

Lala Hajibayova, PhD
School of Information
Kent State University

Seungmin Lee, PhD
Chung-Ang University

Pre-Print
Journal of Documentation, 2025

Authorship in the Age of Artificial Intelligence

Abstract

Purpose

In this paper the evolving boundaries between human and machine-generated content are examined the lens of Foucault's (2002) conceptualizations of authorship and associated ethical considerations in the interest of ensuring quality representation and organization of knowledge.

Findings

The ethical norms and standards of the biosphere are well-aligned with Foucault's conceptualization of work and authorship and promotes a deep understanding of how content can be evaluated.

Originality

This paper argues that implementing the ethical norms and standards of the biosphere within the information environment and incorporating established principles of accountability, fairness, equity, and transparency into knowledge representation, organization, and discovery systems are crucial steps toward fostering ethical and democratic systems of knowledge representation, organization, and discovery.

Keywords

Authorship, Artificial Intelligence, Ethics of knowledge representation and organization.

Introduction

The boom of Artificial Intelligence (AI) technologies and its capabilities to mimic human behaviors and interactions have led to the widespread production of AI-generated content, which extends to widely used bookselling and review platforms on which distinguishing between human and AI-authored works demand an increasingly discerning eye. For example, within days of its release, Maria Arana's well-researched book, *LatinoLand: A portrait of America's largest and least understood minority*, was associated with a summary by Clara Bailey on Amazon of a book entitled *America's largest and least understood minority: a summary of Latinoland*, one of many such summaries that were later removed from Amazon after being flagged as AI-generated content, although it is still available on Goodreads, the Amazon-owned book-review platform (Limpong, 2024). Similarly, to fulfill his long-held dream of becoming a published author, Brett Schickler used ChatGPT to create a 30-page illustrated children's book titled *The Wise Little Squirrel: A Tale of Saving and Investing* in just a few hours, which is now available through Amazon's self-publishing unit (Bensinger, 2023).

It comes as no surprise that one of the main concerns with AI-driven output is a real danger of unscrupulous use, such as the generation of massive amounts of misleading content or misinformation that perpetuates the problem of large-scale persuasion based on groundless claims. In this regard, in their analysis of how AI models mimic human falsehoods, Lin and colleagues (2022) found that the best model was truthful in responding to only 58% of 817 questions across 38 categories (such as health, law, finance and politics), while human performance was 94%. As we increasingly use this data to make decisions on larger and more significant scales, we must remain aware of the underlying structures and our own role in shaping them as interventions in the world (Pine and Liboiron, 2015).

Cultural heritage institutions, such as museums, libraries, and archives, historically serve as guardians of a society's collective memory, which comprises of works that are perceived as accurate and authoritative representations of human history, culture, and knowledge. Thus, maintaining rigorous standards for authenticity and accuracy of the collected work is vital for protecting the integrity of the preserved heritage. Moreover, users of currently publicly available AI tools often do not have enough relevant experience with or knowledge about the ethical implications of using these systems, which is necessary to ensure that these tools are not causing harm. Along with identifying and distinguishing AI-created content from human-authored work, knowledge organization systems need to adopt an ethical and human-centered approach to ensure the quality and accurate representation and organization of knowledge from all sources (Zhitomirsky-Geffet and Hajibayova, 2020).

Building on previous research on ethical principles for knowledge representation and organization (e.g., Mai, 2014), the aim of this paper is to investigate the application of the ethical principles of knowledge representation, organization and discovery systems to the context of AI-generated content. Foucault's (2002) perspective on authorship and discourse is applied to an examination of the evolving boundaries between human and machine-generated content, with emphasis on the associated ethical responsibility to ensure quality representation and organization of knowledge. The ensuing discussion concerns the application of the ethical norms and standards of the biosphere (Floridi and Cowls, 2019) to the information environment and utilization of established principles of beneficence, nonmaleficence, autonomy, justice, and explicability to knowledge representation, organization and discovery systems.

Literature Review

In their classical definition of artificial intelligence, John McCarthy and associates describe AI as "... making a machine behave in ways that would be called intelligent if a human were so behaving" (1955, p. 11). However, the definitions of intelligence and understanding of AI vary among researchers. Schank (1987) argues that the goal of AI research is to build intelligent machines and understand the nature of intelligence. Schank further suggests that intelligence is built upon five key pillars: communication, world knowledge, internal knowledge, intentionality, and creativity, and that the ability to learn is the most essential component of intelligence. Definitions of AI tend to be abstract and open-ended rather than specific to specify a particular technology (Cox and Mazumdar, 2024). This generally reinforces the understanding of AI as an evolving concept, suffused with cultural meaning and significance that as stated by Cox and Mazumdar (2024) "even in its most professional applications, cannot be ignored" (p. 332). While

AI is conceptualized from a range of perspectives, the capacity for machine learning is considered one of the most components of its meaning (Hu *et al.*, 2019). Machine learning is based on statistical techniques that generate models from data without the need to manually program the models' parameters. A basic machine learning approach involves data collection and preparation, selection of relevant features such as which variables to include, choice of machine learning algorithms, selection of models and parameters, training and performance evaluation to enable the machine to discover patterns, set rules and parameters of models by studying the data (Cox and Mazumdar, 2024). Critical to machine learning is training the computer program in a way that allows it to discover aspects of the data that can be developed into a model. In this training process, the machine uses historical data as inputs and learns patterns from them to fine-tune model parameters in an iterative manner to ensure a pre-defined level of accuracy considered reasonable for estimating an output for an unknown input. (Cox and Mazumdar, 2024) For example, a machine learning model might use an author's past publication history, themes, and audience engagement data as input dataset in order to predict the likelihood of the author producing a highly cited paper within the next two years.

Russell and Norvig (2021) examined the conceptualization of intelligence in terms of fidelity to human performance and as rationality, i.e., relying on the application of logic to do the "right things." Human-like intelligence is conceptualized as both acting humanly and thinking in ways similar to human thinking. Acting humanly involves AI capabilities such as successfully processing a natural human language to communicate, knowledge representation for storing and recalling, automated reasoning that plausibly answers questions and draws conclusions, and machine learning to adapt to new situations and detect patterns. Thinking humanly involves understanding how humans think, including introspection (ability to reflect on one's own thoughts), and psychological interpretation (ability to infer human from behavior in action), which may be associated with brain imaging (ability to observe the brain in action). In sum, Russell and Norvig's view of AI intelligence focuses on thinking and acting rationally.

