AI and Ethics (2023) 3:427–439
https://doi.org/10.1007/s43681-022-00169-1

ORIGINAL RESEARCH

Rethinking data infrastructure and its ethical implications in the face of automated digital content generation

Maria Joseph Israel[1] · **Ahmed Amer**[1]

Received: 2 March 2022 / Accepted: 26 April 2022 / Published online: 25 May 2022

Abstract
Recent AI developments have made it possible for AI to auto-generate content—text, image, and sound. Highly realistic auto-generated content raises the question of whether one can differentiate between what is AI-generated and human-generated, and assess its origin and authenticity. When it comes to the processes of digital scholarship and publication in the presence of automated content generation technology, the evolution of data storage and presentation technologies demand that we rethink basic processes, such as the nature of anonymity and the mechanisms of attribution. We propose to consider these issues in light of emerging digital storage technologies that may better support the mechanisms of attribution (and fulfilling broader goals of accountability, transparency, and trust). We discuss the scholarship review and publication process in a revised context, specifically the possibility of synthetically generated content and the availability of a digital storage infrastructure that can track data provenance while offering: immutability of stored data; accountability and attribution of authorship; and privacy-preserving authentication mechanisms. As an example, we consider the *MetaScribe* system architecture, which supports these features, and we believe such features allow us to reconsider the nature of identity and anonymity in this domain, and to broaden the ethical discussion surrounding new technology. Considering such technological options, in an underlying storage infrastructure, means that we could discuss the epistemological relevance of published media more generally.

Keywords Ethics of electronic publishing · Immutable data storage · AI-automated content generation · Authorship and authentication · Explainable AI · AI and machine ethics

1 Introduction

In contrast to the traditional publishing scenario, we are entering into a world, where software agents take more prominent roles, and contribute more substantially to the production and evaluation of written content. The rise of AI-generated[1] textual content is driven by advances in Natural Language Processing (NLP) that give machines an increasing ability to read, and derive meaning from, human language [92]. Of particular note is the Generative Pretrained Transformer[2] (GPT) that has been used by *OpenAI* as a general-purpose language algorithm to translate text, answer questions, and predictively write high quality text.[3] This GPT technology has been widely adapted with applications ranging from routinely generating news articles, to producing creative literature and fine arts, and anything in-between, including email subject lines, blog content, and marketing content, such as advertisements and product descriptions. Some natural language generation tools that can generate entire blocks of text based on brief writing prompts include (not an exhaustive list): *AI-Writer*,[4] *ContentBot*,[5]

✉ Maria Joseph Israel
misrael@scu.edu

Ahmed Amer
aamer@scu.edu

[1] Department of Computer Science and Engineering, Santa Clara University, Santa Clara, CA, USA

[1] The term "AI-generated" is interchangeably used with other terms like computer/software-generated and machine-generated.

[2] Currently, GPT-3 is the third generation of such a system, developed under OpenAI research and deployment company. Its newest edition was introduced in May 2020. For more information, see https://openai.com/.

[3] An example of AI-generated non-fiction text is discussed at https://www.gwern.net/GPT-3-nonfiction.

[4] AI-Writer: https://ai-writer.com/.

[5] ContentBot: https://contentbot.ai/.

Jasper,[6] *Nichess*,[7] *ShortlyAI*,[8] *Text Cortex*,[9] *WriteSonic*,[10] *Automated Insights*,[11] etc. Some of these tools can also be used to generate content at scale, e.g., to the extent of 2000 news stores per second in one case [63]. Mainstream news agencies, such as *The New York Times* and *The Associated Press in the USA, BBC* in the UK, *La Monde* in France have already started utilizing GPT tools to regularly generate news reports [23, 79, 87]. This is not just limited to English language content, but has also extended to other languages, such as Russian, Japanese, French, and Chinese [90]. GPT technology has also been used to generate creative essay writings and poetry [74].

Going beyond the success of GPT in NLP, it has been further extended to processing text and images simultaneously, bridging the domains of computer vision and language processing, not simply for automating object identification and description from images, but for the generation of images, effectively converting natural language description to images depicting what is described. To demonstrate this, *OpenAI* released a multi-modal version of GPT3, called DALL-E, that demonstrates surprisingly effective text-to-image conversion capabilities [18, 22, 67], music creation [30, 49], and radiology reports [5].

With the advent of ever more capable software agents contributing directly or indirectly to our sources of information, it's important to look at not only information dissemination but also the challenges of information attribution. Current electronic publication and scholarship processes and practices are faced with opposing values of anonymity and attribution of authorship that need to be reviewed in light of AI-generated content. Approaches vary in addressing the challenges of ensuring transparency and trust in the digital publishing domain. In most academic publishing today, approaches to author identification can run the gamut, from overtly emphasizing anonymity of authors (as in the "*Journal of Controversial Ideas*") [70, 77],valuing the traditional double blind review process [64],to models of more radical transparency (such as *Open Peer Review*[12]) [71].

Anonymity of authors vs. reviewers differ in their intended purpose, but have traditional purposes that do not necessarily translate smoothly to a world, where content generation can muddy the definition of authorship. Authors' anonymity provides authors' pseudonymity [45] to achieve 'academic freedom' [77], while the blind review ensures anonymity of reviewers to promote 'fair reviews and academic quality/integrity' both during and after the peer review processes [48]. Though these approaches seem to work in the traditional publication context, they cannot meet the complex situations that arise with AI-mediated textual/multimedia content.

According to *OpenAI*'s own statement, some of the potential good use cases of their text generation technologies include: AI writing assistants, more capable dialogue agents, unsupervised translation between languages, better speech recognition systems [65], and better integration of language and vision [67]. On the other hand, this technology also poses threats similar to those posed by "deepfake" media, i.e., phenomena in which it becomes increasingly difficult to discern truth from falsity [56]. Potential malicious use of auto-generated fake news could be used to harm democratic societies through social media manipulation, as was brought to the public eye through events, such as the *Cambridge Analytica* scandal [40], and can even leave some wondering whether the technology did anything good at all for the society [60]. Other threats include, as identified by *OpenAI*: generating misleading news articles, impersonating others online, automating the production of abusive or faked content to post on social media, automating the production of spam/phishing content [65]. Therefore, we must look carefully at the improvement and exploitation (nature and vulnerabilities) of the existing digital communication, authentication, and publication infrastructure. Here we can think of AI in place of an author or an artist, and the whole issue of authorship when facing a potential flood of auto-generated text by the likes of GPT-3 [14, 54] or auto-generated images by the likes of DALL-E [67], and other worries brought about by AI development.

It is hard to address these emerging problems if these AI tools are unleashed on a publishing infrastructure that does not employ digital publishing technologies as robustly, and as effectively, as it should (let alone as it could). For example, if the world is exploitable by anyone with an email address, or social media handle, pretending to be anyone else, then this problem is dramatically compounded when we can no longer assume that the authors of posted content, or the manipulators and disseminators of digitally altered content, are human.

The presumptions and assumptions surrounding human authorship do not always, and increasingly will not, continue to hold in the realm of digital publishing, and even scholarship. How does one attribute authorship when an AI-mediated system copies and/or mixes texts from others' works? We doubt that solely focusing on the detection of altered or synthetic content, as is the main focus of work in this area today [36, 47, 93], will ever be a secure and sustainable long-term solution to this problem. Therefore, digital publication infrastructure has to have more robust accountability, attribution, and transparency mechanisms—mechanisms

[6] Jasper.ai: https://www.jasper.ai/.

[7] Nichess: https://nichesss.com/.

[8] ShortlyAI: https://www.shortlyai.com/.

[9] Text Cortex: https://textcortex.com/.

[10] Writesonic: https://writesonic.com/ai-article-writer-generator.

[11] Automated Insights: https://automatedinsights.com/.

