspectives on the Regulation of Search Engine Algorithms and Social Networks: The Necessity of Protecting the Freedom of Information*, in *Opinio Juris In Comparatione*, 1, 1, 2017.

¹⁹ R. Hirschl, *Comparative Matters: The Renaissance of Comparative Constitutional Law*, Oxford, 2014, 256 and ff.

in section 2, legal and ethical problems related to automated journalism will be analysed by examining the problems of liability and data use. Consequently, one subsection will be dedicated to the issue of liability and another to data utilization. In the concluding remarks, some solutions and guidelines will be proposed for the problems highlighted in the paper.

2. Reasonable and current use of AI in the Press Field: The framework of Automated Journalism

The ‘disruptive innovation’²⁰ of automated journalism has led to ‘automation anxiety’²¹, which is partially unjustified considering AI’s limited effect on the press field. Additionally, the fear of technology is unjustified if it is utilized in sectors where it can operate without dangerous implications. To understand the best field of application of automated journalism, it could be useful to consider the paradigm of freedom of information and its *species* in the Italian legal system as an example²². Freedom of information is regulated quite differently in various legal systems across the world. In Italy, its regulation is well developed and articulated, and, as in the European field, it is based on active and passive features: the freedom to inform and the right to be informed²³.

While examining case law in the Italian legal system, it is possible to identify some *species* of freedom of the press and information²⁴. The first category is the so-called *right of chronicle*, which is the right to merely report facts about something that has happened. This right is related to a journalist’s commitment to and controls over sources of information and, therefore, on the accuracy and truthfulness of the events narrated²⁵. It is possible,

²⁰ C.M. Christensen, *The Innovator’s Dilemma*, Harvard, 1997.

²¹ D. Akst, *Automation Anxiety*, in *The Wilson Quarterly Summer*, 2013.

²² P. Costanzo, *Informazione nel diritto Costituzionale*, in *Dig. Pubbl. VIII*, Torino, 1993, 319, 323. The Italian legal system is peculiar because it has enshrined in laws journalist rules of the profession and the code of ethics (law no. 69/1963), and it has formalized a sort of corporation of journalists that has been considered consistent with the Constitution by the Constitutional Court, which has highlighted the specific role played by journalists in democratic systems (Constitutional Court decision no. 11/1968. Cf. Constitutional Court decision no. 98/1968).

²³ See *ex pluribus* Constitutional Court decision no. 348/1990. Italian scholarship, independently from adhesion to an individualistic theory of free speech (i.e., a sort of American vision) or a functionalist theory of it (i.e., a vision more similar to the German idea of free speech), recognizes that the output of journalism to be considered protected as free speech has to be based on the diffusion of true facts, or – importantly – that subjectivity false reporting is not free speech. See P. Barile, *Il soggetto privato nella Costituzione Italiana*, Padova, 1953, 121; C. Esposito, *La libertà di manifestazione del pensiero nell’ordinamento italiano*, Milan, 1958, 37; A. Pace, *Commentario della Costituzione. Art. 21*, Bologna-Roma, 2006, 89; S. Fois, *Principi costituzionali e libera manifestazione del pensiero*, Milan, 1957, 210-211.

²⁴ Most of the reported Criminal Supreme Court’s (Cass. pen. – *Corte di Cassazione Penale*) caselaw arises from the defamation matter.

²⁵ *Ex pluribus* Criminal Supreme Court (Cass. pen.) decision of the 7/7/1987. Cf. Criminal Supreme Court (Cass. pen.) decision no. 41249/2012.

in any case, that the facts are combined with a critique of the events described or with some comments made by the journalist. The second category is *right of critique* (i.e., the right to criticize), which means freedom to critique an idea, an event, or other aspects of society. In judicial applications²⁶, this right requires minor attention to the truthfulness of events that are commented on and critiqued. *Right of critique* could be considered a sort of political pamphlet made by a journalist. The third category is *investigative reports*, namely pieces of journalism that better correspond to the Press' role as the watchdog of democracy. It consists of exposing theories about facts that are not wholly-verifiable. According to the courts²⁷, in this case, the check on the sources of information is less needed. The field in which automated journalism could operate and be useful in increasing the quality of news seems to be *right of chronicle*. Indeed, in this framework, automated journalism could be a way to improve and make *right of chronicle* more objective by collecting data and reporting them without human input or distortion. It could potentially be the most 'pure' or a perfect form of *right of chronicle*. From the audience's viewpoint, it seems that automated journalism could completely substitute journalists' input in this field²⁸. *The Los Angeles Times' Homicide Report* is a good example²⁹ of how automated journalism could avoid bias in media coverage of murders by reporting all types of murders and informing public opinion with correct data. The project started as a blog in 2008 and was then re-launched in 2010, retrieving homicide data directly from the Los Angeles County Coroner's Office. News stories could then be expanded with details provided by journalists: "crime reporters used the automatically generated stories as initial leads for exploring a particular case in more detail, for example by adding information about the victim's life and family"³⁰. A useful evolution of this system would be the ability to add some details about family stories collected from the public register and other sources.