In library and information science, AI has long been of interest in areas such as information seeking and retrieval (e.g., Smith, 1976; Schultz, Koch and Olbrich, 2024) and knowledge representation and organization (e.g., Golubka, 2021). Recently, information science researchers have also investigated the impact of AI technologies on library operations (e.g., Cox *et al.*, 2019; Cox 2023), everyday information practices (e.g., Brewer, 2023), scholarly communication (e.g., Hosseini *et al.*, 2023) and associated copyright and ethical implications (e.g., Lee, 2023). For instance, Cox and Mazumdar (2022) have outlined the potential of AI to improve the efficiency of library operations and user services through enhancements such as expedited knowledge discovery, automated interactions (e.g., AI chatbots), and better understanding of users' behaviors.

Hirvonen and colleagues' (2023) analysis of the use of AI technologies in everyday information seeking practices reveals that AI systems integrated into search engines, social media platforms, streaming services, and media generations can demand, request, allow, encourage, discourage, or refuse users' actions, shaping their interactions and resulting in both increases and decreases in the diversity of and access to information. As a result, because AI-based systems contribute to the creation of a milieu of affordances for everyday information practices, they influence individuals, communities, societies, and the environment in ways that often go unrecognized

(Hirvonen *et al.*, 2023). To scrutinize the sustainability of these interactions, investigation of how human information practices are mediated in relation to multiple system affordances and each other is needed (Kitzie, 2019; Nathan, 2012). This research is essential to ensure development of affordances that are open and inclusive and that various stakeholders, including diverse combinations of creators, researchers, and designers are actively involved in the process (Hirvonen *et al.*, 2023).

Research has also highlighted the advantages of AI's contributions to the process of creative cultural production. For example, AI can significantly enhance creativity by analyzing existing bodies of work and generating novel ideas that can be tailored to an individual's particular creative predilections, as well as to specific areas of content production such as curation and publishing (Hu, 2023). However, issues surrounding AI-generated content, specifically authorship and associated copyright concerns, are complicated because the line between human and AI-generated content cannot be drawn intuitively, so traditional norms and standards of determining authorship and originality are challenged (Kariyawasam, 2020).

Thus, Lund and colleagues' analysis of application of ChatGPT in research raises ethical concerns related to ownership of AI-generated content and compliance with copyright laws. Along these lines, Wang and associates (2022) suggest that use of AI-generated content in any situation may have legal and ethical implications. The ambiguity in the attribution of AI-generated content will lead to more debates about ownership of cultural production, especially regarding the distinction between original and derivative works, and the role of AI participation as a quantitatively substantive dynamic in the formulation of policies and regulations (Hu, 2023).

Addressing copyright concerns associated with the AI creations, Lu (2021) asserts that AI-generated works could and should be protected by copyright law. Based on the theory of the control of the creative process, Lu argues even though AI implementation may not directly contribute to the creation of a work, they still represent the human elements behind the creation. Therefore, for AI-created works, authorship is "better transferred to a person behind the AI who had control over the creative process in order to safeguard the current copyright system and its founding principles" (p. 2). Lu (2021) further argues that human authorship requirement is in compliance with the current legal system, because only human beings can hold legal rights to copyrights and assume legal obligations, while non-humans are excluded.

Lund and Naheem's (2024) analysis of the AI authorship policies of 300 top academic journals, revealed that the majority (176 journals, 58.7%) have AI authorship policies and guidelines for acknowledging that AI was used in manuscript preparation. Moreover, 170 of these journals (96.6%) allowed for using ChatGPT to improve the quality of manuscripts, while six policies (3.4%) prohibited any use of these models. In terms of inclusion of AI tools like ChatGPT as an author, almost all the reviewed policies (98.9%) specified that no AI tools can be listed as authors. Nonetheless, Nazarovets and da Silva (2024) identified 14 publications that credited ChatGPT as author, co-author or group author in Web of Science and Scopus citation databases, with several papers accruing hundreds of such citations. The inclusion of AI-authorship of publications in mainstream indexing databases raises critical questions about the ethics of scholarly practices, the standards of accountability for both produced and indexed content, and most importantly, the integrity of scholarly output. These concerns prompt deeper

discussions on ethical norms and standards of authorship. Moreover, considering the lack of evaluation standards for language models used in automated content creation and editing, it is essential to critically examine AI use and its impact on scientific and cultural production.

AI technologies have long been an integral part of knowledge representation, organization, discovery, as well as information services, and they have been a sustained area of inquiry within information science research (e.g., Smith, 1980). Mannheimer and colleagues' (2024) have identified three main applications of AI technologies: metadata extraction, recommender systems and reference services, and assessment of digital resources. The application of AI tools to extract and enhance accurate metadata for digital collections, involves techniques such as natural language processing, classification, image recognition, optical character recognition (OCR), and automated object detection. For example, ExLibris Knowledge Center's AI tool, the Alma Metadata Assistant, uses a large language model generative AI to analyze a textual or visual object and generate a metadata that could be used either as a new bibliographic record or to enrich the existing record (ExLibris, n.d.). Similarly, to demonstrate the evolution of gender portrayal in popular media, Schuerkamp and colleagues (2023) applied object detection models to analyze collection of over 30,000 digitized historical postcards to detect depictions of people and classify them by gender. Recommender systems and reference services represent the second most widely implemented AI-driven solutions that aim to facilitate information discovery and respond to the increasing demand for virtual references by employing predictive modeling to evaluate the complexity of users' queries and directing them to the appropriate library personnel. For instance, Library of Congress utilized the Newspaper Navigator application, an experimental AI tool, that used machine learning to enable visual similarity search and browsing of over 1.56 million images extracted from the database of digitized historic newspapers (Library of Congress Labs, n.d.). Such assessment involves utilization of AI techniques to identify latent patterns across vast datasets. For example, Short (2019) describes application of text mining techniques to conduct a subject analysis of a collection of dime novels, including keyword extraction, named-entity recognition, clustering, topic modeling and classification to enhance discoverability of novels and increase the efficiency of digitization of workflow.