[12] https://plos.org/resource/open-peer-review/.

such as immutable data storage with provenance tracking and strong authentication protocols. In other words, in the face of more automated data generation, it is imperative that storage, curation, and presentation of that data should become less dependent on manual processes that were built upon aging infrastructure, and we need to consider the ethical implications of this situation. In this paper, we discuss these issues, and illustrate how deficiencies in our current approaches can be addressed, through an example system model, *MetaScribe*, which is a system aimed at tracking provenance of digital scholarship and preserving this information in immutable data stores.

This paper is structured as follows: In this section we've introduced how digital technologies demand us to rethink our technological positions regarding authentication, attribution, and transparency in digital publishing. Sections 2 and 3 will offer a brief background on machine moral agency and explore related works, respectively. Section 4 will discuss various possible views on the authorship of computer-generated works. Section 5 will deal with digital technology and its responsibility in publication. Section 6 will elaborate on the *MetaScribe*[13] model, a work-in-progress project, as a novel solution to address techno-social aspects of AI in digital publication. In Sect. 7, we will conclude the paper with a discussion of future work.

2 Background: machine moral agency

With AI increasingly generating not just more possible (as human made) content but also serving as a potentially primary means of exploring problem specifications and solutions' probability, the question of where the human originator of content ends and the algorithmic creator of content begins, evokes questions reminiscent of such concepts as '*The Turing Test*' by Alan Turing [81] or '*The Chinese Room Argument*' by John Searle [75]. While the *Turing Test* deals with the question of whether machines can think, the *Chinese Room Argument* argues against the meaningfulness of such claims. While a detailed discussion of these thought experiments is beyond the scope of this study, instead, we focus on the role of the (value-sensitive) design of underlying infrastructure of an AI technology in addressing these moral issues. To set a background for this, we propose to discuss *Machine Moral Agency*[14] to guide discussion of AI ethics in the digital publication/scholarship domain, as machines are increasingly powerful in generating synthetic content that is difficult to attribute to its original source. *Machine Moral Agency* is understood as the capacity for technologies to consider notions of right and wrong and to act on those distinctions. That is to say that artificial agents/entities can 'think' and are capable to do wrong, and may possibly be considered responsible for such wrongdoing [1, 2, 32, 73, 88]. The case for *Machine Moral Agency* has been supported by two factors: (a) the increase in technical capabilities enabling machines to operate autonomously in increasingly broader domain applications (termed as *Domain-Function* by [62], and (b) these functional capacities embody the morally relevant abilities, such as autonomy, intentionality, responsibility, and sensitivity, etc., found in human moral agents (termed as *Simulacrum View by* [62].

In other words, AI-enabled systems have a profound impact on our lives, as they can make decisions that have a moral dimension. Therefore, there is a considerable debate on the questions of how, and to what extent, such artificial moral agents should be included in human practices normally attributing full moral agency and responsibilities to participants. Machine ethicists seriously argue that machines should be given the power to make moral decisions on the premise that machines are deployed in situations in which they make decisions that have a moral dimension [2, 32, 73, 88]. Hence machines should be extended with moral sensibility in situations in which they inevitably find themselves. They also contend that human and artificial morality will be different. They further argue that there is no reason to rule out artificial morality *a priori*, and it is a worthwhile attempt to define and construct such artificial morality.

In contrast, some would argue that mental states, emotions, and social skills are also necessary for moral behavior, and therefore, question whether machines can be said to be moral if they are lacking in this respect [15, 39, 76, 80]. Still others suggest to build machine moral agents that integrate societal expectations about the ethical principles that should guide AI behavior, as surveyed by [4] through their 'Moral Machine' experiment. There are others who critique such suggestions [42, 86] There are other questions that include

[13] The name "*MetaScribe*" is a commonly used term. For example, Fabrice Kordon uses this to denote his work, an Ada-based tool, for implementing program generators. For more details, see: Kordon F. (1999). MetaScribe, an Ada-Based Tool for the Construction of Transformation Engines. In: González Harbour M., de la Puente J.A. (eds) Reliable Software Technologies — Ada-Europe' 99. Ada-Europe 1999. Lecture Notes in Computer Science, vol 1622. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-48753-0_27; Regep, D. & Kordon, F. (2000). Using MetaScribe to prototype a UML to C++/Ada95 code generator, *Proceedings 11th International Workshop on Rapid System Prototyping. RSP 2000. Shortening the Path from Specification to Prototype (Cat. No.PR00668)*, pp. 128-133, https://doi.org/10.1109/IWRSP.2000.855209. This term is also used for an online digital multimedia management platform at https://metascribe.io/. Our usage of the term "*MetaScribe*" refers to the proposed provenance (*Metadata*) tracking system in the scholarly publication domain (*Scribe*). This system is, in fact, an implementation of a use case using the generic provenance framework, *MetaScriptura*.

[14] Machine moral decision making is referred to by a number of names including machine morality, machine ethics, artificial morality, and friendly AI.

whether machines can be assigned rights and duties? Will fear, shame, or punishment have any meaning to machines if they are assigned rights and duties? How is a machine to attribute meaning to a textual/visual cue that it perceives? For example, viewing a social media *like* button that was clicked by Bob for a friendly comment on the topic of immigration may be an act of encouragement, but it could also be viewed as an act of insensitivity if the comment derides immigrants. Knowing how to interpret such comments/contexts, and how to classify them from a moral perspective, is key to creating any kind of moral machine.

Besides these questions, we would like to raise a moral question regarding machine-assisted human agency: Is limiting the use of such automation an unkind act, of denying assistance to those who cannot write as well without the aid of the machine? These kinds of questions can have a big impact on engineering moral decision-making machines.

3 Related work

We are at the "algorithmic turn" [59, 82] witnessing the application of AI algorithms that possess self-learning and autonomous complex decision making abilities at various levels. AI algorithms now have found their way into auto-generated content as they are capable of replacing humans in performing many cognitive tasks regarding auto-generated news content [27, 46]. There are several use case scenarios of auto-generated content and analysis in the areas of automated/robot journalism, games, e-learning, etc. Automated journalism is conceptualized here as "algorithmic processes that convert data into narrative news texts with limited to no human intervention beyond the initial programming" ([17], p.417). The earliest automated journalism was demonstrated in the *TaleSpin* software [55] transforming raw data into intelligible language. The essence of automated journalism is the automation of storytelling [7, 16] as a sequence of continuous narratives [34]. The steps in automated journalism include: locating and identifying relevant data in databases, categorizing the data into key facts while also prioritizing, comparing, and aggregating the data, and then organizing it in a semantic structure of narrative, and finally publishing it as a journalistic output to the public [34, 89]. In other words, in automated journalism, algorithms typically characterize stories that use numbers, such as sports analysis, real-estate market analysis, weather forecast analysis, earnings previews, etc. [17]. Some of these systems include *Robotorial*, *OpenAI*'s[15] *Generative Pretrained Transformers* (GPT-2 and GPT-3), and Deep Learning Networks.

Many studies explore the effects of automated journalism on the human journalists' practices [17, 58, 60, 83, 91] and the effect on readers' and journalists' perceptions [20, 84], but few studies addressed the authorship of synthetically generated content when it copies content from another article violating the journalistic code of ethics. These studies considered two important factors: disclosure transparency and algorithmic transparency. Disclosure transparency deals with the process of revealing how a particular news item is selected and produced, while algorithmic transparency is concerned with the actual process of selecting, constructing, and producing a news item using an algorithm [58].

To underscore the degree of algorithmic involvement in the creative process, and the journalist–machine relation they form, there are two major levels of algorithmic involvement in automated journalistic content creation that include: algorithmic content generation, and integrative content generation. In the former, the textual content is produced without the involvement of a human person (journalist/editor), the latter deals with the textual content generated through the collaborative efforts of a generative algorithm and one or more human persons. A generally accepted good practice is to acknowledge this scenario via an attribution policy that credits the real nature of algorithmic content while describing the software vendor or a programer's role in the organization, and also details the data sources of the particular story and the algorithm methodology [6, 52, 68, 69, 78].

