ALINA TRAPOVA[*] / PÉTER MEZEI[**]

Robojournalism – A Copyright Study on the Use of Artificial Intelligence in the European News Industry

The copyright protectability of outputs generated by, or with the help of Artificial Intelligence (AI) is a hotly debated question in academia and by many institutions. In practice, sophisticated AI algorithms have become a meaningful assistant in the European news industry for the reporting of sports (Retresco's collaboration with the German Football Association), weather (textOmatic's collaboration with FOCUS Online) and finance (the Guardian's 'Guarbot'). Furthermore, for the first time in copyright history a court in China assessed the validity of a company's copyright claim over the articles produced by the corporation's algorithm. The protection with copyright of this 'robojournalism' is no longer just a buzzwordy trend.

From a technological perspective, robojournalism currently relies on assistive, generative and distributive technologies. The first two seem to be the most problematic from a copyright perspective as they challenge the well-rooted human authorship requirement. Experts have been able to agree so far that it does not look like AI technology is going to be a disruptive force in the media industry. However, researching the impact of AI in journalism matters a great deal. There are numerous benefits stemming from the use of AI in the newsroom – from expanding news coverage, through faster content production, all the way to leaving journalists more time for the more 'creative' and investigative tasks where the algorithm remains weak.

This paper addresses, first, the protectability of the outputs of robojournalism under the existing European Union copyright laws. It then goes on to introduce findings related to the practical significance of robojournalism in the European news industry. Here, our focus is on the business, media, and communications studies' perspectives of automated journalism. Our results demonstrate that the extent to which European journalism relies on assistive and generative technologies to produce written output does not justify, from a copyright perspective, the changing of the current anthropocentric copyright system. These findings have wider implications as AI-generated outputs have prompted many to talk about market failure if copyright (or related rights) protection was to be refused for such works.[1]

[*] Assistant Professor in Law and Autonomous Systems, The University of Nottingham, UK. Email: alina.trapova@nottingham.ac.uk.

[**] Associate Professor, University of Szeged, Faculty of Law and Political Sciences, Institute of Comparative Law and Legal Theory. Adjunct Professor (dosentti) of the University of Turku (Finland). Member of the European Copyright Society. Email: mezei.peter@szte.hu.

[1] Parts of the paper were presented at the SLS 2021 Annual Conference and IP After AI Conference 2021. The authors are grateful for the comments of Daniela Simone, Tanya Aplin, Joe Atkinson, Dilan Thampapillai and Jeremie Clos. We are grateful for the excellent support of our research assistant Anushka Tanwar in completing this research.

I. Introduction

Sophisticated Artificial Intelligence (AI) algorithms have become a significant assistant in the European news industry. Going beyond mere computer-assisted reporting, otherwise known as CAR,[2] algorithms are nowadays extensively used in the reporting of sports,[3] weather,[4] or finance.[5] The list of examples from the United States, Australia or China is equally broad.[6] In the US, RADAR – with significant human intervention – creates automated news reports.[7] Automated Insights' and Narrative Science's algorithm reports about sports events.[8] Quakebot, developed by the Los Angeles Times, gives information about earthquakes in California.[9, 10] Furthermore, for the first time in copyright history – albeit in China – a court had to assess the validity of a company's copyright claim over the articles produced by the corporation's algorithm.[11]

---

[2] Bruce Garrison, *Computer-Assisted Reporting* (2nd ed, L Erlbaum Associates 1998).

[3] Compare to Retresco's collaboration with the German Football Association. See 'How the Bundesliga Is Using AI to Increase Brand Reach' (*SportsPro*, 3 March 2020) <https://www.sportspromedia.com/opinions/bundesliga-ai-dfl-deltatre/> accessed 9 February 2022.

[4] Compare to textOmatic's collaboration with FOCUS Online. See 'TextOmatic Und Focus Online Gehen Premium-Partnerschaft Ein' (*Textomat*, 16 March 2018) <https://www.textomat.net/News/detail.205.html> accessed 9 February 2022.

[5] Compare to the Guardian's 'Guarbot'. See Aisha Gani and Leila Haddou, 'Could Robots Be the Journalists of the Future?' *The Guardian* (London, 16 March 2014) <https://www.theguardian.com/media/shortcuts/2014/mar/16/could-robots-be-journalist-of-future> accessed 9 February 2022.

[6] Andreas Graefe, 'Guide to Automated Journalism' (Tow Center for Digital Journalism, Columbia University 2016) 20-22 <https://academiccommons.columbia.edu/doi/10.7916/D80G3XDJ> accessed 9 February 2022.

[7] See <https://pa.media/radar/>. See further Florian De Rouck, 'Moral Rights & AI Environments: The Unique Bond between Intelligent Agents and Their Creations' [2019] GRUR Int 432, 433-34.

[8] Stephen Beckett, 'Robo-Journalism: How a Computer Describes a Sports Match' (*BBC News*, 11 September 2015) <https://www.bbc.com/news/technology-34204052> accessed 9 February 2022; Robert Denicola, 'Ex Machina: Copyright Protection for Computer-Generated Works' (2016) 69 Rutgers University Law Review 251, 257-59; Victor M Palace, 'What If Artificial Intelligence Wrote This? Artificial Intelligence and Copyright Law' (2019) 71 Florida Law Review 217, 224-25.

[9] Will Oremus, 'The First News Report on the L.A. Earthquake Was Written by a Robot' (*Slate*, 17 March 2014) <https://slate.com/technology/2014/03/quakebot-los-angeles-times-robot-journalist-writes-article-on-la-earthquake.html> accessed 9 February 2022; Bruce Boyden, 'Emergent Works' (2016) 39 Colum. JL & Arts 377, 380-81; Denicola (n 8) 257.

[10] Oremus (n 9). See further Boyden (n 9) 380-381; Denicola (n 8) 257.

[11] Decision of the People's Court of Nanshan (District of Shenzhen), 24 December 2019 – Case No (2019) Yue 0305 Min Chu No 14010', (2020) 51 IIC 652 – *Tencent Dreamwriter*, where the Court argued that direct connection (or causal link) existed between the editorial team's creative choices and the final output of the applied algorithm. The selection, judgement and skills of the editorial team's members and the above-the-minimum level of creativity of the outputs ultimately allow for the protection of the news reports by copyright for the benefit of the publisher (the employer of the editors); compare to Li Yan, 'Court Rules AI-Written Article Has Copyright' (*ECNS*, 9 January 2020) <http://www.ecns.cn/news/2020-01-09/detail-ifzsqcrm6562963.shtml> accessed 9 February 2022; Rory O'Neill, 'AI-Written Articles Are Copyright-Protected, Rules Chinese Court' (*World IP Review*, 10 January 2020) <https://www.worldipreview.com/news/ai-written-articles-are-copyright-protected-rules-chinese-court-

This selected list of examples highlights that the topic of robojournalism is no longer just a trend built on buzzwords. Algorithmic or automated content creation seems to be an irreversible part of the data-driven economy.[12] Indeed, 'the computerisation and algorithmisation of news and news work is increasingly becoming the norm'.[13] Journalism cannot evade the consequences of the 'computational'[14] or 'quantitative turn',[15] which necessitates a holistic approach within law, technology and also within media and communications studies.

One of the central issues in this respect is the copyright protectability of outputs generated by, or with the help of AI. This has given rise to masses of academic research, consultations on multiple fora – both nationally and internationally,[16] and various institutional reports.[17] This literature focuses on the authorship and originality issues which underlie copyright protectability. The discussion has pivoted around the ability of the human author to express free and creative choices in the algorithmic process.

From a technological perspective, robojournalism[18] currently relies on assistive, generative, and distributive technologies.[19] The first two seem to be the most problematic from a copyright perspective as they challenge the well-rooted human authorship requirement. While experts agree that it does not look like AI technology is going to be a disruptive force in the media industry, researching the impact of AI in journalism is important. With the help of AI, data collection and processing, news coverage could expand exponentially. From a business perspective, solutions are

---

19102> accessed 9 February 2022; for a detailed analysis of AI under Chinese copyright law see He Tianxiang, 'The Sentimental Fools and the Fictitious Authors: Rethinking the Copyright Issues of AI-Generated Contents in China' [2019] Asia Pacific Law Review 184.

[12] Michael Latzer and others, 'The Economics of Algorithmic Selection on the Internet' in Johannes M Bauer and Michael Latzer (eds), *Handbook on the Economics of the Internet* (Edward Elgar Publishing 2016) 396-397.

[13] Tania Bucher, 'Machines Don't Have Instincts': Articulating the Computational in Journalism' [2017] New Media & Society 918, 920.

[14] David M Berry, 'The Computational Turn: Thinking about the Digital Humanities' [2011] Culture Machine 1.

[15] Caitlin Petre, 'A Quantitative Turn in Journalism?' (*Tow Center for Digital Journalism*, 30 October 2013) <https://blog.chartbeat.com/2013/10/31/quantitative-turn-journalism/> accessed 9 February 2022.

[16] WIPO Secretariat, 'WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI)' (WIPO 2019) WIPO/IP/AI/2/GE/20/1; WIPO, 'Revised Issues Paper on Intellectual Property Policy and Artificial Intelligence' (WIPO 2020) WIPO/IP/AI/2/GE/20/1 REV <https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_ai_2_ge_20/wipo_ip_ai_2_ge_20_1_rev.pdf> accessed 27 November 2020.

[17] See for example Bernt Hugenholtz and others, 'Trends and Developments in Artificial Intelligence – Challenges to the Intellectual Property Framework' (European Commission 2020).