Other examples of implementing *right of chronicle* are *The Los Angeles Times' Quakebot*, which involves earthquake reporting, and *AP's Wordsmith* platform, which concerns corporate earnings stories. Sometimes, the narration is more difficult due to specific elements of the data. Other times, the Wordsmith software is too simplistic to face the topic, as in the case of reporting local governments' activities³¹.

²⁶ *Ex pluribus* Criminal Supreme Court (Cass. pen.) decision no. 15236/2005. And for the issue of the check on the facts: Criminal Supreme Court (Cass. pen.) decision of the 16/4/1993.

²⁷ *Ex pluribus* Criminal Supreme Court (Cass. pen.) decision no. 9337/2012.

²⁸ An experiment has demonstrated that "In sum, the available evidence suggests that the quality of automated news is competitive with that of human journalists for routine tasks". M. Haim, A. Graefe, *Automated News*, in *Digital Journalism*, 2017, 13.

²⁹ N. Lemelshtrich Latar, *The Robot Journalist in the Age of Social Physics: The End of Human Journalism?*, in G. Einav (ed), *The New World of Transitioned Media*, New York, 2015, 74.

³⁰ A. Graefe, cit., 22.

³¹ D. Caswell, K. Dörr, *Automated Journalism 2.0*, cit., 11 and ff.

Thus, it seems that automated journalism is currently limited to *right of chronicle*³² reporting because software and technology are not sufficiently advanced to improve or substitute a human being's critique³³. As a consequence "These limitations, and others, ensure that manual writing will remain the only viable method for producing the most complex, impactful, and valuable journalism for the foreseeable future. Nonetheless (...) it is indeed possible to encode many journalistic events and stories as data, and thereby automate the writing of news that is more complex than routine sports and finance reporting"³⁴. Actually and currently, "Most uses of robot journalism have been for fairly formulaic situations"³⁵. Indeed, even if automated journalism can find new correlations in the data, it cannot explain the reasons for or consequences of these correlations. Although human reasoning is still necessary for non-*right of chronicle reporting*, in the near future³⁶, AI could be able to provide critique if programmers give them additional abilities, such as fact-checking political statements or simulating emotions related to unfair decisions³⁷.

In conclusion, in situations where data mining is impossible³⁸, human journalists are still needed. Additionally, automated journalism could relieve journalists from the task of covering the most straightforward events of public interest and allow them to focus on more complex events or *right of critique*-type reporting³⁹. However, it should be noted that automated journalism should not be considered a form of neutral journalism because the editor still chooses which stories will be published⁴⁰. Editors and publishers could introduce bias by choosing to focus on publishing news about crimes perpetrated by immigrants if they want to attack immigrants or news about problems in the financial market if they want to emphasize problems caused by capitalism.

³² Perhaps, also in some fields of investigative journalism. Cf. S. Parasia, *Data-Driven Revelation: Epistemological Tensions in Investigative Journalism in the Age of 'Big Data'*, in *Digital Journalism*, 2014.

³³ Cfr. K. Dörr, *Mapping the Field of Algorithmic Journalism*, cit.

³⁴ D. Caswell, K. Dörr, *Automated Journalism 2.0*, cit., 16.

³⁵ T. Kent, *An ethical checklist for robot journalism*, in *medium*, Feb 24, 2015 (Updated march 2016).

³⁶ For more information regarding the possible evolution of NGL technology, see D. Caswell, K. Dörr, *Automated Journalism 2.0*, cit.

³⁷ For a very futuristic picture of the journalist's work, see F. Marconi, A. Siegman, *A day in the life of a journalist in 2027: Reporting meets AI*, in *www.cjr.org*, April, 11, 2017.

³⁸ "Despite all the buzz about Big Data and the possibilities of mining huge sets of information, availability of data is actually one of the barriers to increased automation". C-G. Linden, *Decades of Automation in the Newsroom*, in *Digital Journalism*, 5, 2, 2017, 132.

³⁹ A. van Dalen, *The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists*, in *Journalism Practice*, 6 (5-6), 2012.

⁴⁰ Graefe, cit., 16.

3. Legal problems and ethical issues related to Automated Journalism

3.1. Is Automated Journalism protected as free speech, and who is liable for it?

From a legal viewpoint, the two main issues concerning automated journalism are: the legal status of automated journalism and the problem of liability. The two topics are strictly connected since the issue of AI's speech protection has consequences that impact liability. The first issue is the *vexata questio* of an algorithm's protection⁴¹. In the United States, for example, many scholars have debated whether the First Amendment protects algorithms' output. Summarizing the position that can be certainly endorsed, it could be said that the output of an algorithm must be considered protected speech as long as it has a content message⁴². From this perspective, the protection of algorithms could seem legitimate, but the message produced by an algorithm could be illegal. This leads to the second issue: Who is the author of this speech? Alternatively, and perhaps more importantly, who is liable for an algorithm's speech?