To address the potential problems associated with algorithmic news and automated content, an obvious means is to build systems to detect such content. One way to achieve this, is to differentiate it from human-authored content, through technological development of detectors in the form of journalistic robot algorithms [17, 58]. Some examples include: *Amazon*, *Microsoft*, *Facebook* and the non-profit coalition *Partnership on AI* launched the "*Deepfake Detection Challenge*" [28], to build innovative new technologies that can help detect deepfakes and manipulated media; Similarly *Google* released a large database of deepfakes videos to help in researching detection tools [29]. However, research shows that there is no perfect solution. These efforts mainly focused on the visual materials, and lacked efficiency in determining origin text and attribution to AI-generated text. Moreover, some of these tools are based on the same technologies that allowed the media to be created in the first place and most of them still need to be paired with human intelligence to properly identify this type of content [60]. The inevitable deficiency of detectors, in addition to their being imperfect, is compounded by the interplay of generator vs. detector in an escalating arms race, always pushing the inaccuracy of the detector and the persuasiveness of the automated content. Probably a better solution may be to consider the system within which the content is stored and deployed. To that end we need to consider the metadata of authorship and attribution, in other words, the tracking of provenance.

[15] OpenAI: https://openai.com/blog/openai-api/.

This allows us to consider the broader implications of automated content generation, and to revisit the fundamental nature of authorship and citation. On the question of anonymity vs. attribution of authorship and the blind review process, safeguarding academic integrity requires novel solutions. Here the proposed system, *MetaScribe*, is handy to demonstrate how the right data storage infrastructure may help address the problem of authorship and attribution of synthetically generated content. Before we discuss the proposed model, it will be important to take a closer look at the legal–philosophical–ethical views of authorship of computer-generated works (CGW).

4 On authorship of computer-generated works

What is "computer/synthetically generated content"? It refers to content produced by an AI system that is capable of interpreting external data currently, to learn from such data, and to use that learning to achieve specific tasks through flexible adaptation [38]. Therefore, a computer-generated work could be understood as either of the following:

- Written with the assistance of an enabling device (like a word processor), *vs.*
- Generated via algorithms and code and data-sets provided by a third party.

We are referring to the latter situation, as the mere assistance of hardware and software in the transcription of text is not a matter that creates confusion as to the creative origins of the text generated. And so, the question arises as to how authorship can, or should, be attributed when a mechanism has contributed to the generation of the textual content itself, and not merely its superficial form and format.

We identify at least four possible ways of reasonably interpreting the involvement of machines/algorithms. They are:

1. As a creative contribution by its (the algorithm's) creator to the final content.
2. As an enhanced tool, with full attribution going to the author who used the tool.
3. As an enhanced tool, with partial attribution going to the author, and partial to the creator of the tool (perhaps determined by a licensing agreement for use of the tool).
4. If we are willing to consider the question of personhood, or virtual/corporate ownership, by the tool/machine, this would give us further possibilities to consider, which would include the tool as an entity capable of holding ownership/responsibility.

4.1 Legal views on computer-generated works

Legal considerations largely prompt us to focus on the question of copyright, and to dismiss the issue of machine personhood as typically a transitive conduit when it comes to rights, benefits, and liabilities. The first choice above (interpreting content-generation tools as offering a creative contribution by their author), gives great legal pause, as it would suggest that the creator of the tool has some claim to the final output, while the second choice (fully attributing creative authorship to the user of the algorithm) would relegate such a contribution (by the algorithm creator) as no more than a work-for-hire, effectively commissioned (and fully owned) by the author. However, there is more than a question of copyright, and the ownership of intellectual property, as there is the legal question of liability, and a broader ethical question of moral responsibility for content. Therefore, let us focus on the more limited legal questions, and consider, where the legal landscape currently lies in relation to the issue of copyright and machine authorship.

The legal status of computer authorship gradually evolved towards maturity as more extensive AI research gained momentum [12, 13, 46]. Prior to the current legal status of computer authorship, the United States Register of Copyrights distinguished between the cases of using the computer as merely an assisting tool, and the cases of using the computer in the traditional sense of conceiving and executing authorship not by a human person but by a machine [11]. Eventually computer authorship was recognized as true authorship first in the US National Commission on New Technological Uses of Copyrighted Works (CONTU Report, 1976 and then in the Berne Convention, 1988, and the subsequent enactment of the Copyright Act of 1976 [41, 57]. The United Kingdom went even a step further in enacting a copyright statute that makes it irrelevant whether a computer-generated work owes its origin to a human author [85]. The British copyright law states that if a work is produced by a computer rather than by a person, the law simply confers the copyright upon the human being who is responsible for the computer's creation of the work [85]. However, the US Copyrights, Designs, and Patents Act is not definitive in distinguishing authorship of computer generated works when it states that "person by whom the arrangements necessary for the creation of the work are undertaken" [6], p. 222). The Australian Copyright Act (1968) requires the identification of a human as the 'author' of a work. However, some cases have pronounced certain computer-produced output as authorless and, therefore, demonstrate the tension between computerized methods of producing works and the requirement that a copyright work have a human author [52].

Therefore, while it may seem that UK copyright laws have a clear view of authorship in the presence of machine generated content, this is solely in regard to the issue of copyright

and intellectual property, and leaves open considerable questions regarding liability and responsibility, while the majority of the world have no clear definitions of legal standards on this issue. What is needed, as observed by Rajan [66], is the legal clarity on the algorithmic authorship as there is no universal formula for determining AI authorship. In other words, this is a largely unresolved and potentially contentious issue for debate. As such, it is increasingly important to discuss it in a coherent engagement with both the ethical issues it raises, and the technological infrastructure upon which it is (or can be) built.

4.2 Philosophical–ethical views on computer-generated works

The question of computer-generated works' (CGW) authorship could be considered in terms of philosophical–ethical views: originality, intentionality, and creativity criteria. These terms can be understood in the following manner: originality—"the overarching standard of authorship" [35], p. 2002), intentionality—the idea of intention to generate the work; creativity—"ability to generate novel and valuable ideas" [10], p. 24). As per the term originality, the author is considered as the "source of originality" [52], p. 935), as the term is used to refer to the works which are not copied from previous works and are made with minimal effort and expertise. However, the correlation between authorship and originality is inadequate in the case of automated journalism, because the minimal effort in the automated journalism is reduced to the mere decision to generate content rather than true intellectual effort. Similarly the relationship between authorship and CGWs is problematic as the question of intentionality is to be understood as algorithm-initiated or merely scheduled by the programmer to generate the work or the ability to predict the output, are still open to debate [12, 53]. On the question of creativity, it is even more controversial. At the core of the creativity notion, as emphasized by Davis [25], is human-action. Yet, the same idea is understood differently by philosophers, ethicists, and ordinary folks. On the one hand, it is construed to be an inspirational, imaginary, or free-of-rules process, and on the other hand, it also achieves a rational goal-directed one, as stated by Paul and Kaufman [61]. Berglez and Markham argue that it is an essential aspect of modern journalism [9, 51]. When creativity is viewed as something that is different from what went before, Boden's differentiation of creativity into three categories may be of some help here [10]. They are: combinational, exploratory, and transformational. It is combinational when it is the "unfamiliar combination of familiar ideas" [10], p. 24) such as painting collages or combination of varied translated text of an original poem; it is exploratory when it gives occasion to "potential possibilities or limitations in a conceptual space" [10], p. 24), such as inventing a new cuisine or drawing an architectural design of a conceptual building; it is transformational when it transforms its perceptual space by either "altering or dropping one or more of its defining dimensions" [10], p. 25).