[18] This paper uses the expression *robojournalism*, although the available terminology – referring to more or less the same concept – is much broader, ranging from computational, automated or algorithmic journalism to data journalism, journalism as programming or programmer-journalism to open-source journalism to computer-assisted reporting. See Mark Coddington, 'Clarifying Journalism's Quantitative Turn - A Typology for Evaluating Data Journalism, Computational Journalism, and Computer-Assisted Reporting' [2015] Digital Journalism 331, 332; Bucher (n 13) 920.

[19] On these categories, see section III.

mainly provided by external companies that collaborate with news outlets,[20] but not only by these parties/actors. More and more news companies are developing AI internally for the generation of automated news.[21]

Our research targets the news industry for at least two reasons. The use of AI in the fields of music and art is well discussed,[22] but research on robojournalism from the perspective of copyright law, while considered a priority domain,[23] is still far from complete, especially with respect to empirical evidence in this field. Demonstrating the practices adopted in the news industry with respect to AI-generated output and the extent to which the industry implements such solutions seeks to demystify the theoretical analysis and back it up with data. On the other hand, the news industry seems to be rather keen on the use of automated journalism.[24] As indicated above, journalistic tasks carried out by AI primarily include the reporting of finance, sports and weather, where a massive amount of raw data is available.[25] These fields are heavily reliant on numbers and data, which an AI system can process and organise extremely quickly and then generate informational reports – something that is tremendously useful for the wider public interested in this data, but also a tedious task human journalists might dread. Indeed, such work might need more mechanical and less creative input from the journalists. This is further backed by the mere fact that 'data' often has

[20] Examples in the EU here include 'AX Semantics', 'Text-On', '2txt NLG', 'Retresco' and 'Textomatic' operating in Germany, as well as 'Syllabs' or 'Labsense' active in France.

[21] Examples include 'MittMedia/United Robots' (Sweden), 'NTB/Bakken & Baeck' (Norway), 'Austria Press Agency' (Austria), and the 'Berliner Morgenpost' (Germany).

[22] Mark Perry and Thomas Margoni, 'From Music Tracks to Google Maps: Who Owns Computer-Generated Works?' (2010) 26 Computer Law & Security Review 621; Ana Ramalho, 'Will Robots Rule the (Artistic) World? A Proposed Model for the Legal Status of Creations by Artificial Intelligence Systems' (2017) 21 Journal of Internet Law 12; Ana Ramalho, 'Originality Redux: An Analysis of the Originality Requirement in AI-Generated Works' [2018] AIDA 23; Jean-Marc Deltorn and Franck Macrez, 'Authorship in the Age of Machine Learning and Artificial Intelligence' in Sean M O'Connor (ed), *The Oxford Handbook of Music Law and Policy* (OUP 2019) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3261329> accessed 5 September 2019; Gerald Spindler, 'Copyright Law and Artificial Intelligence' (2019) 50 IIC 1049; Péter Mezei, 'From Leonardo to the Next Rembrandt – The Need for AI-Pessimism in the Age of Algorithms' (2020) 2 UFITA Forthcoming; Hugenholtz and others (n 17); P Bernt Hugenholtz and João Pedro Quintais, 'Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?' (2021) 52 IIC 1190-1216 <https://link.springer.com/epdf/10.1007/s40319-021-01115-0> accessed 5 October 2021; Daniel J Gervais, 'The Human Cause' in Ryan Abbott (ed), *Research Handbooks on Intellectual Property and Artificial Intelligence* (forthcoming) <https://papers.ssrn.com/abstract=3857844> accessed 19 November 2021; Tim W Dornis, 'Artificial Creativity: Emergent Works and the Void in Current Copyright Doctrine' (2020) 22 YALE J. L. & TECH 1; Tim W Dornis, 'Of "Authorless Works" and "Inventions without Inventor" – the Muddy Waters of "AI Autonomy" in Intellectual Property Doctrine' (2021) 43 EIPR 570.

[23] Hugenholtz and others (n 17) 33.

[24] See Section V.1.

[25] Elizabeth Blankespoor, Ed deHaan and Christina Zhu, 'Capital Market Effects of Media Synthesis and Dissemination: Evidence from Robo-Journalism' (2018) 23 Review of Accounting Studies 1; Yair Galily, 'Artificial Intelligence and Sports Journalism: Is It a Sweeping Change?' (2018) 54 Technology in Society 47; Andrey Miroshnichenko, 'AI to Bypass Creativity. Will Robots Replace Journalists? (The Answer Is "Yes")' (2018) 9 Information 183.

no language barriers. Sports statistics, stock market or weather information can easily be 'translated' into any written language. As such, digital journalism is less locked into the territory of a certain news agency's linguistic domain. On the other hand, journalists benefit from being able to devote more time for pieces of an investigative, event-driven and storytelling nature where AI continues to struggle.[26]

Against this background, a critical question arises: can humans be replaced by AI to generate mechanical/less creative news reports? This research project seeks to fill that gap in literature by turning to the application of AI to the specific field of journalism and copyright law. The conclusions drawn in this paper combine the law and technology analysis with empirical evidence as well as insights from media and communications studies.

The paper is structured as follows. In Section II, it lays the groundwork of the law by addressing briefly the protectability of the outputs of robojournalism under the existing European Union copyright laws. Section III introduces the technological perspectives of robojournalism. Section IV covers the business realities of robojournalism in the European written news industry. Finally, Section V summarises the key findings of media and communications studies research papers on the implications of AI for journalism.

Our findings generally indicate that most corporations outsource the creation of the relevant technology, and to a certain degree they apply the same available technologies, namely natural language processing. Still, from a copyright perspective our results demonstrate that the extent to which European journalism relies on assistive and generative technologies to produce written output does not justify changing the current anthropocentric copyright system. These findings have wider implications as AI-generated outputs have prompted many to talk about market failure in case copyright (or related rights) protection is refused for such works. We believe that our research finds evidence that the correct argument is to the contrary. Relying on automated journalism has other benefits that go beyond copyright law – being able to report news extremely quickly in the way only 'robojournalists' are capable of satisfies demanding consumer expectations, namely getting news reports extremely quickly on a wide variety of topics. This, coupled with the fact that human journalists will now have more free time to dedicate to creative and investigative journalism, should be seen as a sufficient incentive for the news industry, leaving extended copyright protection aside. An important caveat is nonetheless needed in this respect – the newly introduced press publishers' related right still needs to be tested on the market with respect to robojournalism. It is interesting to see to what extent robojournalism will challenge the operation of this new right.

II. Copyright, AI and journalism – the status quo

[26] David Caswell and Konstantin Dörr, 'Automated Journalism 2.0: Event-Driven Narratives: From Simple Descriptions to Real Stories' (2018) 12 Journalism Practice 477, 478; Aljosha Karim Schapals and Colin Porlezza, 'Assistance or Resistance? Evaluating the Intersection of Automated Journalism and Journalistic Role Conceptions' (2020) 8 Media and Communication 16, 21.

Where copyright and AI are concerned, the big discussion can be divided in two specific categories of issues – upstream and downstream.[27] The former tackle questions with the input of an AI process, namely the legal issues tied to the training data. These include text and data mining,[28] liability for copyright infringing content, adaptation right and derivative works, as well as broader questions of access to data and data ownership.[29] The analysis of these matters lies beyond the ambit of this paper, but it should be acknowledged that they play an important role in determining the legality of the training datasets which are one of the essential pillars in the AI process. The reason why these issues are not analysed in depth in this paper is because robojournalism currently thrives in fields heavy on data, as will appear from the empirical analysis that follows. Data and facts as such are not the object of copyright protection. This cornerstone principle, rooted in the TRIPS Agreement by virtue of the idea/expression dichotomy,[30] often gets overlooked in our data economy reality. In light of this, many of the issues that emerge from text and data mining do not generate difficulties in the practice of robojournalism, even though they theoretically may represent an important legal issue and heavily engage questions of infringement of the reproduction right. Different (i.e. non-copyright) concerns are those linked to access to data, the free flow of public and non-personal data, and data propertisation. These would require going into the in-depth analysis of other legal instruments, which goes beyond the scope of this paper.

One final caveat is necessary. The newly introduced related right for press publishers as per Article 15 of the CDSM Directive[31] will certainly have significant consequences for journalism fuelled by AI.[32] This paper only briefly touches upon the potential impact of the new right. We cautiously point out that instead of copyright protection, a related rights protection of AI-generated output in the field of journalism would perhaps be the appropriate legal right.[33] Before going into this area, the section will focus on the so-called downstream or output issues, and question to what extent copyright protection sustains for robojournalistic output.

---

[27] Burkhard Schafer and others, 'A Fourth Law of Robotics? Copyright and the Law and Ethics of Machine Co-Production' (2015) 23 Artificial Intelligence and Law 217, 219.

[28] For a recent EU discussion on the problems and solutions with respect to text and data mining, see Alain Strowel and Rossana Ducato, 'Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out' (2021) 43 EIPR 322.

[29] Bernt Hugenholtz, 'Data Property: Unwelcome Guest in the House of IP' <https://dare.uva.nl/personal/search?identifier=c5791bb2-e1de-4d7b-9720-68021b5ae5cc> accessed 9 August 2019.

[30] Agreement on Trade-Related Aspects of Intellectual Property Rights as Amended by the 2005 Protocol Amending the TRIPS Agreement, art 9(2).

[31] Directive 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market Official Journal L 130.

[32] Taina Pihlajarinne and others, 'European Copyright System as a Suitable Incentive for AI-Based Journalism?' in Taina Pihlajarinne and Anette Alén-Savikko (eds), *Artificial Intelligence and the Media* (Edward Elgar 2022) <https://papers.ssrn.com/abstract=3853730> accessed 11 January 2022.