Indeed, it could be possible to distinguish the speaker or author of the speech (the Algorithm) from the subject liable for the speech: It could be said, following the aforementioned scholars, that the outputs of an algorithm could be considered free speech⁴³ despite the lack of human intervention, but liability should be identified. In a broader sense, the issue of responsibility seems to be central in the AI field: "Responsibility is therefore essential, in view of what sort of AI we develop, how we use it, and whether we share with everyone its advantages"⁴⁴. The main problem of AI is, legally speaking, liability for the actions of robots or AI⁴⁵ and, in particular, the specific matter of imputation and liability for automated journalism raises different concerns.

⁴¹ See *ex multis*: O. Bracha, F. Pasquale, *Federal search commission – Access, fairness, and accountability in the law of search*, in *Cornell Law Review*, 93, 2008; J. Bambauer, *Is data speech?*, in *Stanford Law Review*, 66, 2014; T. Wu, *Machine speech*, in *University of Pennsylvania Law Review*, 161, 2013; S.M. Benjamin, *Algorithms and speech*, in *University of Pennsylvania Law Review*, 161, 2013; E. Volokh, D. Falk, *Google: First amendment protection for search engine search results*, in *Journal of Law, Economics & Policy*, 8, 2012.

⁴² It is a sort of vessel "for the ideas of a speaker, or whose content has been consciously curated" (Wu, cit., 1498). See Benjamin, cit.

⁴³ Consider the so-called Baidu doctrine: *Jian Zhang v. Baidu.Com Inc*, 10 F. Supp. 3d 433 (2014).

⁴⁴ J. Cowsls, L. Floridi, *Prolegomena to a White Paper on an Ethical Framework for a Good AI Society (June 19, 2018)*. Available at SSRN: <https://ssrn.com/abstract=3198732> or <http://dx.doi.org/10.2139/ssrn.3198732>, 3.

⁴⁵ *Ex multis* between criminal law and private law: P.M. Asaro, *Robots and Responsibility from a legal perspective*, in *Proceedings of the IEEE*, 2007; U. Pagallo, *The Laws of Robots: Crimes, Contracts, and Torts*, Springer, 2013; G. Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems*, Springer, 2015.

Scholars⁴⁶ who have focused on the matter of automated journalism have used a shortcut to solve the problem of liability: the editor's or the fact-checker's responsibility⁴⁷. Indeed, in the Press field, it is quite easy to find individuals who are deputized to check the content of published news, and they are usually held responsible along with the author of the piece of journalism. If the writer of an article is an algorithm, which obviously cannot be charged for defamation, for example, only editors or fact-checkers remain, so all the liability will be placed on them. This position could also be compatible with the theory of an autonomous subjective liability of robots as proposed by Hallevy, who presented a complex twist of models of criminal liability concerning the actions of robots⁴⁸. Hallevy's *perpetration-via-another model*⁴⁹ individuates a robot as an innocent instrument in the hands of an evil programmer or user who decides to commit a malicious crime, while in the *natural-probable-consequence liability model*⁵⁰, liability is connected to the programmer or the user's negligence, and the *direct liability model*⁵¹ considers the AI or robot to have committed a crime independently.

In the automated Press field, as stated previously, the use of AI is not yet developed well enough to create critical journalism or decide what to generate and publish⁵²; for this reason, just the first two categories of models of liability can be met. 'Perpetration via AI' seems the most applicable because if an editor or a programmer decides to alter or use AI to create fake news or defame someone (among other nefarious purposes), his or her crime and the consequent liability seem quite evident. However, if the AI's inaccurate or

⁴⁶ L. Weeks, *Media Law and Copyright Implications of Automated Journalism*, in *N.Y.U. J. Intell. Prop. & Ent. L.*, 4, 81; S.C. Lewis, A. Kristin Sanders, C. Carmody, *Libel by Algorithm? Automated Journalism and the Threat of Legal Liability*, in *Journalism & Mass Communication Quarterly*, 2018.

⁴⁷ The only way in which a programmer could be considered guilty of defamation with actual malice would be when there is a proof that: "human programmers had a 'high degree of awareness' of false statements rather than interrogating the awareness of an algorithm. To do so, the plaintiff would need to show that the programmer knew, or should have known, that the algorithm would produce false statements that would be harmful to an individual's reputation. Such a showing could occur if an algorithm were intentionally programmed to develop and produce false content". S.C. Lewis, A. Kristin Sanders, C. Carmody, cit., 9.

⁴⁸ G. Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems*, Springer, 2015; Id., *When Robots kill*, Northeastern University Press, 2013; Id., *The Criminal Liability of Artificial Intelligence Entities: from Science Fiction to Legal Social Control*, in *Akron Intellectual Property Journal*, 4, 2, 2010.