Yet, the power of AI to facilitate services does not alleviate the concerns about the use of AI-generated content in intellectual and creative production or its impact on the quality and accuracy of representation and organization of knowledge. The blurring of the boundaries between human and AI-generated content significantly challenges the cultural heritage function of information representation, organization, and retrieval: safeguarding trustworthy knowledge and providing ethical, accurate and fair access to authentic cultural artifacts, such as musical, literary, and art works or oeuvres, the intellectual and artistic creations that have always been key entities for information representation, organization, and retrieval (Smiraglia, 2003). The perception of legitimacy and authority derived from authorship is crucial in fostering trust between content creators and consumers, cultivating dedicated audiences, and promoting patronage. This perceived authenticity not only enhances the cultural and intellectual value of the work but most importantly encourages the recognition that sustains a creative and scholarly ecosystem.

In the following discussion, the shifting boundaries between human and machine-generated content are examined through the lens of Foucault's (2002) conceptualization of authorship and

discourse, addressing questions of ethical principles for quality representation and organization of knowledge.

AI, Authorship and Foucault's Conceptualization of Authorship

In an era of proliferation of artificial writing the question of authorship not only persists, but as expressed by Bassett (2025), is “certainly reaching beyond narrow matters of legal ascription, copyright and ownership” (p. 20). Authorship of a text is a key aspect of fair and quality representation of knowledge, as it not only ensures accurate attribution, collocation, and access to the work, but, most importantly, acknowledges the creator's originality and cultivates readership. The authorship of a text is also one of the key criteria for assessing its accountability, ensuring it is not plagiarized and complies with copyright regulations. Moreover, the concept of authorship is important for readers' perceptions, as it aligns with their association of authors' works with their identity and life story (Hongisto, 2025). In knowledge repositories, works are considered to constitute the intellectual or artistic contents of documents (Smiraglia, 2003), which may be represented, organized, and retrieved. Traditionally, authorship has been conceptualized as an acknowledgement of responsibility. According to Resource Description and Access (RDA), a widely adopted international cataloging standard, authorship is defined as the contribution of individuals or entities to the creation of intellectual, creative, and artistic content. This definition characterizes authorship as the entity responsible for producing the intellectual or creative substance of a work, reflecting a stable and traditional understanding of the concept.

Even though widespread discussions of the authorship of AI-generated literature have recently gained traction withing a larger audience, concerns about technology-generated content date back to the mid-20th century. One of the earliest examples of rule-based literary text generation programs *Loveletters*, created by Christopher Strachey (1954), was based on distinct roles for both the programmer and the program. The programmer defined a simple rule-based script that included the concept (i.e., love letters), vocabulary, sentence composition, length, and order of sentence components to generate short love letters. The program selected the words and filled in the programmer's rule-based script, generating a love letter that varied every time the program was run (Strachey, 1954).

These human-machine cooperative rule-based programs were further developed into more sophisticated systems, in which various agents, such as programmers, data curators, and users, could contribute to the generated text (Hongisto, 2025). This practice of co-creation has made the process of writing and artistic expression not only more complex but, more importantly, has obscured attribution of authorship. As noted by Barad (2003), the process of co-creation, widely known in posthumanist artistic discourse as performativity, incorporates “material and discursive, social and scientific, human and nonhuman, and natural and cultural factors” (p. 808). Barad further asserts that the posthumanist perspective calls into question the givenness of the concepts of human and nonhuman, examining the practices through which the boundaries between these concepts are stabilized and destabilized and enabling “a genealogical analysis of the discursive emergence of the “human” as a “part of the world in its open-ended becoming” (p. 821).

Floridi (2025) introduces the concept of *distant writing*, wherein a human author acts as a “meta-author,” one who designs narrative constraints and sets requirements through prompts and curation of AI output. Floridi posits that distant writing (or “wrAIting”) “invites us to reconsider not just how we produce literature but what literature is and might become” (p.24). Thus, distant writing demonstrates that AI-assisted creativity can generate previously inaccessible narrative possibilities and transform literature’s modal space. Yet, the process of human and nonhuman co-creation raises the question of textual creativity as a vital aspect of authorship that weaves together writing, intentionality, and meaning. Bassett (2025) argues that “dominant versions of authorship and creativity draw on naturalized forms of anthropomorphism, and so, are troubled by LLMs’ [Large Language Models] inhuman capacity (or apparent capacity) to generate creative writing” (p. 20). In this regard, Bimo (2025) argues that the text is meaningful “by virtue of a faith in the unique and inherent capacity of the author”, wherein the validity of the text “rests upon and is imparted by the authority of the authorial subjectivity” (p. 54). Meaning then becomes a matter of tracing the intentionality of the author, revealing the communicated message, and facilitating the development of a currently appropriate legal concept of intellectual property (Bimo, 2025). On the other hand, perception of AI as a super-intelligent and transcendent author not only poses a considerable danger to the paradigm of authorship, but, most concerningly, also risks the legitimization of its such harmful outputs such as inaccuracies and fabrications or hallucinations (Bimo, 2025). Knapp and Michaels (1985) argue that the author and the work are indistinguishable, stating that “what a text means and what its author intends it to mean are identical and that their identity robs intention of any theoretical interest” (p. 731), although a work is often treated as an independent entity with its own intrinsic structure and form. However, this perception raises a legitimate question that was posed by Foucault (2002): “If an individual were not an author, could we say that what he wrote, said, left behind in his papers, or what has been collected of his remarks could be called a ‘work’?” (p. 72). Foucault’s question challenges the whole notion of authorship and definition of a work in the context of an individuals’ rights to their intellectual property and creative cultural production. Instead, Foucault argues that an author’s name serves a purpose in narrative discourse by enabling the definition and differentiation of texts and establishment of relationships among them. Foucault explains that:

Hermes Trismegistus did not exist, nor did Hippocrates – in the sense that Balzac existed – but the fact that several texts have been placed under the same name indicates that there has been established among them a relationship of homogeneity, filiation, authentication of some texts by the use of others, reciprocal explication, or concomitant utilization. The author’s name serves to characterize a certain mode of being of discourse: the fact that the discourse has an author’s name, that one can say “this was written by so-and-so” or “so-and-so is its author,” shows that this discourse is not ordinary everyday speech that merely comes and goes, not something that is immediately consumable. On the contrary, it is a speech that must be received in a certain mode and that, in a given culture, must receive a certain status. (p. 74)