[33] For further analysis of the press publishers' right see Ula Furgał, 'The EU Press Publishers' Right: Where Do Member States Stand?' (2021) 16 Journal of Intellectual Property Law & Practice 887; Pihlajarinne and others (n 32).

1. International instruments

Output generated through and with the assistance of AI requires serious consideration of the essence of copyright law. The two key notions are 'authorship' and 'originality'. These are highly interconnected and discussion of one inevitably leads to considerations of the other.[34] Despite the fact that the two concepts have been under stringent scrutiny for decades,[35] none of the international copyright instruments proves clear and straightforward definitions of them.

With respect to authorship, the Berne Convention lacks a correlative definition.[36] This could be due to the fact that the necessity for such a definition is redundant, or even perhaps because it may be considered obvious that the author of a copyright work must be a human being. With this in mind, some academics – and some copyright law statutes – suggest that despite the lack of an explicit internationally agreed definition of an author, generally the author is the person who creates the work.[37] To this end, the substantive provisions of the Berne Convention point towards human authorship. One such indication, according to Sam Ricketson and Jane C. Ginsburg, hinges on the one hand on the fact that copyright duration is linked to the life of the author, and on the other that moral rights only entitle a human. In that respect moral rights are attached to the personality and presence of an author.[38] Thus, the human being is an indispensable element in the equation. Besides, considering that the Berne Convention was inspired by a group of European

[34] Jane C Ginsburg, 'The Concept of Authorship in Comparative Law' (2003) 52 DePaul Law Review 1063, 1072; Jani McCutcheon, 'The Concept of the Copyright Work under EU Law' (2019) 44 European Law Review 767, 779.

[35] See the following among many others Andreas Rahmatian, 'Originality in UK Copyright Law: The Old "Skill and Labour" Doctrine Under Pressure' (2013) 44 IIC 4; Thomas Margoni, 'The Harmonisation of EU Copyright Law: The Originality Standard' in Mark Perry (ed), *Global Governance of Intellectual Property in the 21st Century* (Springer 2016); Eleonora Rosati, *Originality in EU Copyright: Full Harmonization through Case Law* (Edward Elgar 2013).Sam Ricketson, 'The 1992 Horace S. Manges Lecture - People or Machines: The Berne Convention and the Changing Concept of Authorship' (1991) 16 Columbia-VLA Journal of Law & the Arts 1; Adolf Dietz, 'The Concept of Authorship under the Berne Convention' (1993) 155 RIDA 3; Lionel Bently, 'Copyright and the Death of the Author in Literature and Law' (1994) 57 Modern Law Review 973; Lionel Bently, 'R. v. the Author: From Death Penalty to Community Service - 20th Annual Horace S. Manges Lecture, Tuesday, April 10, 2007' (2008) 32 Columbia Journal of Law and the Arts 1; Jane C Ginsburg, 'The Role of the Author in Copyright' in Ruth L Okediji (ed), *Copyright Law in an Age of Limitations and Exceptions* (Cambridge University Press 2017); Martha Woodmansee, 'On the Author Effect: Recovering Collectivity' in Martha Woodmansee and Peter Jaszi (eds), *The Construction of Authorship: Textual Appropriation in Law and Literature* (Duke University Press 1994); Martha Woodmansee and Peter Jaszi, *The Construction of Authorship: Textual Appropriation in Law and Literature* (Duke University Press 1994).

[36] Sam Ricketson, *The Berne Convention for the Protection of Literary and Artistic Works: 1886-1986* (Kluwer 1987) para 6.4.

[37] Antoon Quaedvlieg, 'Authorship and Ownership: Authors, Entrepreneurs and Rights' in Tatiana-Eleni Synodinou (ed), *Codification of European Copyright Law Challenges and Perspectives* (Wolters Kluwer 2012) 198-99; Copyright, Designs and Patents Act 1988, s 9(1) (UK).

[38] Stef van Gompel, 'Creativity, Autonomy and Personal Touch' in Mireille van Eechoud (ed), *The Work of Authorship* (Amsterdam University Press 2014) 127-28.

authors under the leadership of Victor Hugo,[39] it is not surprising that an anthropocentric view on authorship prevailed.[40]
Equally, the term 'originality' is not defined in the Berne Convention. While there is a reference to 'intellectual creations' in Art. 2(5), this is strictly tied to collections of literary or artistic works such as encyclopaedias and anthologies. However, considering the dependence of authorship on originality, which becomes clearer in the brief analysis of the EU setting below, the anthropocentric view of originality comes to the surface.

2. European Union

Turning to the EU – the focus of this paper – human authorship emerges very prominently from the originality standard. Originality has been the subject of a long list of cases from the Court of Justice of the European Union (CJEU), some of which will be briefly analysed here from the perspective of journalism. A detailed and thorough analysis of the two key notions – authorship and originality – usually engages with them separately and studies their origins and evolution independently before bringing them under the same umbrella. However, such an exercise is beyond the scope of this paper and furthermore, as stated above, the two are highly interdependent. Therefore, this section will only briefly examine authorship and originality from an EU law perspective and will then reflect upon what these legal standards mean for the purposes of journalism and more specifically robojournalism.
The standard of originality that the CJEU established necessitates that a work be considered original (and thus, potentially copyright protected) only if it constitutes the author's own intellectual creation.[41] This definition has been criticised as being rather circular.[42] Nonetheless, it puts the figure of the author at the centre stage of EU copyright law. The standard is said to entail two dimensions: normative and causative[43] (also known as a subjective and objective, respectively).[44] The normative focuses on the substance of originality as such, namely a work should reflect an intellectual creation. Present very prominently in civil law jurisdictions, this constitutes the idea that a work should demonstrate the imprint and personal stamp of the author.[45]

---

[39] Sam Ricketson and Jane C Ginsburg, *International Copyright and Neighbouring Rights: The Berne Convention and Beyond Two Volume Set* (2nd ed, Oxford University Press 2006) pt 1.
[40] Madeleine de Cock Buning, 'Autonomous Intelligent Systems as Creative Agents under the EU Framework for Intellectual Property' (2016) 7 Eur. J. Risk Reg. 310, 319; Ricketson, 'The 1992 Horace S. Manges Lecture' (n 35) 6.
[41] Case C-145/10 *Eva-Maria Painer v Standard VerlagsGmbH and Others* ECLI:EU:C:2011:798, para 89; Case C-833/18 *SI and Brompton Bicycle Ltd v Chedech / Get2Ge* ECLI:EU:C:2020:461, para 22. Case C-5/08 *Infopaq International A/S v Danske Dagblades Forening* ECLI:EU:C:2009:465, para 37.
[42] Hugenholtz and others (n 17) 70.
[43] Daniela Simone, *Copyright and Collective Authorship: Locating the Authors of Collaborative Work* (Cambridge University Press 2019) 23.
[44] Mireille van Eechoud, 'Along the Road to Uniformity – Diverse Readings of the Court of Justice Judgments on Copyright Work' (2012) 3 JIPITEC 60, 70.
[45] ibid.

Importantly, the emphasis on intellectual creation and authorial imprint should not be confused with a requirement for a certain degree of aesthetic quality, merit or specific purpose that do not form a requirement of originality under copyright law.[46]
The causative (or objective) dimension pertains to the originating factor.[47] Rooted in UK copyright law, the idea is that a work is not protected unless it originates from a human author. Thus, the emphasis in originality is not on novelty and creativity, but on the fact that a work is created by an author. This is a clear indication of how the originality standard encompasses the authorship notion. Consequently, a work is protected only if it is the product of a human author whose intellectual expression stamps the work,[48] and all of this should result in a subject matter that is sufficiently clear and objective.[49]

3. Copyright implications for journalism

Determining the presence of 'free and creative choices'[50] and thus of the intellectual creation in a work is not a straightforward exercise. Is it a high or a low hurdle to clear?[51] Is it at all possible to assess this objectively?[52] Does the originality test follow the common law or the civil law tradition or is it better described as a mix of both?[53] To that end, the CJEU case-law has provided some insight into the parameters of originality.
Even though the 'author's own intellectual creation' standard is nowadays understood to apply universally to all types of works, an argument can be made that it is of value to determine and bear in mind the type of work in question. Journalistic literary output very often follows a pre-determined style that is imposed by the specific type of publication, newspaper or magazine, its audience, or its subject matter, among other things. There will be norms with which journalists would have to necessarily comply, as a matter of general journalistic practice, but also imposed more specifically by their editors. This discussion is something copyright scholarship has tackled and discussed under the broader label of 'creative constraints'[54] or 'freedom of the creator'.[55]

---

[46] van Gompel (n 38) 103.
[47] Rahmatian (n 35) 12; *University of London Press v University Tutorial Press* [1916] [609-610].
[48] *Painer* (n 41) para 92.
[49] Case C-310/17 *Levola Hengelo BV v Smilde Foods BV* ECLI:EU:C:2018:899, para 40.
[50] *Painer* (n 41) para 90.
[51] van Gompel (n 38) 95.
[52] Estelle Derclaye, 'Wonderful or Worrisome? The Impact of the ECJ Ruling in Infopaq on UK Copyright Law' (2010) 32 EIPR 247, 247.
[53] Ramalho, 'Originality Redux: An Analysis of the Originality Requirement in AI-Generated Works' (n 22) 27; Benoît Michaux, 'L'originalitéen Droit d'auteur, Une Notion Davantage Communautaire a Prés l'arrêt Infopaq' (2009) 5 Auteurs & Media 473, 473.
[54] van Gompel (n 38) 104.
[55] Estelle Derclaye and Marco Ricolfi, 'Opinion of the European Copyright Society in Relation to the Pending Reference before the CJEU in Cofemel v G-Star, C-683/17' (European Copyright Society 2018) 6 <https://europeancopyrightsocietydotorg.files.wordpress.com/2018/11/ecs-opinion-cofemel_final_signed.pdf> accessed 11 January 2022.