⁴⁹ G. Hallevy, *The Criminal Liability of Artificial Intelligence Entities: from Science Fiction to Legal Social Control*, in *Akron Intellectual Property Journal*, 4, 2, 2010, 179 and ff.

⁵⁰ Id., 181 and ff.

⁵¹ Id., 186 and ff.

⁵² In this case, the fine could be directly imposed on the editor, and the AI could be corrected (i.e., modified) to prevent further crimes. The programmer's liability could be less present since "Overall, the (delegated) power of algorithms is rising. A high degree of complexity in the cooperation between algorithmic agents and humans results in low transparency (not only for users, as in the mass media, but also for producers), controllability, and predictability compared to reality construction by traditional mass media. Agency and accountability problems become more important (Chopra and White, 2011) as well as the moral significance (Verbeek, 2014) of algorithms. Even programmers and software engineers increasingly do not know what 'their' algorithmic selection produces (Auerbach, 2015)". N. Just, M. Latzer, *Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet*, in *Media, Culture & Society*, 2016, 253.

libellous output is due to the negligence of programmers or editors ('natural-probable-consequence liability model'), the issue seems more complicated and related to the need for the development of some standard of due diligence.

If the lack of a human writer shifts responsibility to the editor, as in the case of an anonymous article, are programmers or engineers liable for their AI's output? In the literature (and laws), an analysis of automated journalism programmers' liability seems to be absent despite programmers being the controllers and writers of the variables in an algorithm. If it is true that AI could, in some way, act (or, in the future, learn) without human intervention, it does not seem like a current issue in automated journalism because AI's independent decisions are limited by the programmed data choices and instructions to assemble the data. Consequently, would it be possible to discharge the liability on the programmers? Editors or fact-checkers can examine certain aspects of the output, but they cannot check all the aspects of the algorithm and surely not the technical process from which the algorithm produces the news.

If there are significant errors in the programming, the algorithm could ignore pre-determined data, which would distort the output, and the editor and fact-checker would not have the technical skills to recognize it. Editors and fact checkers are probably unable to understand the algorithm's code and trust their engineers and programmers to develop good algorithms⁵³. Indeed, although NLG technology is evolving, making the code more similar to an editorial task characterized by computational thinking than a code for programmers, in any case, it is "a skill set that is not yet a common component of journalism education"⁵⁴. As a consequence, the issue of the programmer's responsibility cannot be ignored as we will see in the next section.

In conclusion, in the field of automated journalism, liability should also be determined for programmers in cases of negligence or actual malice. Some forms of liability should be established by legislators or, at least, developed by the Courts. Implementing and enforcing laws or creating *ad hoc* laws seems to be an urgent and crucial need.

Currently, due to the lack of *ad hoc* laws and rules for automated journalism, the concern arises regarding data use and whether some ethical problems could be resolved by media companies without the intervention of legislators. The types of questions the next section will focus on are those concerning data utilization and the problem of *bad data*⁵⁵.

⁵³ "In Algorithmic Journalism these principles are embedded within code, with journalists and coders working together to fit the product to individual and organizational ethical standards" (K.N. Dörr, K. Hollnbuchner, *Ethical Challenges of Algorithmic Journalism*, in *Digital Journalism*, 2016, 6). From this viewpoint, data analysis and programming are increasingly important skills (V. Mayer-Schonberger, K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, London, 2013), but currently, it is the individual journalist's responsibility to learn these skills (S.C. Lewis, N. Usher, *Code, Collaboration, and the Future of Journalism: A Case Study of the Hacks/Hackers Global Network*, in *Digital Journalism*, 2, 3, 2014).

⁵⁴ D.Caswell, K. Dörr, *Automated Journalism 2.0*, cit., 4.

⁵⁵ Q. E. McCallum, *Bad Data Handbook*, Cambridge, 2012.

3.2. Problems and ethical issues linked to Automated Journalism

Ethical best practices, as well as ethical problems, are often considered by scholars and journalists as important factors through which new technologies utilized by the media can be analysed to improve the work of journalists⁵⁶. Starting with ‘traditional’ mass media ethics⁵⁷, this section will analyse new challenges introduced by automated journalism and the legal framework of *right of chronicle* (i.e., the main field in which automated journalism is applied). The advantage of automated journalism could be improvements to the legal category of *right of chronicle*, which is the main area in which objectivity is essential. The accuracy and objectivity of the news based on facts could be increased by the use of algorithms that are able to produce an article directly from a set of data. Indeed, Graefe⁵⁸ reports that Lou Ferrara, a former vice president and managing editor for entertainment, sports, and interactive media at *The Associated Press*, describe the benefits in term of accuracy derived by automated journalism: “The automated reports almost never have grammatical or misspelling errors, (...) and the errors that do remain are due to mistakes in the source data”⁵⁹. Still, as stressed, automated journalism currently only works well “for fact-based stories for which clean, structured, and reliable data are available”⁶⁰. Automated journalism could also reduce costs and allow journalists to dedicate more time to more important matters and, above all, critique events and even the facts reported by the same automated journalism.