Foucault (2002) identifies four characteristics of discourse containing an author function. The first characteristics is appropriation of the discourse. The historic development of discourse as a

product or an act that was “placed in the bipolar field of the sacred and the profane, the licit and illicit, the religious and the blasphemous” (p. 75) resulted in a system of ownership of text, with associated rights and relations. This process was also coupled with the “possibility of transgression,” which, attached to the act of writing, increasingly took on the form of an “imperative peculiar to literature” (p. 75). Thus, according to Foucault, while the author seemed to be deemed a property of the text, the rediscovery of the “old bipolar field of discourse” (p. 75) entailed the systematically practiced transgression of boundaries that enabled an authentic and open interaction between the text and its audience, reintroducing a sense of risk to writing previously stabilized through notions of ownership. The second characteristic is recognition that the author function does not impact all universal discourses in a consistent way. For instance, in the past the merit of literary texts “whether real or imagined, was regarded as a sufficient guarantee of their status” (p. 75). In other words, the works were acknowledged and valued without any consideration of their authors’ identity and respected despite their anonymity due to such criteria as ancientness. In contrast, contemporary literary discourse is acknowledged and legitimized only when imbued with the author function (p. 76). Foucault’s third characteristic of the author function is acknowledgement of the complex and specific series of “operations” that construct a certain rationale for one’s recognition as an author, such as establishment of the authenticity of the produced texts and their value. In this sense, assurance of being recognized as an author is not self-evident but a result of institutional, cultural and historical processes that regulate perception and acknowledgement. The fourth and final characteristic is that the author role may embrace plurality and “give rise simultaneously to several selves, to several subjects-positions that can be occupied by different classes of individuals (p. 78). This characteristic suggests that authorship is not a stable identity but a shifting function that enables different modes of speaking, writing, and being recognized within discourse, thereby complicating traditional notions of individual authorship and authority.

Foucault (2002) further clarifies that the author function extends far beyond of a given work, arguing that “in the sphere of discourse one can be the author of much more than a book – one can be the author of a theory, tradition, or discipline in which other books and authors will in their turn find a place” (p. 78). These authors are in a position which he called ‘transdiscursive.’

Foucault (2002) also argues that the author is not an “indefinite source of significations which will fill a work; the author does not precede the works; he is a certain functional principle by which in our culture, one limits, excludes and chooses” (p. 81). That is, he asserts that an author is not a behind-the-scenes provider of the contents of a work but one factor by which the audience determines its credibility or worth. He raises the question of whether it matters “who is speaking,” indifference to which is among the “fundamental ethical principles of contemporary writing.” For Foucault it is an ethical question because “indifference is not really a trait characterizing the manner in which one speaks and writes” (p. 81).

Foucault (2002) contends that, in the context of constant change and development, less emphasis is placed on the authorship of a discourse, regardless of its status, form and values, than on such questions as “What are the modes of existence of this discourse? Where has it been used, how can it circulate, and who can appropriate it for himself? What are the places in it where there is room for possible subjects?” ... And behind all these questions...What difference does it make who is speaking?” (p. 81).

Foucault's (2002) questions lead to exploration of the fundamental nature and dynamics of discourse. "What are the modes of existence of this discourse?" prompts consideration of the various forms and manifestations discourse can take. "Where has it been used?" directs attention to the contexts and environments to which the discourse has been applied or has referenced, which could range from academic settings and public forums to online platforms and informal conversations. Understanding the usage of discourse helps illuminate its impact and reach. "How can it circulate?" encourages examination of the mechanisms through which discourse is disseminated. The circulation of discourse affects its accessibility, influence, and the way it shapes public and private thought. "Who can appropriate it for himself?" raises the issue of ownership and control over discourse. It calls for decisions about who has the authority or power to adopt, adapt, and possibly manipulate discourse for self-serving purposes. This question touches on intellectual property, cultural appropriation, and the power dynamics inherent in the control of knowledge. "What are the places in it where there is room for possible subjects?" requires identification of opportunities within the discourse for individual or collective participation and interpretation. It highlights the inclusivity or exclusivity of discourse and whether it allows space for diverse voices and perspectives. Finally, the overarching question, "What difference does it make who is speaking?" calls into question the significance of the author's identity. It suggests that the impact of discourse might be independent of its author, focusing instead on the quality of the content and how it has been situated and perceived. In the context of AI production, the construct of authorship becomes a mechanism that places authority on the meaning of the text and its value. From Foucault's (1984) perspective, the author serves as a mechanism of discursive formulation, situated within specific historical and cultural contexts that bounds the meaning of the text and thereby enables particular formulations of power. In the context of AI-generated text, authorship is construed as a "discursive constructions that govern how... output is given meaning and value" (Bimo, 2025, p. 51).

Foucault's perspective invites an examination of traditional notions of authorship and discourse in the context of ethical representation and organization of knowledge, a central function of which is to ensure equitable access to resources that are authoritative, accurate, and ethically sourced. This function is not only essential for upholding the credibility and accountability in representation, organization and production of knowledge that is critical for fostering trust in information systems and services. In this regard, Fox and Reece (2012) state that "all of the collective must work to attain the end of benevolent user access, which should be the shared axiomatic starting point of any ethical model of IO [information organization]" (p. 382).