The CJEU's guidance in this respect has been instructive. In *BSA*, the CJEU addressed the protectability of a graphic user interface enabling communication between a computer program and the user. The interface may potentially fall within the general protectable subject matter by copyright law pursuant to the Information Society Directive[56] provided that the interface meets the golden 'author's own intellectual creation' standard.[57] The CJEU stressed that if the expression of the graphic user interface's components was dictated by their technical function, the criterion of originality would not be met.[58] In *Football Dataco,* the Court further expanded this notion of functionality and technical limitations.[59] That case concerned a claim of infringement of intellectual property rights – a *sui generis* database as well as copyright as a database – related to fixture lists. While the former intellectual property right is irrelevant for the present analysis (as it pertains to the substantial investment that has gone into the obtaining, verification or presentation of the contents of a database)[60], a database could also be subject to copyright protection if it constitutes the author's own intellectual creation by reason of the selection or arrangement of its contents.[61] Pursuant to Article 3(2) of the Database Directive,[62] read in conjunction with Recital 15, originality here is understood by reference to the structure of the database as opposed to the contents, meaning the elements that constitute its contents. Focusing on the aspect of the intellectual creation, the CJEU emphasised that the effort and skill involved in creating the data remain irrelevant in the assessment of the eligibility of the database itself for copyright protection.[63]

Importantly, the CJEU placed emphasis on the way in which the selection and the arrangement of the data in the databases was carried out. In *Football Dataco,* this was done in accordance with a set of rules, parameters and organisational constraints, along with the specific requests of the football clubs in question.[64] With this in mind, the CJEU turned to analyse whether this process could reach the required originality threshold – would the selection and the arrangement of the data in the fixtures amount to the expression of the author's creative ability in an original manner, through which that author has made free and creative choices and thus stamped the work with their own personal touch? At this stage, the CJEU reaffirmed its position that there will be no room for creative freedom where choices are dictated by technical considerations, rules or constrains.

[56] Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society Official Journal L 167 (hereinafter 'InfoSoc Directive').
[57] Case C-393/09 *Bezpečnostní softwarová asociace – Svaz softwarové ochrany v Ministerstvo kultury* ECLI:EU:C:2010:816, paras 40-42 and 44-46.
[58] ibid 49.
[59] Case C-604/10 *Football Dataco Ltd and Others v Yahoo! UK Ltd and Others* ECLI:EU:C:2012:115.
[60] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases OJ L 77, art 7(1).
[61] ibid, art 3.
[62] ibid.
[63] *Football Dataco and Others* (n 59) para 33.
[64] ibid 35.

Consequently, the CJEU seems to suggest that evaluating the creative elements in the process of producing the copyright work is as important as the final creative features of the product itself.
A crucial aspect in this discussion is the available room for creativity, i.e. the creative constraints. Limiting the author by certain creative constraints is not a sufficient reason to deny that author copyright protection.[65] Yet, this is a very a delicate point. Some constraints might be too rigid leaving the author little or no space for creativity. Others may actually stir creativity – too much freedom can 'paralyse' creativity as the creative space becomes too wide to control and choices become too complex.[66]
The creative freedom of journalists could also be dictated by some very specific constraints. Journalists often strictly follow an editorial statute and/or an ethical code.[67] For example, the Reuters Handbook of Journalism lists the following as aspects guiding their journalistic outputs: story length, basic story structure, consistency of style, key words, language that must be avoided.[68] All these constraints, if very diligently followed, risk excessively restraining the free and creative choices of the human journalist and thus it could be argued that some journalistic pieces do not qualify for copyright protection themselves as they would follow too strictly pre-determined rules. Put differently and in the words of the CJEU in *Football Dataco* case, if the selection and arrangement of data is done in accordance with a set of rules, parameters and organisational constraints, then it can be convincingly argued that there would be little room for copyright protected subject matter.[69] In *Funke Medien*, the CJEU underlined that the so-called 'Afghanistan Papers' (i.e. military status reports on the deployment of the Federal German armed forces) would benefit from the economic rights in the InfoSoc Directive only provided they are original in the sense of the 'author's own intellectual creation'.[70] This was a finding of fact for the national court, so the CJEU did not engage with that point in the preliminary ruling. However, AG Szpunar was slightly more explicit, raising doubts as to the copyrightability of the military reports in light of their 'unusual nature […] to the extent that their content is purely informative'.[71] Having said that, a note of caution is necessary when drawing a close parallel between the copyright protection of databases as elaborated in *Football Dataco* and journalistic output. Even though journalistic conventions may impact on the room for creative freedom of journalists, it is fair to say that there is still some (and not so limited) room for creativity even when editorial handbooks prescribe specific parameters to be followed closely. In this respect, each instance must be assessed on its own merits. It could be that the more restrictions there are, the more creative an author is

[65] Hugenholtz and others (n 17) 73.
[66] van Gompel (n 38) 107.
[67] ibid 116.
[68] 'Reuters Handbook of Journalism' (2008) <https://www.trust.org/contentAsset/raw-data/652966ab-c90b-4252-b4a5-db8ed1d438ce/file> accessed 11 January 2022.
[69] *Football Dataco and Others* (n 59) para 35.
[70] Case C-469/17 *Funke Medien NRW GmbH v Bundesrepublik Deutschland* ECLI:EU:C:2019:623, paras 19-20.
[71] C-469/17 *Funke Medien NRW GmbH v Bundesrepublik Deutschland* ECLI:EU:C:2018:870, Opinion of AG Szpunar, para 21.

pushed to be.[72] Yet, all will depend on the intellectual creation that was the product of free and creative choices of a human author.

Inevitably these lead to one very straightforward and simple conclusion, often undermined in the context of AI. Not all output, even if directly the product of human hands, shall receive copyright protection. Free and creative choices and expression of intellectual creation must be present. Turning to the realities of robojournalism, in identifying these choices – which trigger copyrightability – it is necessary to lift the technical veil and unpack the basic process behind the production of journalistic pieces with the aid of (or allegedly entirely autonomously by) AI.

III. Technological perspectives/levels of creativity of robojournalism

A 2020 report, commissioned by the European Commission and carried out by the Institute for Information Law (IViR) and the Joint Institute for Innovation Policy (JIIP), studied the specific IPR challenges from the perspective of copyright and patent law. It identified three specific domains as priority ones – pharmaceutical research, science/meteorology, and journalism.[73] The present research turns to the specific implications of robojournalism from the perspective of copyright law and uses as a starting point the technological classification presented in that report.

From a technological perspective, four specific applications of robojournalism appear to have come to the forefront as most relevant: automated content production; data mining; news dissemination; and content optimisation.[74] While all of these have produced fascinating discussions, the focus of this paper is on the first application. Automated content production can entail assistive or generative machine learning techniques.[75] The former rely heavily on the involvement of a human being in the production of content – thus, it can be presumed that the control is still entirely in the hands of the human journalist. The copyright law issues that would emerge in this respect would not differ from those one often seen in the context of photography and classic video games.[76]

Generative technologies create the most difficulties. The presumption is that these are 'capable of creating media content largely autonomously and with very little human intervention'.[77] In this part we seek to unpack these technologies and identify the degree of autonomy in the robojournalism process where generative technologies are used. The available technologies have become mainstream in the field of descriptive reporting tasks but are still not capable of being directly applied to other more complicated forms, such as storytelling journalism. This study turns to the technological reality behind descriptive reporting only. However, it is worth mentioning that

[72] van Gompel (n 38) 107.
[73] Hugenholtz and others (n 17) 33.
[74] Efthimis Kotenidis and Andreas Veglis, 'Algorithmic Journalism – Current Applications and Future Perspectives' (2021) 2 Journalism and Media 244, 246.
[75] Hugenholtz and others (n 17) 57.
[76] Jane C Ginsburg and Luke Ali Budiardjo, 'Authors and Machines' (2019) 34 Berkeley Technology Law Journal 343, 378.
[77] Hugenholtz and others (n 17) 57.

the reason why more sophisticated journalism is still out of reach for robojournalism is the lack of data models suitable for encoding event-driven narratives.[78] Some progress has been made and models have recently been suggested that aim to target event-driven natural language generation.[79] With that in mind, our study digs into the technology behind robojournalism that thrives on dry statistics and numbers, such as sports, weather and finance. From the outside, it appears that the technology in these fields has indeed become rather mainstream and accessible to many. This is a proposition we test with the empirical study in Section 4. Here, we seek to unpack their functioning in order to address the copyright protectability issues. In this discussion, we start from the main protagonist – 'natural language generation', or 'NLG'.