The question is then what aspects of the creative process(es) could be attributed to algorithmic authorship. There is no consensus on this among the scholars and they are divided on several aspects of algorithmic creativity including the ability of AI algorithms to be creative [10] and the need for incentives, such as copyrights protection to be given to machines [21, 72]. As automated journalism challenges fundamental aspects of computer authorship, the debate becomes even more relevant in light of ongoing research extending AI capacities. And so, the debate on computer authorship is fiercely fought. At one part of the spectrum of the CGWs debate, there is a push for accepting algorithmic authorship in light of the evolving automatic capabilities of AI in media content generation. At the another end of the debate is an equal push for the primacy of anthropomorphic authorship in automated journalism as the current practice on the disclosure and transparency of authorship attribution is tailored to a human journalist. The lack of legal clarity on computer authorship further complicates this debate. In light of the computer authorship controversy, instead of validating 'truth' on either side of the debate, we focus on the most neglected aspect of this critical authorship debate: the underlying infrastructure in digital publication. To this end, we discuss *MetaScribe*, a model that focuses on the underlying infrastructure that aids in recording of content that accommodates both scenarios without bias.

5 Digital technology and responsibility for publication

If we do not make our content ecosystems more robust in the face of increasingly sophisticated software agencies (more capable AI systems), are we not abdicating a measure of responsibility for the harm such software can cause? In other words, if we live in a world that is increasingly dangerous, is the burden to act for good solely on the shoulders of those developing the potentially harmful systems, or is the burden shared by those who, while not directly building the AI systems, are designing and maintaining the infrastructure upon which they will be unleashed? Do we not have a responsibility to restructure existing systems to mitigate the harms of disruptive new technologies, and to be more resilient in the face of such potential harms? To illustrate this, we can use an analogy—the current threat of deepfake multimedia content on social media [26, 43]. One may embrace the reality of more prevalent deepfake software, and may demand that such software be constructed in a manner that makes it

detectable, or expend endless resources in a detection and detection-defeating arms race. Regardless of which approach is used, there remains an increasing responsibility to consider mechanisms that track authorship and authenticity as a means of mitigating the harm of deceptive content, i.e., holding accountable the sources of such deepfake videos in the first place.

In this analogy, depending solely on deepfake detection would mean that we would have foregone pursuing a potential solution (by insisting on a particular way of addressing the problem that is not complementary to the technology, i.e., by focusing on a narrow, less holistic view). We used this analogy here to force us to think about this ethical question: Is it upon individual technology developers, or is it upon society (as a system within which the technology works), to address the harms of new AI technologies? We used the example of deepfake videos to stress that this is a current problem within the digital dissemination of information, and our focus is on the processes of scholarship because of the existence, within that domain, of a strong expectation (if not a contemporary or consistent mechanism) of establishing and evaluating provenance. We will, therefore, look at one possible system (*MetaScribe*) in the next section.

6 *MetaScribe*

We consider the general case of disseminating and publishing information in the digital age by discussing one proposed infrastructure, the *MetaScribe* model (a proof-of-concept system in early development) as an example of how an easily overlooked systems' issue, specifically the possibility of building trusted provenance infrastructure, can be helpful. To this end, we particularly focus on the implications of provenance and trust in enhancing the integrity of, and trust in the integrity of, disseminated or stored data.

6.1 *MetaScribe* architecture

MetaScribe is a general provenance framework [3, 8, 24, 33], used to capture richer (and extensible) metadata in unstructured textual data sources, such as scholarly publications including literary texts, commentaries, translations, and digital humanities. Specifically, it's intended to demonstrate the feasibility of capturing and representing expressive provenance metadata (including context, scope, and such additional attributes as an author's declared or implied intent in citing another work), while also supporting subsequent tagging/addition of such richer metadata by third parties, be they human or automated. In pursuit of this goal, we also ensure that *MetaScribe* is not simply a bibliographic citation application, but that it offers an architecture for a more general data storage system, one that is suitable for immutable storage [19] and supporting flexible authentication schemes. In other words, *MetaScribe*, being a holistic system that both tracks provenance and provides the necessary infrastructure to do so effectively in a trusted manner, is a good vehicle for exploring the ethical questions concerning attribution, authentication, and authorship. In particular it allows us to consider alternative means of dealing with the veracity of the data being presented by this platform, and an alternative means of identifying auto-generated texts (data) and annotations, and does not strictly limit us to its intended usage domain (of digital scholarship evaluation and dissemination).

Originally intended for the application domain of digital scholarly publication, and specifically, where the focus is on preserving accurate and semantically rich representations of citations and attributions, results in a system that includes a trustworthy data repository. Such a repository is provided in the form of an immutable secure database. For the original *MetaScribe* proof of concept, we limited three parties, as shown in Fig. 1, to the following: the original authors who prepare and publish their manuscripts with their research claims supported by their usage of external citations/references; users who can record their views on the cited works in a research paper; and software agents that attempt to automatically extract semantics of cited works in a paper (e.g., the purpose of comparing authentic interpretations of cited works by third-party humans with third-party software agents.)

We start by defining what provenance means in the context of the *MetsScribe* use case and also present an abstract schema of the framework that is designed to achieve the stated goals of deep tracing and preserving authenticity and authorship of auto generated and/or annotated text. To characterize the model we define data provenance of a data object as the documented history of actors, processes, operations, communications, user access controls, and preferences related to the creation and modification of data objects. Subsequently, the relationships between provenance entities form the provenance graphs of the data objects. Figure 2 shows the proposed *MetaScribe* architecture consisting of four layers with entities and their interactions among them in each of the layers. The four layers are: User Application Programming Interface (User API), Data Model, System Software Infrastructure, and Data Storage. Each layer is intrinsically linked either to the layer above and/or below it by transforming data objects and transferring them onto the next layer.

The User API receives inputs of unstructured textual data objects, specifically scholarly articles and passes them to the data model layer for data operations. It also provides an interface to display user's query results, such as a citation's provenance annotations, as shown in Fig. 1. The data model layer accepts documents and queries regarding those