However, automated journalism also creates many ethical dilemmas. Setting aside the use of automated journalism in creating fake news⁶¹, the central ethical problems linked to AI’s creation of news articles can be analysed. The main issue of the current use of AI in writing articles is the quality and correctness of the data used. Concerning this topic, the problems that could be highlighted are linked to the data used to generate an article, especially the identity (i.e., the source) of the data and their publicity, the accuracy of data, and the need to respect its integrity by avoiding manipulation.

The first issue concerns the identity of the data used or, more precisely, the transparency of data sources. If protecting sources of information is fundamental to traditional journalism, what is the ethically most desirable solution for data sources in automated journalism? As stressed by Dörr&Hollnbuchner “it is questionable whether source protection is possible or even desired as service providers and their journalistic clients should disclose

⁵⁶ S.J. Ward, *The Invention of Journalism Ethics. the Path to Objectivity and beyond*, Montreal, 2006.

⁵⁷ D.S. Horner, *Understanding Media Ethics*, Brighton, 2013. Using a holistic approach as suggested by M. Ananny, *Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timelines*, in *Science, Technology, & Human Values*, 41, 2015.

⁵⁸ Graefe, cit., 23.

⁵⁹ Ibidem.

⁶⁰ Id., 15.

⁶¹ W. Knight, *Fake news 2.0: personalized, optimized, and even harder to stop*, in *MIT Technology Review*, 27 mar 2018.

all data sources in terms of data transparency”⁶². Indeed, it could be ethically desirable to make readers conscious of data sources used by AI to produce an article. Using data from a ‘political’ source or an independent/public authority could be acceptable even if – probably – the parameters used by a political actor and an independent one to collect these data are different⁶³. The most important thing is that the reader is aware of the source of the dataset used by AI. From this viewpoint, it has to be stressed that a dataset used by AI could be a public and open source or public but based on the necessity to request access⁶⁴, and sometimes, data utilization could be illegal or barely legal⁶⁵. Using these terms, the data source could sometimes be similar to an informant’s identity.

However, to be consistent with the principle of transparency, given the risk of abuse of AI, the data source should always be made clear. Being transparent about the provenience of the dataset could make discovering the ‘informant identity’ of the data (or the chink in the security system)⁶⁶ easier, but the accountability of the news transmitted seems to be more important as long as it informs the public debate and the political arena of facts, and the risks of abuse in AI journalism are high in term of misinformation and disinformation. Furthermore, it is also undeniable that factual information could come from a politically oriented source⁶⁷. The consequence, from an ethical point of view, should be that as the audience has the right to know the political position of a newspaper, people also have a similar right to know the sources of data used by automated journalism, as well as their political orientation.

The second issue regards the ‘quality’ of the data employed (i.e., the accuracy and correctness of data from which the article is generated). “The reason for the lower error rate is that algorithms don’t make typos or arithmetic miscalculations,” said AP’s global business editor, Lisa Gibbs. “The errors are generally because of a problem with the data. If the data’s bad, you get a bad story”⁶⁸. Regarding this problem, an ethical duty should be the necessity to only use correct, objective, and accurate data: “automation works particu-

⁶² K.N. Dörr, K.Hollnbuchner, *Ethical Challenges of Algorithmic Journalism*, in *Digital Journalism*, 2016, 9.

⁶³ Cfr. M. Schudson, *Political Observatories, Databases & News in the Emerging Ecology of Public Information*, in *Daedalus*, 139 (2), 2010.

⁶⁴ “An American data journalist explained that getting interesting and useful data is often the result of long struggles based on Freedom of Information Act (FOIA) requests”. C-G. Linden, *Decades of Automation in the Newsroom*, cit., 132.

⁶⁵ For example, collaboration with hackers to detect financial scandals, from which a daily news automated journalism outputs could be generated: P. Bradshaw, *Data Journalism*, in L. Zion, D. Craig, *Ethics for Digital Journalists: Emerging Best Practices*, New York, 2014, 207.

⁶⁶ Cf. Id., 212.

⁶⁷ As stressed: “But factual information can be molded by the provider to its liking. Imagine if political campaigns began to offer data feeds of candidate speeches – location of speech, size of crowd, main points, key quote, etc. Even if a news company’s algorithm added background information on the candidate, poll numbers, etc., would we feel comfortable basing a news story on what the campaign considered the most significant things he said? How would a story like this be different from a press release?”. T. Kent, cit.