1. NLG – the key driver in robojournalism

NLG is a subcategory of Natural Language Processing. It is the major technology in respect of robojournalism due to its capability of transforming data into text. Caswell and Dörr have defined NLG as 'the automatic creation of text from digitally structured data'.[80] NLG systems can be either rule-based, where all rules are pre-coded *ex ante*; or machine learning, whereby the system learns with example after having been exposed to a large quantity of learning material.[81] The latter have been the source of a genuine revolution since rule-based systems entailed heavy pre-coding for all articles in a specific domain.[82] While ML techniques have been around since the 1940s,[83] their popularity and widespread application in various domains, including journalism, picked up exponentially in the past decade. NLG is now mainstream and accessible even to those without specialised technical training, entering the newsrooms of various scale companies to automate certain routine tasks.[84] Outputs of NLG come very close to being automatically generated by the system directly, but not entirely. News pieces are often the final product of human and algorithmic collaboration.[85]

The study in this paper turns to unravel NLG techniques as applied to the specific journalistic fields of weather, sport, finance, and real estate reports. The reason behind the selection of these specific domains of journalism is that they focus on telling us 'what happened or is happening' since 'the limitation of only answering the 'what', rather than the 'why', is due to the inability of computer systems to analyse events against contextual life-world knowledge'.[86] An important

[78] Caswell and Dörr (n 26) 478-79.
[79] ibid 483.
[80] ibid 477.
[81] Stefanie Sirén-Heikel and others, 'Unboxing News Automation' (2019) 1 Nordic Journal of Media Studies 47, 49.
[82] ibid.
[83] Warren McCulloch and Walter Pitts, 'A Logical Calculus of Ideas Immanent in Nervous Activity' (1943) 5 Bulletin of Mathematical Biophysics 114, 114.
[84] Caswell and Dörr (n 26) 478.
[85] Anja Wölker and Thomas E Powell, 'Algorithms in the Newsroom? News Readers' Perceived Credibility and Selection of Automated Journalism' (2021) 22 Journalism 86, 88.
[86] Sirén-Heikel and others (n 81) 50.

development fuelling robojournalism in these fact-driven fields is datafication. Society is used to digitalisation – there is barely a domain still operating on analogue. With respect to news, however, the fact that data are constantly generated, rendered open and available has pushed literature to discuss digitisation evolving into datafication, which becomes particularly relevant for the news industry.[87] The accuracy and reliability of data is pertinent, especially in a field such as journalism which is tied to many ethical responsibilities. This is usually referred to broadly as 'data-to-text' generation.[88] Many authors and journalists spend a significant amount of their time producing documents from data and this is often not their primary task. Once this can be either entirely or to a large extent delegated to an algorithm, the journalists' productivity and morale are automatically enhanced.[89] It must, however, be borne in mind that developing an NLG system is costly and not all news companies can afford it. The decision of whether to invest is very often an economic one.[90]

2. Dissecting NLG

The central question is – what is the degree of human involvement in the generative technology process itself and does that degree justify a copyright claim to arise? In these complex technical processes, human involvement can take place at various stages of the process – pre-production, during the NLG process and post-production. This is something Hugenholtz and Quintais refer to as conception, execution, and redaction.[91] This corresponds neatly to the CJEU's analysis in *Painer*, where the CJEU emphasised that authorial creative choices can take place in three different stages when a photograph is taken: at the preparation stage; when taking the photograph; and post-production when choosing a developing technique and method.[92] Applying this to the NLG process, it is important to dissect the typical NLG process.

As with all technologies based on machine learning, there is no one-size-fits all technical model for all NLG systems. Yet, it appears that a consensus exists that in any NLG process six basic activities need to be performed; these go all the way from input data to a final output text.[93] Even though the order of these may vary, and some of them may be merged, these stages always come back in one way or another as they represent the stages of any text generation. Reiter and Dale define them in the following manner.[94]

---

[87] ibid.

[88] Albert Gatt and Emiel Krahmer, 'Survey of the State of the Art in Natural Language Generation: Core Tasks, Applications and Evaluation' (2018) 61 The Journal of Artificial Intelligence Research 65, 66.

[89] Ehud Reiter and Robert Dale, 'Building Applied Natural Language Generation Systems' (1997) 3 Natural Language Engineering 57, 59.

[90] ibid 61; Meredith Broussard and others, 'Artificial Intelligence and Journalism' (2019) 96 Journalism & Mass Communication Quarterly 673, 677-78.

[91] Hugenholtz and Quintais (n 22) 1202.

[92] *Painer* (n 41) para 91.

[93] Reiter and Dale (n 89) 64.

[94] ibid 64-68.

1) Content determination – the 'process of deciding what information should be communicated in the text'.
2) Discourse planning – the ordering and structuring of the text into a coherent form; for example ensuring there is a beginning, middle and end.
3) Sentence aggregation – the actual grouping of messages and information into sentences, which is not always a necessary step, but it often eases 'fluency and readability' of the text.
4) Lexicalisation – the 'process of deciding which specific words and phrases should be chosen to express the domain concepts and relations which appear in the messages'.
5) Referring expression generation – the selection of specific words or phrases to identify certain information.
6) Linguistic realisation – the step that ensures that the text is grammatically coherent, following rules of syntax, morphology, and orthography.

Each stage entails its own individual peculiarities, depending on several elements, including the type of text to be produced, the style of writing, and the target audience. For instance, the editorial constraints relating to the Reuters Handbook example discussed above would certainly play a heavy role in the setting up of the technical specifications in each of the six stages. The beauty of a process of this kind is that it provides the editors, journalists and the computer scientists involved with a wide freedom to tweak and adjust. In that regard, a parallel could be made with what Lehr and Ohm underline with respect to machine learning in general – these complex processes are not monoliths; on the contrary, before coming into a final output form, the work 'dances back and forth' across the various steps and stages instead of proceeding through them linearly.[95] Furthermore, several of these six tasks can be combined when building the architecture of the system, for which there are numerous existing models.[96]

3. Copyright implications for NLG

Matching this technical analysis with the copyright discussion above, cases like *Painer* have already stressed the fact that originality can take place at different stages.[97] What matters is not one single epiphany-like moment. Instead, creativity and originality can take place at different moments of the NLG process. Consequently, one is prompted to seek the choices that authors involved in the process make (as opposed to the system itself) in order to satisfy the human authorship requirement and determine whether these choices indeed are 'free and creative' to constitute an intellectual expression. The main instrument in this analysis will be notion of constraints introduced above in Section 2. The essential question to be asked is whether the imposed technical constraints excessively limit the creators' freedom in each of these stages to the extent that there is no copyright claim subsisting.

[95] David Lehr and Paul Ohm, 'Playing with the Data: What Legal Scholars Should Learn About Machine Learning' (2017) 51 U.C. DAVIS L. REV. 653, 669.
[96] Reiter and Dale (n 89) 68.
[97] *Painer* (n 41).

Literature has categorised these six tasks in two stages – early and late.[98] Early decisions are directly tied to the input data. In this respect, Gatt and Krahmer pivot the early decisions around the question of which information to convey to the reader, while the late decisions are strictly tied to the choice of words to use in a particular sentence and how to put them in their correct order.[99] The first stage – content determination – is an early task and it can be suggested that the decision of which data to insert in the NGL process is not immediately the type of free and creative choice that triggers a copyright claim. More importantly, this content determination in the NLG process is typically carried out through automated means where the process leaves little room for human intervention.

Content determination does not appear to entail any free and creative choices that would trigger a copyright claim. This is due to the idea/expression dichotomy, according to which copyright does not protect ideas, but only expression.[100] While this has always been a very difficult line to draw in reality, it can be safely stated that deciding what information should be communicated in the text may stay closer to the idea side of the spectrum. Admittedly, it can be argued that there may be some free and creative choices in the selection of the information that would go into the NLG process, it must be stressed that words in isolation would not constitute the author's own intellectual creation. *Infopaq* taught us that it is 'only through the choice, sequence and combination of those words that the author may express his creativity in an original manner and achieve a result which is an intellectual creation'.[101] Thus, regardless of whether content determination is an activity carried out by a human author or automatically by the system itself, it would not have any bearing on the copyright claim.

On the other hand, discourse planning – the second activity – may cover some important features for the copyright claim. The ordering and structuring of the text into a coherent form whereby logical connections between the beginning, the middle and the end of the text are present would certainly entail free and creative choices. These could be limited by the editorial constraints imposed by the specific journalistic output, but regardless of this would be the type of activity that triggers a copyright claim that goes beyond simple idea, dictated by functionality.

The next stage – sentence aggregation – does not appear to have any impact on the copyright claims, especially considering that this is not always a necessary stage and would typically entail the grouping of the sentences together. Arguably, these are not choices that would entail sufficient intellectual creation in a free and creative manner as required by copyright law. Most likely, these choices would be heavily influenced by the information that is being conveyed.

Next comes the lexicalisation phase, which appears to be particularly important from a copyright perspective. Lexicalisation entails the process of deciding which specific words and phrases should be used to express the domain concepts and relations.[102] It looks like lexicalisation can be carried

[98] Gatt and Krahmer (n 88) 71.
[99] Gatt and Krahmer (n 88).
[100] TRIPS (n 29) art 9(2).
[101] *Infopaq* (n 41) para 45.
[102] Reiter and Dale (n 89) 67.

out by hard coding, a process in which humans determine in advance which words would come to represent any specific concept or domain. Arguably, the decision of using one word instead of another could reflect free and creative choices made by the author. Yet, it is questionable whether merely choosing one word could constitute the authorial choice sufficient to trigger originality. The CJEU case-law has not established a minimum, nor a *de minimis* rule; thus, a case-by-case analysis is required here.

As for the task of referring generation expression, it does not look like any copyright relevant free and creative choices would take place here as this is when certain phrases or words are selected to be identified with others. Deciding to use 'the team' and 'they', or 'the score' and 'it' interchangeably are minimal choices which do not contribute to creative expression.

Finally, during the linguistic realisation task, grammar, syntax, morphology, and orthography are revised. Once again, none of these pertain to the copyright-relevant intellectual creativity – these decisions are mostly dictated by certain rules and therefore the creative freedom for such choices is rather restricted.