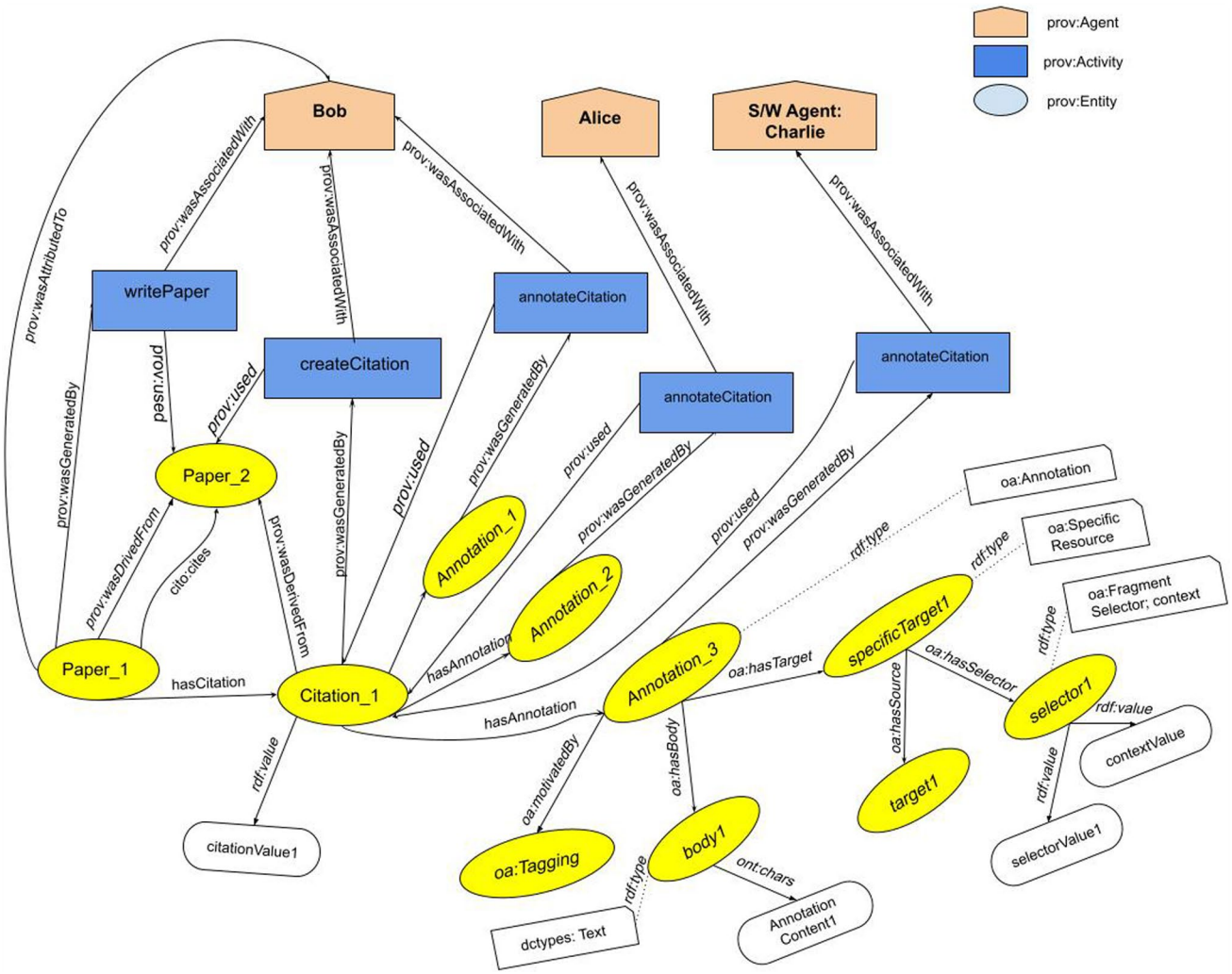

Fig. 1 A section of *MetaScribe* data model of provenance and usage scenarios in scholarly publications

documents and returns results. Therefore, this layer deals with the task of transforming unstructured textual documents into machine readable data entities based on existing bibliographic citation ontologies, and semantic web technologies and data models. At this layer, data creation or manipulation is performed by a series of operations initiated either by a user or a process as shown in Fig. 1 on various scenarios of *MetaScribe*. Particularly, the semantic metadata preprocessor engine performs text analysis, identifies relevant entities and metadata, and represents them in RDF/XML formats. For execution of SPARQL queries, it fetches the necessary data from the data storage layer and prepares the data in the format suitable for SPARQL queries. The systems software infrastructure layer engages with files and records and provides guarantees and checks to confirm any modification and verification of data and provenance metadata. Therefore, the systems software infrastructure layer is built upon a layer that provides a client–server records access. This layer supports data operations including data extraction, logging provenance metadata records, and immutable data storage services. Immutability of data is archived through a lightweight application of blockchain technology. A crucial aspect of this layer is to associate each data object and relevant data operations to their lineage/provenance records. The data storage layer accepts requests for storage and retrieval of data. Accordingly, this layer deals with the raw random access storage and provides a structure for storing records and assists in retrieving those records both locally and remotely in RDF stores across distributed systems. In this system, user activities are monitored and recorded, and data objects also could be tagged with annotations in realtime (with timestamps) that help deep trace the lineage of data elements of interest, when needed.

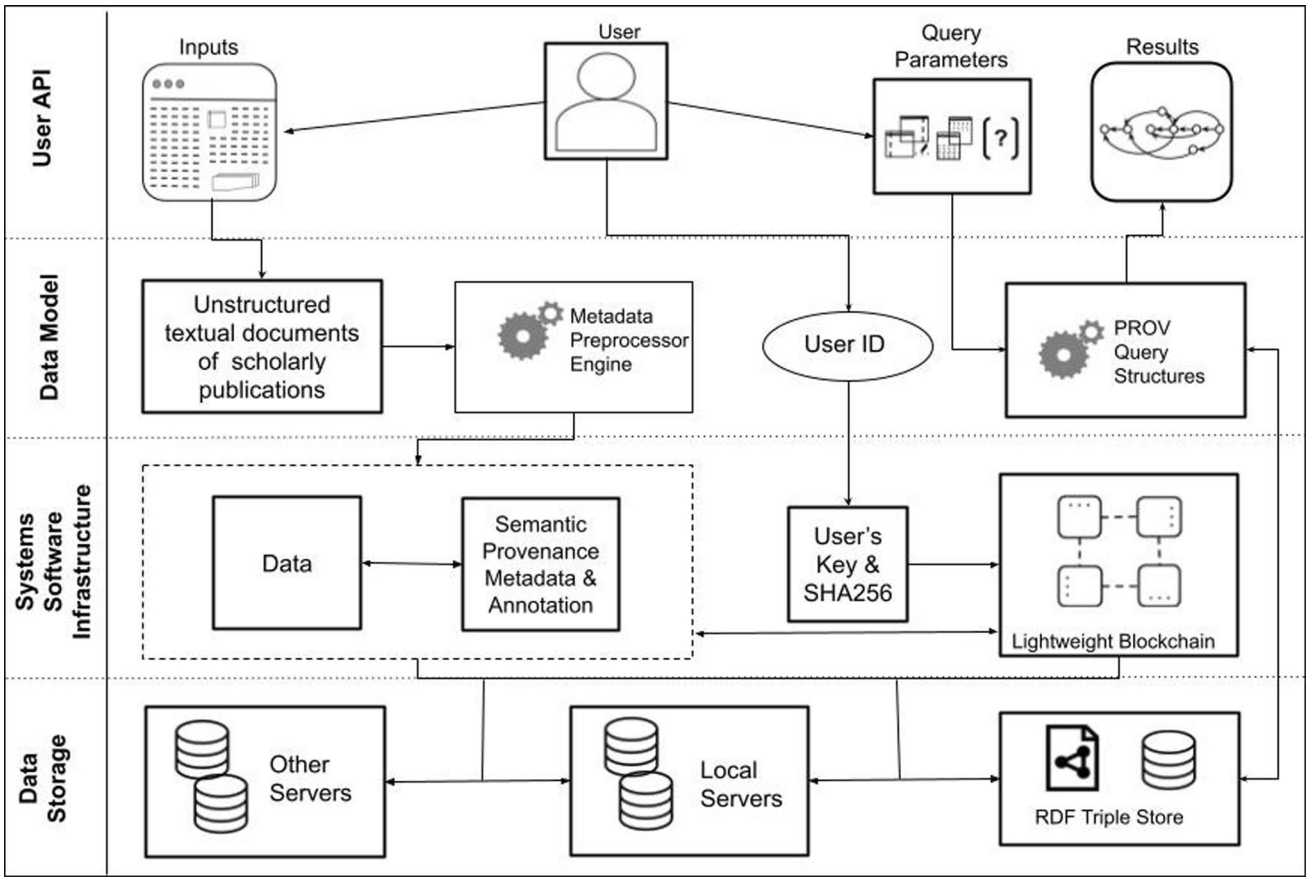

Fig. 2 *MetaScribe* Architecture Overview

6.2 Application of *MetaScribe*

MetaScribe is an example of a holistic system-wide solution, offering the ability to strengthen the underlying storage infrastructure while serving as a more reliable home for data that needs to be more carefully scrutinized. This is a growingly general need, thanks to AI-based autonomous systems becoming more capable at generating more, and more convincing, content for human consumption. This means that we have to deal with the fact that authors (of original data, or commenters and manipulators thereof) can be hyper-productive when they are no longer solely human (as they are augmented or replaced by AI tools). This requires that the mechanisms, which track how they are contributing to each other, and citing each other, have to become more robust and machine-compatible than existing infrastructure. Otherwise we would have to worry about not only deepfake videos and images, but increasingly about fake reports, correspondence, reviews, and journal papers (at least more so than we already do). In this context, *MetaScribe* could be beneficial by enabling provenance-tracking and authentication in cases, where the rate of output of passable content, being submitted to human review and consumption, is dramatically accelerated through the use of software tools (such as OpenAI's GPT-3 and Deep Learning Networks).