⁶⁸ F. Marconi, *A guide for newsrooms in the age of smart machines*, cit., 18.

larly well in domains such as finance, sports, or weather, where data providers make sure that the underlying data are accurate and reliable. Needless to say, automation cannot be applied to domains where no data are available. Automation is challenging in situations where data quality is poor⁶⁹. Hence, choosing a dataset should be guided by the ethical principle of accuracy because choosing a reliable source, especially if it is politically oriented, is a central ethical point. Consider a dataset produced by a company or collected by a political organization. These sources could also be used if they are ‘politically’ oriented⁷⁰, but regardless the fact they only collect data they want or need (e.g., collecting just the criminal records of immigrants, the advantages of the commercialization of a product, or the aptitude test results of a private school), the data’s accuracy must be undoubted. The programmer/editor/journalist must check the accuracy and correctness of the data⁷¹. Of course, this verification must also be conducted if the dataset is managed by the State or a public authority⁷².

On the contrary, the consequence of using unreliable data could be the diffusion of fake news, such as *The Los Angeles Times*’s Quakebot reporting an earthquake that did not happen or inaccurate reports in the financial field due to an erroneous reading of the data, such as the error concerning Netflix’s second-quarter earnings in 2015⁷³. Searching for the most accurate data available and attempting to cross-check different datasets could be an excellent way to avoid making embarrassing mistakes or spreading fake news.

Additionally, automated journalism could be used only when the data are trustworthy and predictable events are involved; regardless, an ethical duty should be monitoring the process of producing and publishing the output of automated journalism. This last aspect leads us to the third ethical principle that should be at the base of automated journalism: Monitoring by a fact-checker or an editor could avoid the inevitable errors that a machine might make. The absence⁷⁴ of monitoring and validating⁷⁵ the produced output could be one of the most dangerous aspects of automated journalism. Conscientious monitoring and validating could eliminate or minimize errors due to, for example, unpredictable events⁷⁶ or misleading data: “The accuracy of these supervised learning systems is, of course, important. The two most common errors in this sort of machine learning are terms that we

⁶⁹ Graefe, cit., 17.

⁷⁰ Of course, as said before, readers have to be informed about the sources of the dataset and its political orientation.

⁷¹ It is very important to verify data before using it and should be an ethical duty of every journalist. P. Bradshaw, *Data Journalism*, in L. Zion, D. Craig, *Ethics for Digital Journalists: Emerging Best Practices*, New York, 2014, 203.

⁷² See, for instance, the mistake made by *The Texas Tribune*: Id., 204.

⁷³ Graefe, cit., 24 and 25.

⁷⁴ Regarding the lack of a monitoring process, see Dörr, Hollnbuchner, cit., 10. The newsroom should have it as claimed by T. Kent, cit.

⁷⁵ N.L. Latar, *The Robot Journalist in the Age of Social Physics: The End of Human Journalism?*, in G. Einav (ed.), *The New World of Transitioned Media*, New York, 2015.

⁷⁶ Regarding unpredictability, see Graefe, cit., 24.

borrow from statisticians – Type I (false negative) and Type II (false positive) errors”⁷⁷. A verification process for AI-generated output should be mandatory in the Press field, which would make it possible to consider fact-checkers or editors liable for negligence.

The fourth ethical issue is linked to the problem of data distortion due to either someone modifying the results in bad faith⁷⁸ or the existence of bias in the AI algorithm. Journalism ethics has already dealt with this first aspect: the creation of fake news, as well as the distortion of facts and data, is prohibited⁷⁹. For this reason, the solution is quite banal and straightforward: banning this type of conduct.

The second one is, instead, quite innovative as bias in the AI could influence the reading of data and, as a consequence, the correctness and accuracy of the news article. “Dan Keyserling, head of communications at Jigsaw, a technology incubator created by Google, explains the overarching concern – that algorithms are prone to bias, just like humans: «We need to treat numbers with the same kind of care that we would treat facts in a story,» Keyserling said. «They need to be checked, they need to be qualified and their context needs to be understood»”⁸⁰.

Linked to the problem of bias, there is also the problem of how ethical values⁸¹ of journalism can be coded into an algorithm. These two correlated aspects lead to the necessity of a dual ethical solution: granting the accountability of the algorithm while making the programmers ethically bounded. Thus, the first issue could be resolved by making the algorithm accountable so that it becomes not only transparent but also understandable⁸², which would make identifying bias easier⁸³. Reputational pressure on the Press should then force the newspaper to resolve and eliminate possible bias by discovering it and

⁷⁷ F. Marconi, *A guide for newsrooms in the age of smart machines*, available at the website <https://insights.ap.org/uploads/images/the-future-of-augmented-journalism_ap-report.pdf>, 2017, 9.

⁷⁸ AI could be a very dangerous tool in the wrong hands. AI could be a very dangerous tool in the wrong hands: cf. M. Taddeo, *The limits of deterrence theory in cyberspace*, in *Philos. Technol.*, 2017. T. King, N. Aggarwal, M. Taddeo, and L. Floridi, (2018, May, 22), *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*. available at the website <SSRN: <https://ssrn.com/abstract=3183238>>.