4. Interim conclusion

As a result of this brief dissection of the NLG process, it appears that there are at least two specific stages (discourse planning and lexicalisation), where the choices that are being made could be free and creative enough to trigger a copyright claim. However, this is not guaranteed as it may be that editorial policy imposes strict restrictions on creative freedom even during the discourse planning and the lexicalisation. For example, sports reports always contain a certain type of information, which needs to be communicated and which is often presented in the same manner, using the same terms. This would take away the freedom in these two specific tasks. Therefore, if the creative choices during these tasks are too commonplace and banal, it may not even matter whether NLG processes were utilised or whether the entire report was written by a human being. Copyright would simply not subsist in a work deprived of intellectual creativity. Therefore, it may be that robojournalism and copyright law are much ado about nothing.

IV. Business perspectives

1. Lack of empirical data so far

So far, a limited amount of literature has been published that empirically tests NLG service providers. One of the relevant sources pointed out that the algorithmic content industry (ranging from the selection through recommendation to creation) is a massively developing field of business.[103] In 2014, Latzer et al. nevertheless found that automation was ancillary for the news

[103] Latzer and others (n 12) 396.

industry.[104] However, some studies have started shedding light on the use of NLG. Graefe[105] and Dörr[106] discussed the functionality and offers of 11 and 13 NLG service providers, respectively, while Fanta interviewed 15 news agencies on their use of AI tools.[107]
None of these research papers focused on the copyright law aspects of robojournalism. However, this does not mean that the researchers' findings cannot inform – at least indirectly – a research project on the copyright implications of robojournalism. One of the most important findings of these research papers is that media corporations generally outsource the development of AI tools with which they might generate the final literary outputs. As Graefe noted, '[m]any newsrooms, however, lack the necessary resources and skills to develop automated journalism solutions in-house. Media organisations have thus started to collaborate with companies that specialize in developing natural language generation technology to automatically generate stories from data for a variety of domains.'[108] The involvement of NLG service providers in the production of media outputs questions the manner and extent to which media companies could claim copyright protection of output generated with the assistance of NLG.

2. Targeted empirical analysis of NLG service providers

This led us to conduct a targeted empirical analysis of selected European NLG service providers under various factors. We checked how widely they support news publishers with automated journalism tools. We analysed 10 service providers: *AX Semantics, Retresco, Textomatic* from Germany; *Syllabs* and *Labsense* from France; *United Robots* from Sweden; *Bakken & Baeck* from Norway; *Arria* and *RADAR* from the United Kingdom; and *Connexun* from Italy.[109]
We also paid attention to seven variables:

(i) general information of the service (especially the ways these corporations offer their service to their clients, e.g. SaaS, CaaS);
(ii) the role of humans in the process of content generation (especially whether the service is fully automated or requires substantive human control);
(iii) the number of available languages;
(iv) the number of confirmed clients;
(v) the sectors where the given corporation is actively present (besides media & publishing);
(vi) the use of service in journalism/best examples; and,

[104] ibid 403.
[105] Graefe (n 6).
[106] Konstantin Nicholas Dörr, 'Mapping the Field of Algorithmic Journalism' [2016] Digital Journalism 700.
[107] Alexander Fanta, 'Putting Europe's Robots on the Map: Automated Journalism in News Agencies' (University of Oxford 2017) <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-09/Fanta%2C%20Putting%20Europe%E2%80%99s%20Robots%20on%20the%20Map.pdf> accessed 5 August 2021.
[108] Graefe (n 6) 20.
[109] Our focus on the 'European' news industry was therefore not limited to European Union Member States only.

(vii) the availability of the terms of use of the selected corporation's NLG (and if so, what these terms practically include).

The collection of data has evidenced a significant overlap of functionalities and market presence of the distinct service providers, as well as a huge difference among the service providers with respect to the transparency/availability of data regarding the distinct factors/variables we paid attention to.

To start with the commonalities (or available information): most services are offered on a software-as-a-service (SaaS) basis, although some corporations provide for a content-as-a-service (CaaS), hyper-personalised, or custom-built solutions. Most service providers offer NLG, but several corporations also provide for NLP solutions.

The majority of analysed services claim to be fully automated, although a minority necessitates editorial control of the final output (e.g. in case of RADAR). Here, however, we lack information to a significant degree because 4 out of 10 service providers did not indicate whether their product is fully automated or not, and what is in fact understood by 'fully automated' (considering that the term is rather loaded).

Only half of the analysed services published data on the number of available languages in which they offer their services. Where data was available with respect to this variable, the numbers vary heavily: from 6[110] (United Robots) to 110 (AX Semantics). If the language variations (e.g. in Hungarian or Bulgarian) are as effective as NLG services in the leading languages (e.g. in English, German or French), this richness of languages might guarantee a sensible market benefit for early innovators of multi-language NLG service providers.

Corporations are simultaneously present with their solutions in multiple market segments, ranging from e-commerce to national government communications. In general, our empirical findings confirm that the most relevant services are connected to data-driven markets: telecommunication; the financial sector; weather forecasts; sports; and real estate, etc.

The number of confirmed clients of the service providers varies significantly. Three corporations do not provide data on the number of their clients,[111] but the rest report from between many dozens to 800+ clients. At the same time, these numbers are not fully comparable. Some corporations publish the overall number of their partners, while others also specify the number of the news industry clients. For example, Syllabs has 800+ and AX Semantics has 500+ clients overall; on the other hand, Labsense and United Robots report 100+ media clients (also including, however, radio/audiovisual corporations). Finding a correlation between the various factors was not the purpose of the present paper. It would be interesting to explore further how the various service providers' language variations, or their market presence, correlates with the reported client numbers. Only a much deeper empirical analysis – with a direct focus on the given corporation's business strategy – would be capable of shedding light on the correlations.

[110] It is hard to measure whether Syllabs' 'multiple languages' means more or less than six.

[111] These are Textomatic, Bakken & Baeck and Connexun. We nevertheless know that they have existing (and famous) collaborations: Textomatic has built a fruitful collaboration with FOCUS Online (compare with n 4 above); and Bakken & Baeck has collaborated with NTB on football sport reports.

The best available examples of the use of the selected services tend to focus on (the often mentioned) sports or financial reports, or weather forecasts.[112] However, other important elements of the online publishing process (e.g. SEO visibility or topic management) are supported as well.[113] This aspect once again underlines the growing prominence of robojournalism in these specific fields.

Only four out of ten corporations made the terms and conditions of the use of their NLG service available online. Unsurprisingly, these terms are generally silent on copyright-relevant questions. Importantly, two service providers (AX Semantics[114] and Retresco[115]) expressly state that the user of the service shall use their own data for the creation of the relevant content. Another service provider (Textomatic) clarifies that its News-Alert-System's database is filled up by licensed data, open data, and contents from media partners.[116] The fourth service provider (Connexun) notes that its API system relies on data from publicly available websites, including content protected by

[112] The *Stuttgarter Zeitung* uses AX Semantic's service to generate sport, and live fine dust and air quality reports (<https://en.ax-semantics.com/portfolio/stuttgarter-zeitung/>); *FOCUS online* automatically generates weather and finance news with the help of Textomatic's solution (<https://www.pt-magazin.de/de/wirtschaft/innovation/roboter-journalismus---ist-nicht-mehr-wegzudenken_jknpci4d.html>); *Mediafin* automatically generates stock market news feed with Syllabs technology (<https://www.syllabs.com/en/client/lecho-automatically-generates-stock-market-newsfeed>); *Ouest France* generates reports on weather and upcoming cultural events by using Syllabs' solution (<https://www.syllabs.com/en/client/Ouest-France-boosts-its-local-information>); 60.000 local soccer games were reported on during the first "COVID season" in the Netherlands (<https://www.unitedrobots.ai/for-newsrooms/news/how-dutch-ndc-will-cover-60000-regional-football-matches?hsLang=en>); *Bonnier News Local* also automated live sports reporting (<https://www.unitedrobots.ai/for-newsrooms/news/automating-live-sports-at-bonnier-news-local?hsLang=en>); *Bakken & Baeck* and *NTB's* collaboration was also centered around digital football reporters (<https://medium.com/bakken-b%C3%A6ck/building-a-robot-journalist-171554a68fa8>). All websites accessed on 10 November 2021.

[113] The *FAZ.NET* opts for an audience-first experience to increase SEO visibility and topic management (<https://www.retresco.de/wp-content/uploads/2020/09/Retresco-TMS-Case-Study-FAZ.pdf>); *TF1* uses Labsense's service to generate automated editorial content (<https://www.lalettrea.fr/medias_presse-ecrite/2019/05/20/tf1-fait-appel-a-l-intelligence-artificielle-de-labsense-pour-rediger-des-textes-automatiques,108357671-art>). All websites accessed on 10 November 2021.

[114] AX Semantics' Master Subscription Agreement, §2.2 ('The customer may only process his own data, or data he has legal access and usage rights for, for his own purposes. All rights to the data provided by the customer remain with him.') Available via <https://assets.ax-semantics.com/terms-and-conditions.pdf>. All websites accessed on 10 November 2021.

[115] Retresco's Terms & Conditions, G.2 ["Retresco will store (duplicate) and process (catalogue or prepare and summarise for the semantic search function) the Customer's data and content solely on behalf of the Customer and, unless expressly agreed otherwise, for use by the Customer."] Available via <https://www.retresco.com/terms-conditions/>. All websites accessed on 10 November 2021.