It is not unimaginable to see similar tools becoming increasingly significant contributors to the body of human scholarship, especially when we already see the impact of automatically generated mathematical proofs, hypotheses, texts, summaries, and evaluations of large textual corpora. In the face of such a potential deluge, and thanks to its increasingly ambiguous nature of origin, a system that improves our ability to describe provenance relationships, and automates the tracking and evaluation of authorship and attribution, is arguably needed now more than ever. However, it is also our contention that this direction of attacking the problem of auto-generated content, is easily overlooked when research is focused exclusively on mitigation techniques that look more directly at detecting the involvement of automated agents, to the exclusion of the infrastructure that might be adapted to better embrace them. To offer a bolder statement in support of such an approach, we contend that robust and flexible provenance tracking, and the means to trust its secure storage and evaluation, is a potential boon to examining the epistemological relevance of any and all electronically published media. The advantage of *MetaScribe* is that it gets

around these questions, by attributing the text to an author when the text has a human author as its origin and, if it's a machine-assisted human agency, the *MetaScribe* system does not hide that fact.

Therefore, while existing tools may aim to help detect whether a given block of text, or other media, is AI-generated after it is created, they lack the capacity to reliably trace the origins of AI-generated news stories, and other potentially AI-generated creative writings. In attempting to detect "fake" content, they neglect to attempt to address the more fundamental problem of reliable attribution at the time of creation of data items. Our proposed *MetaScribe* architecture supports a reliable mechanism to record, declare, track, and interrogate the provenance of media at the point of its creation, and through its transfers and transformations. Establishing and maintaining such provenance records allows a reliable mechanism to determine the source of original authorship, and to defer the questions of copyright and agency regarding the human and algorithmic authors of the work thus preserved.

6.3 Ethical implications

Our approach to the problem of automatically generated content, and our focus on the data storage, provenance, and presentation infrastructure as a starting point, is borne from the sound engineering perspective that it is wrong to focus solely on the individual technical problem without looking at the broader system within which it resides, i.e., the broader technical context of the problem. Neglecting such a consideration can be construed as a moral failing on our part, just as it would be a negligent oversight by a good civil engineer who fails to evaluate the ground upon which they plan to erect a building. It is really a question of what lens through which we view the system, and what philosophical assumptions and questions we follow concerning the role and nature of digital scholarship (and more generally data generation and preservation). Therefore, instead of solely debating the ethics of using, or not using, software agents (AI), we see the need to ask how existing human activities can be made more resilient to the advent of such AI-enabled automation. As the current ethical discussions tend to be limited to the technology, and its immediate impact as applied to existing systems, we, therefore, feel that it is critical to broaden our outlook on how we might be able to change existing systems to better respond to the disruptive nature of automation technologies.

Therefore, the ethical discussion must include a consideration, not just of the morality of proposed solutions within existing system infrastructures, but must also consider the potential (re)engineering of the underlying infrastructures used to generate, represent, and disseminate digital content. As we mentioned above, we feel this is easily related to the basic epistemological value of published media, and that the abilities and trustworthiness of underlying data storage and presentation infrastructure (e.g., in the publishing domain) will impact that value. In other words, the general question of attribution and identity [50], is impacted by changes in the technical abilities of underlying storage systems. If we care about truth in media, be it the media of scholarship, or the broader media of mass consumption, then we have to develop and embrace more robust solutions to the epistemological problems of provenance and attribution [44] whose current best solutions are too exclusively human-based. Human-based authentication systems, and human-verified provenance mechanisms, will all invariably start increasingly failing as it becomes easier and easier to produce more and more content that convincingly seems human-generated [this is especially obvious with deepfakes [31, 43], in general, but is even a problem with fake journal papers [26, 37].

Another ethical question that we would like to raise is this: is limiting the use of such algorithmic automation an unfair act of limiting those who otherwise cannot write well without the aid of the machine? When we focus on underlying infrastructure using a model, such as *MetaScribe,* then the answer is an affirmative 'yes'. Limiting access to automated content generation because of ethical concerns, which could be largely addressed if a means of identifying and tracking the use of automation tools (a secure provenance framework) was available, would constitute unnecessarily hindering those people who could not communicate their views (as well, or at all) without the use of automation tools. For example, Stephen Hawking, the famous scientist, who used an Augmentative and Alternative Communication (AAC) device to generate text for his speech, would be unnecessarily hindered and silenced if denied the use of such a device. However, automated generation of media that could convincingly pass as human-authored only seems more threatening that such an assistive technology as long as there is a concern that the origins of its authorship could be deceptive. Novel systems, such as *MetaScribe,* are, therefore, needed if humans are no longer the prime generators of content, and in the presence of such systems the limiting of access to content generating technologies raises an ethical concern regarding the impact of such limitation on those whose lives and abilities are most enhanced by them.

Therefore, we have to look at not just the content, but the broader ecosystem in which the content exists. If we do not do this, we may not just lose truth in digital content, but fail to serve truth as fully as we could. We might not just leave ourselves unnecessarily vulnerable to an ever-growing flood of ever-more convincing, and potentially malicious, automatically generated content, but we might also fail to support the positive benefits of such technologies (thereby cruelly denying such benefits to those who'd most benefit from them).

7 Conclusions

As major organizations already use automated journalism, it arouses enormous interest and promises great potential benefits to more organizations, especially considering the mounting financial pressure on media outlets and their continuous quest for more rapid content generation with lower marginal costs. However, this practice has moral implications beyond journalism and media, and raises fundamental questions on algorithmic authorship and attribution. Several prevention and detection tools are being developed and deployed to fight against maliciously manipulated, or deceptively AI-generated media, but aside from the fact that most of them still need human assistance and intelligence to be effective in determining attribution and authenticity, such detection algorithms represent an incomplete technical approach to the problem.

As discussed, algorithmic authorship is a complex issue, involving crucial philosophical, ethical, and theoretical concerns, but the technical approaches to addressing these questions are broader than mere detection of automated content. Though many issues related to machine morality have no common established opinions, approaches, or answers, there are efforts to model moral decision-making logic and learning into algorithms. Meanwhile, the shift in the machine moral debate should be expanded from solely focusing on which moral philosophies one should use in constructing artificial moral agents to the question of to what extent machine moral agents could be efficiently utilized. These questions can be better addressed if we broaden our technical solutions to consider not just detection, but also the broader infrastructure within which data is stored and disseminated. With infrastructure that includes secure provenance tracking, a model like our proposed *MetaScribe*, can address the machine moral dilemma of attribution and authorship of auto-generated content, by allowing transparency and confidence in the provenance of the presented media (e.g., clearly ascribing the text to an author when it originates from a human author and if it's a machine). In future work, we plan to extend the discussion on moral epistemology and the requirements of full moral agency for content generation algorithms (especially within a more robust data storage and sharing infrastructure than is currently available with existing systems).

Author contributions All authors contributed to the study conception and design. The first draft of the manuscript was written by Maria Joseph Israel and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding The authors did not receive support from any organization for the submitted work.

Data availability The manuscript does not contain any associated data.

Code availability Not applicable.

Material availability Not applicable.

Declarations

Conflict of interest All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Ethics approval Not applicable.

Consent Not applicable.