AI and Ethics of Authorship

The debate surrounding the ethics of AI is nearly as old as the establishment of artificial intelligence as a field of academic research, dating back to the 1950s (McCarthy *et al.*, 2006; Samuel, 1960). Due to the growing impact of technologies on human lives, ethical considerations serve as enduring principles that guide the development and acceptance of technology (Friedman and Kahn, 2007). Ethical considerations include application of the existing moral theories to clarify relevant issues and prescribe appropriate norms of behavior. While principles and guidelines for ethical artificial intelligence reflect a global convergence around five core ethical values—transparency, justice and fairness, non-maleficence, responsibility, and privacy—there remains significant divergence in how these principles are interpreted, why they are considered

important, which issues, domains, or actors they address, and how they should be implemented (Jobin *et al.*, 2019). Friedman and Kahn (2007), for example, argue that integration of ethics and sociotechnical analysis with value-sensitive design accounts for human values that center on human well-being, human dignity, justice, welfare, and human rights. The value-sensitive approach is principled in that it maintains that such values have moral epistemic standing independent of whether a particular person or group upholds such values. Bostrom (2014) raises an underlying question of whether AI can be developed to “understand” the idea of “moral rightness” while humans themselves have an “imperfect understanding of what is right and wrong, and perhaps an even poorer understanding of how the concept of moral rightness is to be philosophically analyzed” (p. 267). Hagendorff (2020) suggests a “situation-sensitive ethical approach based on virtues and personality dispositions, knowledge expansions, responsible autonomy and freedom of actions” to accommodate “individual situations and specific technical assemblages” (p. 114). Hagendorff (2020) further calls to “emancipate” moral actors from “potential inability to act self-responsibly on the basis of comprehensive knowledge, as well as empathy in situations where morally relevant decisions have to be made” (p. 114).

AI ethical principles are also often aligned with the four fundamental principles of bioethics, i.e., beneficence, nonmaleficence, autonomy and justice, as these principles are closely related to digital ethics in dealing with new forms, agents, clients and environments (Floridi, 2013). Floridi (2008) contended that ethical standards should be uniformly applied to both human and non-human entities, irrespective of their existence in either the cyber or biosphere, encompassing all areas of the planet that sustain life. Based on previously developed AI guidelines and principles, Floridi and Cowls (2022) have further developed these ideas and recommended a unified framework of five principles of AI in society: beneficence, nonmaleficence, autonomy, justice and explicability. In their formulation, beneficence emphasizes the importance of promoting the well-being of both people and the planet through AI; nonmaleficence (do-no-harm) stresses the need to prevent infringements on individuals’ privacy and security; autonomy involves balancing decision-making power between humans and artificial agents; justice calls for elimination of unfairness and discrimination, the promotion of diversity, and the prevention of new threats to justice; and, explicability entails incorporating both the epistemological sense of “intelligibility” (how does it work?) and the ethical sense of “accountability” (who is responsible for the way it works?).

In recent years, cultural heritage institutions have increasingly focused on the critical analysis of the implications of applying AI and machine learning to digital collections (Fleischmann *et al.*, 2017; van Strien *et al.*, 2022). This process has led to the development of various tools, such as checklists, toolkits, and impact assessments designed to evaluate various stages of machine learning projects, the stages of which encompass dataset creation, model training and auditing, and model implementation in operational settings (Lee, 2023). The checklists are composed to target a broad audience ranging from researchers and practitioners to personal users of the models (Lee, 2023). These considerations are crucial in addressing both scientific and ethical concerns: scientifically, there is the problem of generalizability of results; ethically, there is the risk of significant real-world harms (Bender and Friedman, 2018). For instance, Lee’s (2023) checklist includes four central components: the cultural heritage collection as data (including data composition, the collecting process and curation rationale, the digitization pipeline, data provenance, and crowd labor); the machine learning model as instrumentation (including

training, evaluation, deployment, release and environmental impact); organizational consideration (e.g., stakeholders' interests, use of machine learning, and project deployment); and security measures (e.g., copyright, transparency, documentation, maintenance, and privacy).

Implementing the ethical norms and standards of the biosphere within the information environment and incorporating established principles of accountability, fairness, equity, and transparency into knowledge representation, organization, and discovery systems are crucial steps toward fostering systems that are ethical and democratic systems of representation, organization, and discovery (Hajibayova, 2019). Applying Floridi and Cowls' (2022) principles to information representation and organization, the following tenets of ethical application of AI in authorship are proposed:

- **Beneficence (do only good):** Given the significant environmental impact of AI technologies, it is crucial that the utilization of AI advancements is aligned with values that prioritize environmental sustainability and human well-being. Moreover, careful consideration must be given to the entire AI lifecycle used in knowledge organization systems, from data processing to hardware production, which involves designing, manufacturing, assembling, and distributing the physical components that power AI systems.
- **Nonmaleficence (do no harm):** Warrant that knowledge representation, organization and discoverability systems are based on principles of respect for and prevent infringement on individuals' personal privacy and security, and that they avoid any negative impacts of overusing or misusing AI. This provision also implies that individuals' original creative works should be safeguarded from misuse by AI technologies, including the replication, modification, and distribution of individuals' creative works in violation of their copyright agreement. The potential overuse or misuse of AI could also be addressed through more refined bibliographic descriptions of the resources, which may require revision of existing bibliographic standards to ensure transparent and consistent documentation of AI utilization.
- **Autonomy (maintain human precedence):** Acknowledge that knowledge representation, organization and discovery systems are based on the principles of balance between human- and AI- driven decision- making power, wherein humans retain authority to generate and implement associated decisions concerning ways to represent, organize and provide access to knowledge. This provision implies that original human-created works are represented, organized and made accessible, and AI-generated content is clearly distinguished with information regarding provenance and copyright. For example, source and provenance of the resource could be specified using field 588 and/or subfield \$7 of the MARC bibliographic standard (PCC, 2025).
- **Justice (promote social justice):** Ensure the unbiased and fair representation, organization and discovery systems that promote diversity and justice. Eliminate any unfair treatment and discrimination arising from the use of AI technologies, particularly concerning author attribution, author intellectual property rights, subject analysis and metadata creation.
- **Explicability (support user criticality):** Ensure the comprehensibility, transparency, and accountability of AI-driven solutions by pursuing understanding of the ways in which AI may support and/or constrain individuals' autonomy as well as taking measures to avoid the latter. This provision also implies accountability and transparency in the application of AI-based solutions for representing, organizing and discovering knowledge.

Floridi and Cowls (2019) argue that the “development and use of AI hold the potential for both positive and negative impact on society, to alleviate or to amplify existing inequalities, to cure old problems, or to cause new ones” and that the socially preferable course will depend on well-designed regulations and standards as well as the “use [of a] framework of ethical principles, within which concrete action can be situated” (Floridi and Cowls, 2019, p. 10). Similarly, the application of AI solutions in knowledge representation, organization and discovery systems creates opportunities to enhance knowledge discovery; however, it could also exacerbate the biases and limitations inherent in current knowledge organization systems. Thus, it is now a critical time develop and apply ethical and human-centered AI solutions to ensure equitable, fair and transparent access to knowledge.