[116] Textomatic's Cooperation Agreement for News-Alert-System (NAS) and rob.by-Chatbot, Preamble and Definitions ['The databases of the NAS system are filled with licensed data (e.g. Tradegate/Deutsche Börse/VWD, DFB, Wetterkontor) and Open Data (e.g. Wikipedia) or with content from media partners.'] Available via< https://newsletter.textomatic.ag/en/Contract/NAS/index.html>. All websites accessed on 10 November 2021.

copyright. At the same time, the Connexun's clients are bound to allow the public use of the output in case it contains any protected subject matter.[117]

3. Interim conclusion

This empirical research demonstrates that the NLG services market is thriving. Outsourcing the development of algorithms is the standard solution in robojournalism. It comes as no surprise that the use of algorithms is generally present only in data-driven fields such as finance, weather forecasts or sports reporting. Most of the analysed service providers obscure their contractual practices. The publicly available and relevant documents almost all require the client to provide the source data and allow the use of the content without claiming any copyright interest in the input content. Indeed, it is plausible to believe that the other service providers which failed to disclose their service contracts, follow the same logic. Furthermore, although most services advertise the underlying algorithm as fully automated, the final publication of the given content necessitates some degree of human intervention in the newsrooms. Hence, the copyright protection of the relevant media outputs might effectively arise as a consequence of the potential free and creative choices made at the level of editing, after the NLG process has taken place. These choices will certainly vary widely from process to process – each newsroom is orchestrated differently, so the amount of postproduction creative effort necessary to bring the NLG output to a readable journalistic piece is not always the same in all circumstances. These practices are discussed later in this paper. For the purposes of the interim conclusion on the business reality it suffices to say that advertising a service as automated may turn out to be simple window dressing when one studies the reality in the newsroom. The algorithmic creation of contents fits perfectly into the existing copyright business logic and necessitates no extension to any external parties or to the robots themselves.

V. The implications of robojournalism on journalism

To comprehensively understand the key implications of robojournalism, copyright lawyers should also take a close look at the topic from the angle of media and communications studies. This perspective is of crucial importance, especially since those are the journalists and the news publishers who decide on whether and how they want to rely on algorithms in producing and disseminating news to the public in the first place. It has been established that the news outlet's decision to adopt automatic journalism techniques depends on two specific variables – 'expected

[117] Connexun's Terms & Conditions, API Data usage ['Data accessible through Connexun may contain Third Party Content (such as text, images, videos obtained from various news sources). This content will remain the sole responsibility of those who make it available. In some cases content accessible through our Services may also be subject to intellectual property rights. In these cases you are allowed only to perform actions and activities that are awarded to you by the owner of the content.'] Available via <https://connexun.com/terms-and-conditions> accessed 31 March 2022.

effects' and consumer receptivity.[118] The former pertains to the business performance brought about by robojournalism, while the latter centres on the willingness of customers to digest news written by robojournalists.[119] Furthermore, user expectations have a direct effect on the journalistic activities. Whatever is best suited to the needs of the clients of news portals also has implications for the creation and dissemination of news.

In other words a holistic approach is needed in deciding whether outputs of robojournalists should be subject to copyright protection. Such protection is heavily dependent on the purpose, role and the practical availability of algorithms in newsrooms. For that purpose, we reviewed the relevant (first, the European, and second, the US) media and communications studies literature to find patterns that have relevance for copyright law.[120] In the following, we will introduce the implications of robojournalism for (1) journalists; (2) publishers; and (3) readers/consumers.

1. Implications for journalists

There is an understanding among some AI researchers that the biggest threat to the development of AI is the human fear of the effects of such changes.[121] Such a 'Frankenstein Complex' is certainly also present with respect to robojournalism. Journalists inescapably meet the challenge of 'resistance versus assistance', that is, whether they believe robojournalists will replace or only supplement them.

Media and communication studies literature tends to indicate that the typical (and optimistic) reaction is that algorithms will only supplement rather than replace human authors. Usually, robojournalism is treated to be 'a means of upgrading and equipping journalism for the demands of the 21st century'.[122] This optimistic view has its roots in history: the earliest form of robojournalism, CAR, which was applied from as early as the 1950s and was at its peak around the 1970s in the US, never led to the extinction of human reporters.[123]

Indeed, the general trend among journalists is to argue that '[a]lgorithms make possible journalistic practices that would not be feasible based on human labor alone. Algorithmic systems help news sites determine quality reader comments, find important stories on social media platforms, and use data sets to generate stories'.[124] The empirical research of Schapals and Porlezza showed that

---

[118] Daewon Kim and Seongcheol Kim, 'Newspaper Companies' Determinants in Adopting Robot Journalism' (2017) 117 Technological Forecasting and Social Change 184, 188.

[119] ibid.

[120] At the same time, we will not be discussing professional questions such as the ethical aspects of automated journalism, as well as issues related to objectivity, bias or newsworthiness.

[121] Lee McCauley, 'The Frankenstein Complex and Asimov's Three Laws' (*AAAI*, 10 May 2007) <https://www.aaai.org/Library/Workshops/2007/ws07-07-003.php> accessed 9 February 2022.

[122] Bucher (n 13) 920.

[123] Seth C Lewis and Nikki Usher, 'Open Source and Journalism: Toward New Frameworks for Imagining News Innovation' [2013] Media, Culture & Society 602. See Coddington (n 18), 338, who notes that '[d]ata is similarly seen within CAR as entirely secondary, to human-oriented aspects of a story'.

[124] Matt Carlson, 'Automating Judgment? Algorithmic Judgment, News Knowledge, and Journalistic Professionalism' [2018] New Media & Society 1755, 1762.

journalists tend to defend their positions by referring to expressions like creativity, context or uniqueness to describe their work. Journalism is regularly treated by journalists themselves as 'an 'art' or a 'craft' rather than some manual task on an assembly belt'.[125] Human experience and know-how is argued to be irreplaceable,[126] especially as algorithms are only a form of programmed logic.[127] As Coddington stated, '[d]ata journalism retains an emphasis on editorial selection and professional news judgment in analysing and presenting data, but it does so while also building around a recognition that expertise in analyzing and drawing meaning from that data often exists outside of the profession, among the audience'.[128] Some estimates suggest that only about 15% of journalists' and 9% of editors' jobs might be replaced by automated technologies.[129]

Furthermore, robojournalists are usually designed to free up human journalists for more sophisticated workplace tasks,[130] allowing them the chance and time to 'produce a better story'.[131] Arguably, this refers to practices such as creative writing, investigative journalism, and clever interviewing, where the creative intellectual effort of the journalist is indispensable to the final piece. Another study found that the journalists' three key motives for using AI were: making their own work more efficient; delivering more relevant content; and improving business efficiency.[132] Each of these is directly linked to the speed and coverage that AI systems can reach. It is without doubt that NLG can generate written output extremely quickly. In addition, AI systems can process immense volumes of information allowing it to generate statistical correlations much more profoundly than human beings. One must not take all of this without the necessary qualifications – AI systems still do not make logical causal relationships between the information they process. Delivering more relevant content is therefore certainly a strong benefit of the AI system, but 'the genuine relevance' of the information gets verification from a human journalist. Similarly, American researchers found that AI is particularly helpful in three categories of activities: 'finding needles in haystacks'; identifying patterns; and providing a good subject for a story itself.[133]

Human journalists' primacy over algorithms is also connected to the abilities and qualities of AI itself. Current NLG technologies are unable to observe society and fulfil journalistic tasks, e.g. orientation and public opinion formation. In short, AI is currently capable of focusing on 'what' instead of 'why'.[134] Algorithms are able to focus on the raw data rather than the 'bigger picture',

[125] Schapals and Porlezza (n 26) 23.
[126] Bucher (n 13) 925.
[127] ibid 924.
[128] Coddington (n 123) 339.
[129] Broussard and others (n 90) 680.
[130] Graefe (n 6) 34 and 597; Schapals and Porlezza (n 26) 21-22.
[131] Lewis and Usher (n 123) 605.
[132] Charlie Beckett, 'New Powers, New Responsibilities. A Global Survey of Journalism and Artificial Intelligence' (*Polis*, 18 November 2019) 7 and 32-34 <https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities/> accessed 9 February 2022.
[133] Mark Hansen and others, 'Artificial Intelligence: Practice and Implications for Journalism' (2017) 8 <https://academiccommons.columbia.edu/doi/10.7916/D8SN0NFD/download> accessed 9 February 2022.
[134] Graefe (n 6) 597.

the context of the issue.[135] And this is where human journalists step in prominently to work with the AI.

As the interviews made by Schapals and Porlezza showed, the journalists' craft can 'best be described by linguistic eloquence, stylistic nuance and a general need to not merely convey facts objectively, but to contextualise them, that is, to take readers by the hand and help understand the deeper meanings, possible consequences and wider (societal) significance of the factual information they are consuming. [The journalists] also stressed the need for a human editor to double-check and to validate accounts of sports or financial news coverage'.[136] Finally, as Graefe pointed out, journalists should focus on tasks that algorithms cannot perform. The authors suggest that in the future human and automated journalism is likely to become closely integrated and form a relationship that Reginald Chua refers to as a 'man-machine marriage', whereby algorithms will analyse data, find interesting stories, and provide a first draft, which journalists will then enrich with more in-depth analyses, interviews with key people, and behind-the-scenes reporting.[137] As the technological reality section below will demonstrate, this is already the reality.