References

1. Allen, C., Varner, G., Zinser, J.: Prolegomena to any future artificial moral agent. J. Exp. Theor. Artif. Intell. **12**(3), 251–261 (2000)
2. Allen, C., Wallach, W.: Moral machines: contradiction in terms, or abdication of human responsibility? In: Lin, P., Abney, K., Bekey, G. (eds.) Robot ethics: the ethical and social implications of robotics, pp. 55–68. MIT Press, Cambridge (2011)
3. Altintas, I., Barney, O., & Jaeger-Frank, E.: Provenance collection support in the Kepler scientific workflow system. In: International Provenance and Annotation Workshop (IPAW), 118–132 (2006)
4. Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., et al.: The moral machine experiment. Nature **563**(7729), 59–64 (2018)
5. Babar, Z., van Laarhoven, T., Zanzotto, F.M., Marchiori, E.: Evaluating diagnostic content of AI-generated radiology reports of chest X-rays. Artif. Intell. Med. **116**, 102075 (2021)
6. Bainbridge, D. I.: Software copyright law. Fin. Times Manag. (1992)
7. Baluja, T.: Robot Reporters: The New Frontier in Journalism? The Canadian Journalism Project (2013). Retrieved on May 5, 2021. http://j-source.ca/article/robot-reportersnew-frontier-journalism
8. Bavoil, L., Callahan, S.P., Crossno, P.J., Freire, J., Scheidegger, C.E., Silva, C.T., & Vo. H.T.: Vistrails: enabling interactive multiple-view visualizations. In: IEEE Visualization (VIS), 135–142 (2005)

9. Berglez, P.: Inside, outside, and beyond media logic: Journalistic creativity in climate reporting. Media Cult. Soc. **33**(3), 449–465 (2011)
10. Boden, M.A.: Computer models of creativity. AI Mag. **30**(3), 23–23 (2009)
11. Bodenhausen, G.H.: United States copyright protection and the berne convention. Bull. Copyright Soc. USA **13**, 215 (1965)
12. Bridy, A. Coding Creativity: Copyright and the Artificially Intelligent Author. Stanford Technology Law Review, 2012, 5–28 (2012)
13. Bridy, A.: The Evolution of authorship: work made by code. Colum. JL Arts **39**, 395 (2015)
14. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al.: Language models are few-shot learners. arXiv preprint (2020). arXiv:2005.14165
15. Bryson, J.: Robots should be slaves. In: Wilks, Y. (ed.) Close engagements with artificial companions: key social, psychological, ethical and design issue, John Benjamins Publishing, Amsterdam, 63–74 (2008)
16. Bunz, M.: The silent revolution: how digitalization transforms knowledge, work, journalism and politics without making too much noise. Basingstoke: Palgrave Pivot (2014)
17. Carlson, M.: The robotic reporter: automated journalism and the redefinition of labor, compositional forms, and journalistic authority. Digit. Journal. **3**(3), 416–431 (2015)
18. Cetinic, E., She, J.: Understanding and creating art with AI: review and outlook. ACM Trans. Multimed. Comput. Commun. Appl. (TOMM) **18**(2), 1–22 (2022)
19. Chowdhury, M. J. M., Colman, A., Kabir, M. A., Han, J., & Sarda, P.: Blockchain as a notarization service for data sharing with personal data store. In 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE). 1330–1335 (2018)
20. Clerwall, C.: Enter the robot journalist: users' perceptions of automated content. Journal. Pract. **8**(5), 519–531 (2014)
21. Clifford, R.D.: Intellectual property in the era of the creative computer program: Will the true creator please stand up. Tul. L. Rev. **71**, 1675–1703 (1996)
22. Crane, L.: Creative differences. New Sci. **241**(3211), 18–19 (2019)
23. Danzon-Chambaud, S.: A systematic review of automated journalism scholarship: guidelines and suggestions for future research. Open Res. Eur. **1**(4), 4 (2021)
24. Davidson, S.B. & Freire, J.: Provenance and scientific workflows: challenges and opportunities. In: ACM Conference on the Management of Data (SIGMOD), 1345–1350 (2008)
25. Davis, R.: Intellectual property and software: the assumptions are broken. In Proceedings of the WIPO Worldwide Symposium on the Intellectual Property Aspects of Artificial Intelligence, Stanford, March 25–27 (1991)
26. De Vries, K.: You never fake alone. Creative AI in action. Inf. Commun. Soc. 1–18 (2020)
27. Diakopoulos, N.: Algorithmic accountability reporting: on the investigation of black boxes. Digit. J. **3**(3), 398–415 (2014)
28. Dolhansky, B., Bitton, J., Pflaum, B., Lu, J., Howes, R., Wang, M., & Canton Ferrer, C.: The deepfake detection challenge dataset. arXiv e-prints, arXiv-2006 (2020). https://arxiv.org/abs/2006.07397
29. Dufour, N. & Gully, A.: Contributing Data to Deepfake Detection Research (2019). Retrieved May 10, 2021, from http://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html
30. Fei, J., Xia, Z., Yu, P., Xiao, F.: Exposing AI-generated videos with motion magnification. Multimed. Tools Appl. **80**(20), 30789–30802 (2021)
31. Floridi, L.: Artificial intelligence, deepfakes and a future of ectypes. Philos. Technol. **31**(3), 317–321 (2018)
32. Formosa, P., Ryan, M.: Making moral machines: why we need artificial moral agents. AI Soc. **36**(3), 839–851 (2021)
33. Freire, J., Koop, D., Santos, E., Silva, C.T.: Provenance for computational tasks: a survey. Comput. Sci. Eng. **10**(3), 11–21 (2008)
34. Ghuman, R., Ripmi, K.: Narrative science: a review. Int. J. Sci. Res. (IJSR) **2**(9), 205–207 (2013)
35. Ginsburg, J.C.: The concept of authorship in comparative copyright law. DePaul L. Rev. **52**, 1063–1092 (2002)
36. Gragnaniello, D., Marra, F., & Verdoliva, L.: Detection of AI-Generated Synthetic Faces. In Handbook of Digital Face Manipulation and Detection. Springer, Cham, pp. 191–212 (2022)
37. Habal, M.B.: Artificial intelligence and machine learning in the identification of authentic and fake data presentation. J. Craniofacial Surg. **30**(6), 1617–1618 (2019)
38. Haenlein, M., Kaplan, A.: Siri Siri in my hand: Who's the fairest in the land? On the interpretations, illustrations and implications of artificial intelligence. Bus. Horiz. **62**(1), 15–25 (2019)
39. Johnson, D.G.: Computer systems: moral entities but not moral agents. Ethics Inf. Technol. **8**(4), 195–204 (2006)
40. Kaiser, B.: I blew the whistle on Cambridge Analytica—four years later Facebook still hasn't learnt its lesson. (2020). Retrieved May 5, 2021, from https://www.independent.co.uk
41. Karjala, D.S.: United States adherence to the berne convention and copyright protection of information-based technologies. Jurimetrics **28**(2), 147–152 (1988)
42. Kochupillai, M., Lütge, C., Poszler, F.: Programming away human rights and responsibilities? "The Moral Machine Experiment" and the need for a more "humane" AV future. NanoEthics **14**(3), 285–299 (2020)
43. Korshunov, P. & Marcel, S.: Vulnerability assessment and detection of Deepfake videos, 2019 International Conference on Biometrics (ICB), Crete, Greece, 1–6 (2019). https://doi.org/10.1109/ICB45273.2019.8987375
44. Kruglanski, A.W.: Causal explanation, teleological explanation: on radical particularism in attribution theory. J. Pers. Soc. Psychol. **37**(9), 1447–1457 (1979). https://doi.org/10.1037/0022-3514.37.9.1447
45. Lahman, M.K., Rodriguez, K.L., Moses, L., Griffin, K.M., Mendoza, B.M., Yacoub, W.: A rose by any other name is still a rose? Problematizing pseudonyms in research. Qual. Inquiry **21**(5), 445–453 (2015)
46. Latar, N.L., Nordfors, D.: Digital identities and journalism content-how artificial intelligence and journalism may co-develop and why society should care. Innov. Journal. **6**(7), 3–47 (2009)
47. Liu, Y., & Wu, Y. F. B.: Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In Thirty-Second AAAI Conference on Artificial Intelligence (2018)
48. Locascio, J.J.: Results blind science publishing. Basic Appl. Soc. Psychol. **39**(5), 239–246 (2017)
49. Louie, R., Coenen, A., Huang, C. Z., Terry, M., & Cai, C. J.: Novice-AI music co-creation via AI-steering tools for deep generative models. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–13 (2020)
50. Love, H.: Attributing Authorship: An Introduction. Cambridge University Press, Cambridge (2002)
51. Markham, T.: The politics of journalistic creativity: expressiveness, authenticity and de-authorization. Journal. Pract. **6**(2), 187–200 (2012)
52. McCutcheon, J.: The vanishing author in computer-generated works: a critical analysis of recent Australian Case Law. Melbourne Univ. Law Rev. **36**(3), 915–969 (2013)