⁷⁹ See T. LAITILA, *Journalistic Codes of Ethics in Europe*, in *European Journal Of Communication*, 1995.

⁸⁰ F. Marconi, *A guide for newsrooms in the age of smart machines*, available at the website <https://insights.ap.org/uploads/images/the-future-of-augmented-journalism_ap-report.pdf>, 2017, 3.

⁸¹ F. Kraemer, Felicitas, K. van Overveld, M. Peterson, *Is There an Ethics of Algorithms?*, in *Ethics and Information Technology*, 13 (3), 2011. Cf. M. Del Campo, A. Fure, W. McGee, S. Manninger, A. Flexer, *Autonomous Tectonics – A Research into Emergent Robotics Construction Methods*, in F.Scheurer, J. Nembrini, A. Kilian, C. Gengnagel, *Rethinking Prototyping: Proceedings of the Design Modelling*, Berlin, 2013.

⁸² Cf. M. Coddington, *Clarifying Journalism’s Quantitative Turn: A Typology for Evaluating Data Journalism, Computational Journalism, and Computer-Assisted Reporting*, in *Digital Journalism*, 3 (3), 2015. More broadly see M. Turilli, L. Floridi, *The Ethics of Information Transparency*, in *Ethics and Information Technology*, 11 (2), 2009.

⁸³ It would be possible, for instance, to develop an AI that can discover bias or analyse an AI program to verify which variables it uses. Cf. A. Caliskan, J.J.Bryson, A. Narayanan, *Semantics derived automatically from language corpora contain human biases*, in *Science*, 356 (6334), 2017, 183-186.

aiming “to move toward a model of algorithm ethics by asking when, how, and for whom NIAs work”⁸⁴.

The second issue should be overcome by also applying a code of ethics to programmers⁸⁵ who will be part of the new technological world of journalism. Indeed, with the advent of automated journalism, they will become proto-journalists: “a further challenge will be to develop and codify ethical guidelines building an ethical background for non-journalistic actors involved. This also applies to media organizations as they have to develop and adopt ethical codes of conduct for Algorithmic Journalism”⁸⁶. Thus, the correctness of automated journalism is possible as long as the algorithm at the base of the AI that ‘composes’ (i.e., writes) the article is free from any ideological bias and programmers do not distort the data. Given the increasingly relevant role of engineers in the Press field, it is necessary to think about forms of ethical responsibility and the legal liability of programmers⁸⁷. The idea that the programmers are free of any bias is naïve or, at least, too optimistic⁸⁸.

Finally, there is the problem of attributing the news generated by algorithms. Usually, in traditional journalism, when a piece of news is anonymous, it means it was created by the whole board of the newspaper and is attributed to the whole team (also, the liability remains just on the editor). In automated journalism, an issue arises concerning the necessity to highlight that the writer is AI and not a human being⁸⁹. Ethically speaking, identifying whether a human being or a machine is the writer of a piece of news seems to be necessary so the reader will be aware when reading an article and allow him or her the freedom to choose between traditional journalism and automated journalism.

All these ethical issues and, most importantly, good practices⁹⁰ that avoid the diffusion of low-quality journalism could be enshrined in an ethical code or, even better, in a code of conduct based on the model of those required by the *EU General Data Protection Regulation* (art. 40). Although this solution could regulate automated journalism without the ‘plaster effect’ of a law, guaranteeing the possibility of modifying the code as AI continues

⁸⁴ M. Ananny, *Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness*, in *Science, Technology, & Human Values*, 2015, 7.

⁸⁵ According to some editors, they could be considered the true authors of the news. T. Montal, T., Z. Reich, *I, robot. You, journalist. Who is the author?*, in *Digital Journalism*, 5, 2016, 13.

⁸⁶ K.N. Dörr, K.Hollnbuchner, *Ethical Challenges of Algorithmic Journalism*, cit., 11.

⁸⁷ “New leaders can be expected to run the newsrooms – they will be the data silo managers and software writing engineers. Arthur Sulzberger, publisher of the NYT, was recently asked what he would do today in his media organization, given his experience: Arthur Sulzberger surprised some people recently when asked what he would do differently in the digital transition, given hindsight. Hire more engineers, he said”. Latar, cit., 75. All of those involved in automated journalism should probably be subjected to a code of ethics. See S.C. Lewis, O. Westlund, *Actors, Actants, Audiences, and Activities in Cross-Media News Work: A Matrix and a Research Agenda*, in *Digital Journalism*, 2014.

⁸⁸ Cfr. D. Lazer, R. Kennedy, G. King, A.Vespignani et al., *The Parable of Google Flu: Traps in Big Data Analysis*, in *Science*, 6176, 2014.

⁸⁹ Cfr. T. Montal, T., Z. Reich, cit., 13. K. Dorr, *Mapping the Field of Algorithmic Journalism*, in *Digital Journalism*, 2015.