The ethical norms and standards of the biosphere are also well-aligned with Foucault’s conceptualization of work and authorship, which promotes a deep understanding of how content can be evaluated on its own merits, irrespective of its authorship. By adhering to universal ethical standards, knowledge representation and organization systems can ensure that within their domain AI technologies are developed and utilized in ways that are ethical, transparent, and beneficial. The application of AI to facilitate understandable, fair, transparent and accountable access to work is central to knowledge representation and organization. Applying principles of accountability and transparency to the use of AI helps to streamline the process of knowledge representation and organization, facilitating users’ discovery and engagement with intellectual content. From an applied perspective, this may require a more nuanced and comprehensive representation of work that includes description of not only the provenance and source of the work, but also processes and methodologies behind the creation. By supporting transparency and accountability, this enhanced documentation of the work would enable users to better assess the credibility, authenticity, and ethical grounding of the materials they access. By embedding these elements into knowledge representation and organization guidelines and standards, information professionals can foster informed use, preserve contextual integrity, and strengthen trust in the information environment.

Conclusion

In light of the growing impact of AI on people’s daily lives, it is crucial to understand the role of AI in knowledge representation, organization and discovery, and to establish appropriate ethical AI principles. Many users of currently available AI tools lack sufficient experience with them or knowledge about the ethical implications of their usage, potentially leading to unintended harm caused by these systems. It is certainly naïve to expect AI to fully simulate human creativity, yet it is important to embrace the technological advancements that shape the new world of creative co-creation. Given the evolving nature of authorship and works, and the increasingly blurred lines between human and machine-generated content, it becomes crucial to ensure that the application of AI is supportive of the ethical, inclusive and supportive of equitable representation and organization of knowledge (Zhitomirsky-Geffet and Hajibayova, 2020). The design and development of ethically responsible AI involves navigating competing values, understanding the regulatory and policy context into which applications are placed, and designing systems that can effectively respond to ethical evaluation (Slota *et al.*, 2023)

Foucault's (2002) question, "What difference does it make who is speaking?" (p. 81), which is fundamental to this inquiry and to which there is no single or simple answer, encourages a deeper evaluation of the inherent complexities of the characteristics and functions of discourse. In the context of AI development and its application in cultural production, Foucault's perspective raises awareness of the broader social, cultural, and political contexts within which discourse operates. Moreover, it challenges the custodians of knowledge to examine how discourse functions, how it is disseminated, and how it both influences and is influenced by various actors and platforms within and across different contexts. This approach promotes a nuanced and comprehensive understanding of discourse, encouraging its assessment not only through the lens of authorship, but also on its own merits as a path to knowledge. That is, one should ask not only who is speaking but also what they are saying, to whom, and why. Foucault's perspective shifts the focus from the author to the conditions and contexts in which discourse is produced and disseminated, emphasizing the importance of understanding the provenance, usage, and appropriation of texts (Wilson, 2004) in an era increasingly shaped by AI-generated content, which challenges traditional notions of authorship and highlights the evolving nature of intellectual work. The ethical norms and standards of the biosphere are also well-aligned with Foucault's conceptualization of work and authorship, which promotes a deep understanding of how content can be evaluated on its own merits, irrespective of its authorship, "at the same time acknowledging the importance of clarifying authorship as a component of a work's identity" (Pugh, 2025, personal communication).

References

- Arthur L. S. (1960), "Some Moral and Technical Consequences of Automation—A Refutation." *Science*, Vol. 132 No. 3429, pp. 741-42. doi:[10.1126/science.132.3429.741](https://doi.org/10.1126/science.132.3429.741)
- Barad, K. (2003), "Posthumanist performativity: Toward an understanding of how matter comes to matter", *Journal of Women in Culture and Society*, Vol. 28 No. 3, pp. 801-831.
- Bassett, C. (2025), "The author, poor bastard: Writing, creativity, AI", Slocombe, W. and Liveley, G. (Eds.), *The Routledge Handbook of AI and Literature (Routledge Literature Handbooks)*, Routledge, New York, NY, pp. 19-26.
- Bender, E.M. and Friedman, B. (2018), "Data statements for natural language processing: Toward mitigating system bias and enabling better science", *Transactions of the Association for Computational Linguistics*, Vol. 6, pp. 587-604. https://doi.org/10.1162/tacl_a_00041
- Bensinger, G. (2023), "ChatGPT launches boom in AI-written e-books on Amazon", *Reuters*, 21 February, available at: <https://www.reuters.com/technology/chatgpt-launches-boom-ai-written-e-books-amazon-2023-02-21/> (accessed 6 August 2025).
- Bimo, S. (2025), "Emerging models of AI 'Authorship' in popular discourse", Slocombe, W. and Liveley, G. (Eds.), *The Routledge Handbook of AI and Literature (Routledge Literature Handbooks)*, Routledge, New York, NY, pp. 49-59.
- Brewer, R. (2023), "Understanding voice-based information uncertainty: A case study of health information seeking with voice assistants", *Journal of the Association for Information Science and Technology*, Vol. 75 No. 10, pp. 1041-1057. <https://doi.org/10.1002/asi.24854>
- Bostrom, N. (2014), *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, Oxford, UK.