No doubt not all journalists are happy with the recent changes. Those with less training in technology might find their future in the news industry more vulnerable. Empirical evidence also shows the fears of gradual disappearance of data-intensive newsroom jobs,[138] especially those related to sports, weather, and financial reports.[139]

2. Implications for publishers

Graefe pointed out that '[i]n automating traditional journalistic tasks, such as data collection and analysis, as well as the actual writing and publication of news stories, there are two obvious economic benefits: increasing the speed and scale of news coverage. Advocates further argue that automated journalism could potentially improve the accuracy and objectivity of news coverage. Finally, the future of automated journalism will potentially allow for producing news on demand and writing stories geared toward the needs of the individual reader'.[140] Reading this opinion in conjunction with other sources, the key motivation of publishers in introducing NLG solutions might be the speedy creation of new products, rather than cutting human workload costs. [141] Indeed,

[135] Fanta (n 107) 10; Neil Thurman, Konstantin Dörr and Jessica Kunert, 'When Reporters Get Hands-on with Robo-Writing' (2017) 5 Digital Journalism 1240, 1246-48.
[136] Schapals and Porlezza (n 26) 21.
[137] Graefe (n 6) 35.
[138] Matt Carlson, 'The Robotic Reporter – Automated Journalism and the Redefinition of Labor, Compositional Forms, and Journalistic Authority' (2015) 3 Digital Journalism 416, 422-24.
[139] Graefe (n 6) 33-34; Schapals and Porlezza (n 26) 22.
[140] Graefe (n 6) 22.
[141] Thurman, Dörr and Kunert (n 135) 1249-50; Even South-Korean media researchers found that the top concerns of newspaper companies are, first, the business performance of their companies brought about by the introduction of robojournalism, and, second, consumers' willingness to read algorithmic news stories. Companies are found to be rather insensitive regarding the possible sunken costs stemming from the introduction of AI in the newsrooms, see Kim and Kim (n 118).

there is a sensible 'profit trap' in NLG solutions. On the one hand, publishers' struggle for profitability, and NLG solutions can reduce some transaction costs due to process automation.[142] On the other hand, collaborations between journalists and computer scientists necessitate extra resources.[143] The development expenses of robojournalism, including the hiring of trained technical experts or the internal training of them, are barriers to entry and further expansion.[144]
Another key factor is that '[c]omputers never get tired. Thus, algorithms are less error-prone'.[145] We do not believe that the latter necessarily flows from the former. Computers do crash and the code could be flawed, and the data with which the machine learning algorithm is fed could be biased and lacking in objectivity. Yet, the absence of physical and emotional tiredness of which even the keenest and most dedicated human journalists suffer makes the machine more efficient in contrast to humans. While this is a factor that publishers typically tend to consider from the perspective of users' expectations rather than from the perspective of the creation of news outputs, it must be highlighted that this accuracy and speed certainly render the use of robojournalists more attractive to publishers and should in itself be seen as a benefit.
Automated journalism is mainly limited to elite/well-resourced news organisations, and small organisations are unable to fully employ NLG solutions.[146] This can be tied to the cost of developing the necessary software, which most publishers do not have the economic capacity to do. As our empirical findings also evidenced, many NLG service providers necessitate the use of the media company (the client) to provide its own data for the generation of the output. Fanta also found that media companies are not only under-resourced but are also far behind digital innovations.[147] It is a general problem that small-sized media corporations simply do not have the necessary resources to collect publicly unavailable data that might form the basis of algorithmic content creation.

3. Implications for readers/consumers

[142] Latzer and others (n 12) 407.
[143] Carlson, 'Automating Judgment? Algorithmic Judgment, News Knowledge, and Journalistic Professionalism' (n 124) 1762.
[144] Fanta (n 107) 11.
[145] Andreas Graefe and others, 'Readers' Perception of Computer-Generated News: Credibility, Expertise, and Readability' [2018] Journalism 595, 597.
[146] Schapals and Porlezza (n 26) 18; The same experience is present in the US, see Hansen and others (n 133).
[147] Fanta (n 107) 15.

The increasing potential of NLG has led to rising user expectations. Such expectations are related to the quality of news,[148] transparency,[149] trustworthiness of robojournalists,[150] the personalisation of media coverage[151] or 'news on demand'[152] among many others. The importance of these values becomes even greater. This is essentially because NLG algorithms can generate outputs that the readers/consumers identify with human messages.[153]

At the same time, there is a perceptible danger of an 'information overload'.[154] It is more than a hypothesis that robojournalism multiplies 'the number of available stories well beyond present limits'.[155] There is a significant risk that '[t]his expansion of stories necessarily reduces the odds that any single story will be read'.[156] Tied to this is the well-known danger of not being able to determine the authenticity and trustworthiness of information, as well as the possibility of falling into a filter-bubble.[157] If so, the negative externalities of NLG-based news production can heavily outweigh the benefits of robojournalism.

4. Interim conclusion

Journalists seem to primarily think 'about' rather than think 'with' algorithms.[158] The 'point of no return' is not here yet. Computing can support journalists to focus on in-depth, investigative activities that give them a competitive advantage,[159] rather than taking over their creative role. Human workload (both at the writing and the editorial level) is and (except for certain fields such

[148] Graefe (n 6) 40.

[149] Thurman et al. have empirically shown that journalists also favour transparency, see Thurman, Dörr and Kunert (n 135) 1252. Graefe (n 6) 36-42; As Fanta pointed out, however, 'not all use of automation is made transparent to customers and readers. Reuters, AP and NTB usually tag their robot stories, However, this does not apply to single-line alerts, so-called snaps, which Reuters sends out. At least two news agencies produce partial stories from templates without mentioning the robot as a co-author', see more at Fanta (n 107) 11.

[150] Inge Graef, Raphaël Gellert and Martin Husovec, 'Towards a Holistic Regulatory Approach for the European Data Economy: Why the Illusive Notion of Non-Personal Data Is Counterproductive to Data Innovation' (2018) TILEC Discussion Paper No 2018-029 599 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3256189> accessed 31 March 2022.

[151] As Graefe says, to '[t]ell the same story in a different tone depending on the reader's needs', see Graefe (n 6) 22.

[152] ibid 27.

[153] Christer Clerwall, 'Enter the Robot Journalist – Users' Perceptions of Automated Content' [2014] Journalism Practice 519.

[154] Graef, Gellert and Husovec (n 150) 596.

[155] Carlson, 'Automating Judgment? Algorithmic Judgment, News Knowledge, and Journalistic Professionalism' (n 124) 1763.

[156] Carlson, 'Automating Judgment? Algorithmic Judgment, News Knowledge, and Journalistic Professionalism' (n 124) 1763.

[157] Eli Pariser, *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think* (Penguin 2012).

[158] Bucher (n 13) 927-29.

[159] Lewis and Usher (n 123) 606.

as sports, weather, and financing) will remain fundamental and inevitable, at least in the near future, and probably longer.

News publishers certainly see possibilities in NLG services, but financial considerations play a frustrating role in this regard. Due to the large quantity of resources needed to set up a functioning robojournalism newsroom (including the building of human-robot collaboration in the creative phase), only bigger media corporations are currently in a position to take even the first steps to switch to NLG solutions. At the same time, cost reduction seems to remain a daydream, which is another reason for small players to think twice before investing in robojournalism.

It is not yet possible to measure whether the externalities of robojournalism will be mainly positive or negative for users. As a general consequence, however, we can conclude that bulk news consumption, in conjunction with the generational shift towards tweets or Tik-Tok videos rather than in-depth writing,[160] might contribute to a substantial devaluation of journalism.

Taking into account the long-lasting need for human involvement in news creation, the limited switch to NLG by the bigger media corporations, and the hardly predictable outcomes of robojournalism for users, we argue that there is no convincing evidence in media and communications studies to introduce copyright protection of automated news for the benefit of artificial intelligence or its developers.

VI. Conclusion/recommendations

This paper looked at the implications of robojournalism from the perspective of copyright law. It studied the techniques of NLG as applied to journalism and established that there may be several stages in the process where there is room for free and creative choices that would trigger a valid copyright claim. Yet, this should not be taken at face value. Most of the journalistic fields in which NLG is applied relate to rather dry, data-heavy, fact-based fields such as sports, weather, and finance. It is therefore questionable whether it could actually trigger a valid copyright claim, even if the journalistic output in those fields was written by a human author, completely excluding the presence of any NLG system. Basic principles of copyright law dictate that what is subject to protection are the expression of ideas and facts belong to the public domain. Additionally, from the perspective of business, developing an NLG system is particularly costly. This is backed up by the empirical analysis underlying this paper which proved that outsourcing the development of NLG – due to the lack of resources and/or the lack of expertise – is the standard practice. Looking into the practices in the editorial room it appears that postproduction plays a significant role. Therefore, at the end of the day even backed up with NLG processes, news editors are strongly in control of the output that they communicate. Finally, from the perspective of journalists, publishers, and readers, it appears that robojournalism is already making a huge impact. While NLG costs for news publishers are rather high, journalists are adapting to work with algorithms to

[160] Christian Montag, Haibo Yang and Jon D Elhai, 'On the Psychology of TikTok Use: A First Glimpse From Empirical Findings' (2021) 9 Frontiers in Public Health 1-6.

meet the demanding expectations of consumers, while still balancing important values such as transparency and news quality.

The three perspectives studied in this paper – technological, business, and media and communications – demonstrate that copyright law is not to be extended to cover output generated by NLG. The current copyright framework is rooted in the presence of a human author and that should remain so. The absence of free and creative choices should not be artificially compensated for by considerations of potential market failures if copyright protection does not arise for robojournalism output. It can be concluded that robojournalism follows well the negative spaces theory.[161] Being the first one to utilise generative techniques that are trustworthy, transparent, accurate, zeroing discrimination brings sufficient benefits to companies resorting to NLG techniques even in the absence of intellectual property, especially copyright protection.

[161] Chris Sprigman and K Raustiala, 'The Piracy Paradox: Innovation and Intellectual Property in Fashion Design' (2006) 39 Cardozo Arts & Entertainment Law Journal 535, 538, according to which certain creative fields thrive regardless of the protection of intellectual property.