⁹⁰ D. Craig, *Journalism Ethics And Best Practices*, L.Zion, D. Craig (eds.), *In Ethics for Digital Journalists: Emerging Best Practices*, New York, Routledge, 2014.

to evolve⁹¹, it could bind both programmers and journalists. In the complex mix of approaches guaranteeing rights and democracy on the Internet (e.g., digital governance, digital ethics, or digital regulation⁹²), the utilization of a code of conduct, at least in Western democracies⁹³, could be the Aristotelian golden mean to safeguard technological evolution and progress, as well as integrate automated journalism in the Press field as a watchdog of democracy. The accuracy of news diffused through automated journalism is necessary for being framed as news output and for being consistent with the passive feature of freedom of information as contained in the Italian and European paradigms of the press and freedom of information (i.e., the right to be informed).

4. Final Remarks

The issue “is not whether data, computers, and algorithms can be used by journalists in the public interest, but rather how, when, where, why, and by whom”⁹⁴. Automated journalism could be a very effective way to improve the quality and accuracy of news related to *right of chronicle*, but some ethical and legal rules need to be established. From this perspective, enshrining the aforementioned ethical principles in a code of conduct could assist judges in applying laws yet in force, interpreting them to apply to AI, and understand the respect of good practices in trials concerning liability.

De lege ferenda, the issue of liability could be resolved by assigning responsibility to the editors⁹⁵, who can develop tools for monitoring the results of automated journalism. For example, they could apply a monitoring process or check the sources of data to avoid being deemed negligent in a defamation case⁹⁶. The application of criminal and civil liability for abuse and negligence in the use of automated journalism could be the best instrument

⁹¹ F. Pizzetti, *La protezione dei dati personali e la sfida dell'Intelligenza Artificiale*, in Id. (ed.), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Torino, 2018, 181.

⁹² “On the governance of the digital, there is much to be said, and even more still to be understood and theorised, but one point is clear: the governance of the digital (henceforth digital governance), the ethics of the digital (henceforth digital ethics, also known as computer, information or data ethics (Floridi and Taddeo 2016)) and the regulation of the digital (henceforth digital regulation) are different normative approaches, complementary, but not to be confused with each other, in the following sense”. L. Floridi, *Soft Ethics and the Governance of the Digital*, in *Philosophy & Technology*, 2018, 3.

⁹³ This solution could also accommodate the doubts of Floridi. Id., 6 and ff.

⁹⁴ A.B. Howard, *The Art and Science of Data-driven Journalism*, New York: Tow Center for Digital Journalism, Columbia University, 2014, 4.

⁹⁵ In the United States, Section 230 of the Communications Decency Act cannot act as a shield for automated journalism (Weeks, S.C. Lewis, A. Kristin Sanders, C. Carmody, cit., p. 12); the same could be said in the Italian legal system, where an online newspaper was considered a common newspaper (for additional information regarding warranty in the criminal trial, see Criminal Supreme Court [Cass. pen., Sez. Un.], decision no. 31022/2015) and under the European Union's legislation (Directive 2000/31/EC of the European Parliament and the Council of 8 June 2000).

⁹⁶ Cf. S.C. Lewis, A. Kristin Sanders, C. Carmody, cit., 10.

to guarantee the accuracy of automated journalism, imposing a process of monitoring and fact-checking algorithms' outputs. This seems to be the easiest way to grant the respect of ethical principles, at least until the 'offenses' committed are checkable by a non-programmer. In this perspective, some *ad hoc* laws could be necessary to assign liability also to programmers.

Given as a pacific statement the unavoidable political nature of choosing which news stories to publish, automated journalism raises many questions concerning the future of the Press and journalism, but above all, it involves many ethical questions. The main problems are related to the way data are utilized, but they could be solved with the application of the best practices and some *ad hoc* law in the field of liability. The aforementioned ethical principles could, hence, be enshrined in an ethical code that should be enforced on workers involved in new journalism technologies and, preferably, linked to innovative law regarding the liability of programmers. Of course, some form of self-regulation of the category of techno-journalists, including programmers, such as the *IFJ Code of Principles*⁹⁷ or the *Charter of Poynter*⁹⁸, would also be desirable as a complementary tool. To sum up, with these rules, automated journalism could bring benefits to the world of journalism, but it cannot substitute for journalists. The role of human beings in journalism will not disappear because the kind of critique needed to read the facts currently cannot be performed by a machine. No AI can currently make the governments accountable for their choices. "No robot journalist can become a guardian of democracy and human rights. It is therefore extremely important that human journalists should understand the dramatic developments in their professions and make sure these changes serve them in ways that will preserve and strengthen their very important social function"⁹⁹.

⁹⁷ IFJ Declaration of Principles on the Conduct of Journalists, available at <<http://www.ifj.org/about-ifj/ifj-code-of-principles/>>.

⁹⁸ Poynter Code of Principles, available at <<https://ifcncodeofprinciples.poynter.org>>.

⁹⁹ Latar, cit., 78.