- Cox, A. (2023), "How artificial intelligence might change academic library work: Applying the competencies literature and the theory of the professions", *Journal of the Association for Information Science and Technology*, Vol. 74 No. 3, pp. 367-380.
- Cox, A. M. and Mazumdar, S. (2022), "Defining artificial intelligence for librarians", *Journal of Librarianship and Information Science*, Vol. 56 No. 2, pp. 330-340. <https://doi.org/10.1177/09610006221142029>
- Cox, A., Pinfield, S.M. and Rutter, S. (2019), "The intelligent library: Thought leaders' views on the likely impact of artificial intelligence on academic libraries", *Library Hi Tech*, Vol. 37 No. 3, pp. 418-435. doi:10.1108/LHT-08-2018-0105
- ExLibris Knowledge Center (n.d.). The AI metadata assistant in the metadata editor, available at:
[https://knowledge.exlibrisgroup.com/Alma/Product_Documentation/010Alma_Online_Help_\(English\)/Metadata_Management/005Introduction_to_Metadata_Management/The_AI_Metadata_Assistant_in_the_Metadata_Editor](https://knowledge.exlibrisgroup.com/Alma/Product_Documentation/010Alma_Online_Help_(English)/Metadata_Management/005Introduction_to_Metadata_Management/The_AI_Metadata_Assistant_in_the_Metadata_Editor) (accessed on 1 August 2025)
- Fleischmann, R.K., Hui, C. and Wallace, W.A. (2017), "The societal responsibilities of computational modelers: Human values and professional codes of ethics", *Journal of the Association for Information Science and Technology*, Vol. 68 No. 3, pp. 543–552. <https://doi.org/10.1002/asi.23697>
- Floridi, L. (2025). Distant writing: Literary production in the age of Artificial Intelligence. *Minds & Machines*, Vol. 35 No. 30, pp. 1-26. <https://doi.org/10.1007/s11023-025-09732-1>
- Floridi, L. and Cowls, J. (2022), "A unified framework of five principles for AI in society", Carta, S. (Ed.), *Machine Learning and the City: Applications in Architecture and Urban Design*, John Wiley & Sons, Hoboken, NJ, pp. 535-545.
- Floridi, L. and Cowls, J. (2019), "A Unified Framework of Five Principles for AI in Society." *Harvard Data Science Review*, Vol. 1 No. 1, pp. 1-14. <https://doi.org/10.1162/99608f92.8cd550d1>
- Floridi, L. (2013), *The Ethics of Information*, Oxford University Press, Oxford, UK.
- Floridi, L. (2008), "Foundations of information ethics", Himma, K. E. and Tavani, H.T. (Eds.), *The Handbook of Information and Computer Ethics*, John Wiley & Sons, Hoboken, NJ, pp. 3-24.
- Foucault, M. (2002), "What is an Author?", Arnold, D. (Ed.), *Reading Architectural History*, Routledge, London, UK, pp. 71-81. <https://doi.org/10.4324/9780203164471>
- Fox, M.J. and Reece, A. (2012), "Which ethics? Whose morality? : An analysis of ethical standards for information organization. *Knowledge Organization*, Vol. 39 No. 5, pp. 377-383.
- Friedman, B. and Kahn Jr, P.H. (2007), "Human values, ethics, and design", Sears, A. and Jacko, J.A. (Eds.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, Lawrence Erlbaum Associates, New York, NY, pp. 1241-1266.
- Golub, K. (2021), "Automated Subject Indexing: An Overview", *Cataloging & Classification Quarterly*, Vol. 59 No. 8, pp. 702-719. <https://doi.org/10.1080/01639374.2021.2012311>
- Hagendorff, T. (2020), "The ethics of AI ethics: An evaluation of guidelines", *Minds and Machines*, Vol. 30 No. 1, pp.99-120.
- Hajibayova, L. (2019), "Guardians of the knowledge: relevant, irrelevant, or algorithmic?" *Information Research* Vol. 24 No. 4. Paper 839. <http://InformationR.net/ir/24-4/paper839.html>

- Hirvonen, N., Jylhä, V., Lao, Y. and Larsson, S. (2023), "Artificial intelligence in the information ecosystem: Affordances for everyday information seeking", *Journal of the Association for Information Science and Technology*, Vol. 75 No. 10, pp. 1152-1165.
<https://doi.org/10.1002/asi.24860>
- Hosseini, M, Rasmussen, L.M. and Resnik, D. B. (2023), "Using AI to write scholarly publications", *Accountability in Research*, Vol. 31 No. 7, pp. 715-723.
doi:10.1080/08989621.2023.2168535
- Hongisto, T. (2025), "A Brief History of Computer-generated Literature: In Search of the Author", Slocombe, W. and Liveley, G. (Eds.), *The Routledge Handbook of AI and Literature*, Routledge, New York, NY, pp. 36-48.
- Hu, Y. (2023), "Literature in the Age of Artificial Intelligence", in Mohd Hussain, R.B.B., Parc, J., & Li, J. (Eds.), *Proceedings of the 2023 9th International Conference on Humanities and Social Science Research (ICHSSR 2023)*, Beijing, China, pp. 1781-1787. doi: [10.2991/978-2-38476-092-3_228](https://doi.org/10.2991/978-2-38476-092-3_228)
- Hu, Y., Li, W., Wright, D., Aydin, O., Wilson, D., Maher, O. and Raad, M. (2019), "Artificial intelligence approaches", Wilson, J.P. (Ed.), *The Geographic Information Science & Technology Body of Knowledge* (3rd Quarter 2019 Edition).
doi:[10.22224/gistbok/2019.3.4](https://doi.org/10.22224/gistbok/2019.3.4).
- Jobin, A., Ienca, M. and Vayena, E. (2019), "The global landscape of AI ethics guidelines", *Nature Machine Intelligence*, No. 1, pp. 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Kariyawasam, K. (2020), "Artificial intelligence and challenges for copyright law", *International Journal of Law and Information Technology*, Vol. 28 No. 4, pp. 279-296.
<https://doi.org/10.1093/ijlit/ehaa023>
- Kitzie, V. (2019), "That looks like me or something I can do": Affordances and constraints in the online identity work of US LGBTQ+ millennials." *Journal of the Association for Information Science and Technology*, Vol. 70 No. 12, pp. 1340-1351.
<https://doi.org/10.1002/asi.24217>
- Lee, B. C. G. (2023), "The 'Collections as ML Data' checklist for machine learning and cultural heritage", *Journal of the Association for Information Science and Technology*, Vol. 76 No. 2, pp. 1-22. <https://doi.org/10.1002/asi.24765>
- Library of Congress Labs (n.d.). Newspaper Navigator, available at: <https://labs.loc.gov/work/experiments/newspaper-navigator> (access 1 August 2025).
- Limbong, A. (2024), "Authors push back on the growing number of AI 'scam' books on Amazon", *NPR*, 13 March, available at: <https://www.npr.org/2024/03/13/1237888126/growing-number-ai-scam-books-amazon> (accessed 6 August 2025)
- Lin, S., Hilton, J. and Evans, O. (2022), "TruthfulQA: Measuring how models mimic human falsehoods", in Muresan, S., Nakov, P., & Villavicencio, A. (Eds.), *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, Dublin, Ireland, pp. 3214-3252. doi:10.18653/v1/2022.acl-long.229
- Lu, B. (2021), "A theory of 'authorship transfer' and its application to the context of Artificial Intelligence creations", *Queen Mary Journal of Intellectual Property*, Vol. 11 No. 1, pp. 2-24. <https://doi.org/10.4337/qmjip.2021.01.01>
- Lund, B.D., Wang, T., Mannuru, N.R., Nie, B., Shimray, S. and Wang, Z. (2023), "ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of