53. McCutcheon, J.: Curing the authorless void: protecting computer-generated works following IceTV and Phone Directories. Melb. UL Rev. **37**, 46 (2013)
54. McGuffie, K., & Newhouse, A.: The Radicalization Risks of GPT-3 and Advanced Neural Language Models. arXiv preprint (2020). arXiv:2009.06807
55. Meehan, J.R.: TALE-SPIN, an interactive program that writes stories. IJCAI **77**, 91–98 (1977)
56. Metz, C., Collins, K.: How an A.I. "Cat-and-Mouse Game" Generates Believable Fake Photos. The New York Times, New York (2018)
57. Miller, A.R.: Copyright protection for computer programs, databases, and computer-generated works: Is anything new since CONTU? Harvard Law Rev. **106**(5), 977–1073 (1993)
58. Montal, T., Reich, Z.: I, robot. You, journalist. Who is the author? Authorship, bylines and full disclosure in automated journalism. Digit. Journal. **5**(7), 829–849 (2017)
59. Napoli, P.M.: Automated media: an institutional theory perspective on algorithmic media production and consumption. Commun. Theory **24**(3), 340–360 (2014)
60. Partadiredja, R. A., Serrano, C. E., & Ljubenkov, D.: AI or human: the socio-ethical implications of AI-generated media content. In: 2020 13th CMI Conference on Cybersecurity and Privacy (CMI) - Digital Transformation - Potentials and Challenges, 1–6 (2020). https://doi.org/10.1109/CMI51275.2020.9322673
61. Paul, E.S., Kaufman, S.B. (eds.): The Philosophy of Creativity: New Essays. Oxford University Press, Oxford (2014)
62. Powers, T.M.: On the moral agency of computers. Topoi **32**, 227–236 (2013). https://doi.org/10.1007/s11245-012-9149-4
63. Pressman, L.: The automated journalism. Automated Insights Blog (2017). Retrieved April 5, 2021. https://automatedinsights.com/blog/the-automated-future-of-journalism/
64. Pritchard, S.M.: Double-blind review: a commitment to fair editorial practices. Portal: Libraries Acad. **12**(2), 117–119 (2012)
65. Radford, A., Wu, J., Amodei, D., Amodei, D., Clark, J., Brundage, M., & Sutskever, I.: Better Language Models and Their Implications (2019). Retrieved May 10, 2021. https://openai.com/blog/hetter-language-models/
66. Rajan, M.T.S.: Moral rights: principles, practice and new technology. Oxford University Press (2011)
67. Reddy, M. D. M., Basha, M. S. M., Hari, M. M. C., & Penchalaiah, M. N.: Dall-e: Creating images from text. *UGC Care Group I Journal*, **8** (14), 71–75 (2021)
68. Reich, Z.: Constrained authors: bylines and authorship in news reporting. Journalism **11**(6), 707–725 (2010)
69. Reich, Z., Boudana, S.: The fickle forerunner: the rise of bylines and authorship in the French press. Journalism **15**(4), 407–426 (2014)
70. Rosenbaum, M.: Pseudonyms to protect authors of controversial articles, BBC News (2018). Retrieved May 5, 2021, from https://www.bbc.com/news/education-46146766
71. Ross-Hellauer, T.: What is open peer review? A systematic review. F1000Research (2017). https://doi.org/10.12688/f1000research.11369.2
72. Samuelson, P.: Allocating ownership rights in computer-generated works. U. pitt. L. rev. **47**, 1185–1228 (1985)
73. Scheutz, M.: The need for moral competency in autonomous agent architectures. In V. C. Müller (Ed.) Springer International Publishing, pp. 515–525 (2016)
74. Schober, R.: Passing the turing test? AI generated poetry and posthuman creativity. Artif. Intell. Hum. Enhanc. **21**, 151 (2022)
75. Searle, J.: Minds, brains and programs. Behav. Brain Sci. **3**, 417–457 (1980)
76. Sharkey, A.: Can robots be responsible moral agents? And why should we care? Connect. Sci. **29**(3), 210–216 (2017). https://doi.org/10.1080/09540091.2017.1313815
77. Singer, P.: Setting the record straight on the Journal of Controversial Ideas. The Guardian (2018). Retrieved May 5, 2021, from https://www.theguardian.com/world/2018/nov/18/setting-the-record-straight-on-the-journal-of-controversial-ideas
78. Stokes, S.: Digital copyright: law and practice. Bloomsbury Publishing (2019)
79. Tang, Y.: A robot wrote this?: An empirical study of AI's applications in writing practices. In: The 39th ACM International Conference on Design of Communication, 380–381 (2021)
80. Tonkens, R.: A challenge for machine ethics. Mind. Mach. **19**(3), 421–438 (2009). https://doi.org/10.1007/s11023-009-9159-1
81. Turing, A.: Computing machinery and intelligence. Mind **59**(236), 433–460 (1950)
82. Uricchio, W.: The algorithmic turn: photosynth, augmented reality and the changing implications of the image. Vis. Stud. **26**(1), 25–35 (2011)
83. Van Dalen, A.: The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists. Journal. Pract. **6**(5–6), 648–658 (2012)
84. Van der Kaa, H., Krahmer, E.: Journalist versus news consumer: the perceived credibility of machine written news. In: Proceedings of the Computation and Journalism Conference, vol. 24, pp. 25–29. Columbia University, New York (2014)
85. Van Houweling, M.S.: Author autonomy and atomism in copyright law. Va. L. Rev. **96**, 549 (2010)
86. Van Wynsberghe, A., Robbins, S.: Critiquing the reasons for making artificial moral agents. Sci. Eng. Ethics **25**(3), 719–735 (2019)
87. Waddell, T.F.: A robot wrote this? How perceived machine authorship affects news credibility. Digit. Journal. **6**(2), 236–255 (2018)
88. Wallach, W., Allen, C.: Moral Machines: Teaching Robots Right from Wrong. Oxford University Press, Oxford (2008)
89. Weiner, A.: Fantasy Football and the Cold Future of Robot Journalism. The Daily Dot (2014). Retrieved May 5, 2021. http://kernelmag.dailydot.com/issue-sections/features-issue-sections/10097/fantasy-football-and-the-cold-future-of-robot-journalism/
90. Wu, Y., Mou, Y., Li, Z., Xu, K.: Investigating American and Chinese Subjects' explicit and implicit perceptions of AI-Generated artistic work. Comput. Hum. Behav. **104**, 106186, 1-11 (2020)
91. Young, M.L., Hermida, A.: From Mr. and Mrs. outlier to central tendencies: computational journalism and crime reporting at the Los Angeles Times. Digit. Journal. **3**(3), 381–397 (2015)
92. Yse, D. L.: Your Guide to Natural Language Processing (NLP) (2019). Retrieved May 5, 2021 from https://towardsdatascience.com/your-guide-to-natural-language-processing-nlp-48ea2511f6e1
93. Zhou, X., Zafarani, R., Shu, K., & Liu, H.: Fake news: fundamental theories, detection strategies and challenges. In Proceedings of the twelfth ACM international conference on web search and data mining, 836–837 (2019)