- the large language models in scholarly publishing”, *Journal of the Association for Information Science and Technology* Vol. 74 No. 5, pp. 570-81.
<https://doi.org/10.1002/asi.24750>
- Lund, B.D. and Naheem, K.T. (2024), “Can ChatGPT be an author? A study of artificial intelligence authorship policies in top academic journals”, *Learned Publishing* Vol. 37 No. 1, pp. 13-21.
- Mai, J.-E. (2014), “Ethics, values and morality in contemporary library classifications”, *Knowledge Organization* Vol. 40 No. 4, pp. 242-53.
- Mannheimer, S., Bond, N., Young, S. W. H., Kettler, H. S., Marcus, A., Slipher, S. K., ... Sheehy, B. (2024), “Responsible AI practice in libraries and archives: A review of the literature”, *Information Technology and Libraries*, Vol. 43 No. 3.
<https://doi.org/10.5860/ital.v43i3.17245>
- McCarthy, J., Minsky, M. L., Rochester, N. and Shannon, C.E. (2006), “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955”, *AI Magazine*, Vol. 27 No. 4, p. 12. <https://doi.org/10.1609/aimag.v27i4.1904>
- Nathan, L. P. (2012), “Sustainable Information Practice: An Ethnographic Investigation”, *Journal of the American Society for Information Science and Technology* Vol. 63 No. 11, pp. 2254-68. <https://doi.org/10.1002/asi.22726>
- Nazarovets, S. and Teixeira da Silva, J. A. (2024), “ChatGPT as an ‘Author’: Bibliometric Analysis to Assess the Validity of Authorship”, *Accountability in Research*, pp.1–11. doi:10.1080/08989621.2024.2345713.
- Short, M. (2019), “Text mining and subject analysis for fiction; or, using machine learning and information extraction to assign subject headings to dime novels”, *Cataloging & Classification Quarterly*, Vol 57 No. 5, pp. 315–336.
<https://doi.org/10.1080/01639374.2019.1653413>.
- Schuerkamp, R., Barrett, J., Bales, A., Wegner, A. and Giabbanelli, P.J. (2023), “Enabling new interactions with library digital collections: automatic gender recognition in historical postcards via deep learning”, *The Journal of Academic Librarianship*, Vol. 49 No. 4, pp.1027-1036.
- Smiraglia, R.P. (2003), “The History of ‘The Work’ in the Modern Catalog”. *Cataloging & Classification Quarterly*, Vol. 35 No. 3-4, pp. 553-567. doi:10.1300/J104v35n03_13
- Pine, K.H. and Liboiron, M. (2015), “The politics of measurement and action”, *CHI’15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Seoul, Korea, pp. 3147-3156. <https://doi.org/10.1145/2702123.270229>
- Program for Cooperative Cataloguing (PCC). (2025), “PCC Participants Meeting”, available at: <https://www.loc.gov/aba/pcc/documents/PCC-Participants-2025-02-13-slides.pdf> (accessed 6 August 2025).
- Russell, S.J. and Norvig, P. (2016), *Artificial Intelligence: A Modern Approach*, Pearson, New York, NY.
- Schank, R.C. (1987), “Where's the AI, Anyway?” *AI Magazine*, Vol. 8 No. 4, p. 59.
doi:<https://doi.org/10.1609/aimag.v8i4.623>
- Schultz, D. C., Koch, C. and Olbrich, R. (2024), “Dark sides of artificial intelligence: The dangers of automated decision-making in search engine advertising”, *Journal of the Association for Information Science and Technology*, Vol. 75 No. 5, pp. 550–566. <https://doi.org/10.1002/asi.24798>

- Slota, C. S., Fleischmann, K.R., Greenberg, S., Verma, N., Cummings, B., Li, L. and Shenefiel, C. (2023), "Locating the work of artificial intelligence ethics", *Journal of the Association for Information Science and Technology*, Vol. 74 No. 3, pp. 311-322. <https://doi.org/10.1002/asi.24638>
- Smith, L. C. (1976), "Artificial intelligence in information retrieval systems", *Information Processing & Management*, Vol. 12 No. 3, pp. 189-222. [https://doi.org/10.1016/0306-4573\(76\)90005-4](https://doi.org/10.1016/0306-4573(76)90005-4).
- Strachey, C. (1954), "The 'thinking' machine", *Encounter*, Vol. 3, pp. 25–31.
- Slocombe, W. and Liveley, G. (2025), *The Routledge Handbook of AI and Literature* (Routledge Literature Handbooks), Routledge, New York, NY.
- Zhitomirsky-Geffet, M. and Hajibayova, L. (2020), "A new framework for ethical creation and evaluation of multi-perspective knowledge organization systems", *Journal of Documentation*, Vol. 76 No. 6, pp. 1459-1471. <https://doi.org/10.1108/JD-04-2020-0053>
- van Strien, D., Bell, M., McGregor, N.R. and Trizna, M. (2022), "An Introduction to AI for GLAM", in Kinnaird, K.M., Steinbach, P., & Guhr, O. (Eds.), *Proceedings of the Second Teaching Machine Learning and Artificial Intelligence Workshop*, pp. 20-24. <https://proceedings.mlr.press/v170/strien22a.html>
- Wang, X, Lin, X. and Shao, B. (2022), "Artificial intelligence changes the way we work: A close look at innovating with chatbots", *Journal of the Association for Information Science and Technology*, Vol. 74 No. 3, pp. 339–353. <https://doi.org/10.1002/asi.24621>
- Wilson, A. (2004), "Foucault on the 'Question of the author': A critical exegesis", *Modern Language Review*, Vol. 99 No. 2, pp. 339-363.