



Information Research – Vol. 31 No. 3 (2026)

The information professional as an ethical intermediary: information quality criteria and the ethics of generative AI

Thomas D. Wilson and Elena Maceviciute

DOI: <https://doi.org/10.47989/ir31365045>

Abstract

Introduction. The rapid adoption of generative AI raises significant ethical concerns for information professionals whose intermediary role requires them to ensure information quality standards that users cannot independently verify. Professional ethics statements address quality only superficially, and the ethical implications of generative AI have not been examined from the perspective of professional quality obligations.

Method. A scoping review assessed 644 records across three search strands, retaining 120 papers. Retained papers were analysed against ten information quality criteria and assessed for the strength of their ethical framing.

Analysis. Ten practice-derived quality criteria are shown to be groundable in a Kantian deontological framework. The six distinctive characteristics of generative AI are shown to constitute systematic violations of those criteria. A Marxist political economy analysis explains why these violations are structural rather than incidental, and why market forces will not correct them voluntarily.

Results. The criteria attract variable coverage in the literature: some, such as accuracy and objectivity, are well covered with developed ethical treatment, while others, such as currency, receive little attention. This asymmetry is itself a significant finding, reflecting visibility and professional tradition rather than the relative importance of the obligations involved.

Conclusions. The information professional's intermediary role is an ethical necessity. Market forces will not remedy AI's quality failures voluntarily; regulation, informed by professional expertise, is required. Professional organisations are called upon to engage in AI governance, through policy advocacy, standards development and regulatory consultation, using the criteria framework developed here.

Introduction

Artificial intelligence (AI) systems have been in use for years in many kinds of organisations. For example, the field of medicine was an early user with expert systems such as DENDRAL for chemical analysis and MYCIN for medical diagnosis, as early as 1971 and 1975 (Wu et al., 2025). With the emergence of the Internet and the World Wide Web, applications began to affect everyday users, exemplified by voice recognition systems like Siri (O'Shaughnessy, 2024) and recommender systems used by platforms such as Netflix (Dong et al., 2022). However, the release of ChatGPT in November 2022 gave users a system they could interact with, ask questions about everyday issues, and get well-constructed, intelligible answers. ChatGPT's success in attracting users was phenomenal: in a little more than three years it had reached 900 million weekly users (OpenAI, 2026), and there was a rapid proliferation of competitors offering similar conversational interfaces based on generative pre-trained transformers. These are highly sophisticated systems but, as one prominent AI scientist has argued, they are fundamentally statistical pattern matchers rather than reasoning agents (Adam, 2026).

It did not take long before ethical issues began to be raised about the nature of AI systems, their outputs and their use. In fact, concerns were raised before ChatGPT was released (see for example, Pazzanese, 2020), and since then the concerns have increased. The concerns operate on two levels: the ethics built into (or failing to be built into) the design of the systems; and the ethical use of such systems. These two levels interact, of course: for example, bias in the training data may result in biased outputs in response to a prompt. While the two levels interact, the literature tends to treat them separately. But how can ethics be understood in this framework?

The ethical framework

A number of ethical frameworks exist, particularly virtue ethics (based on the work of Aristotle and concerning the development of the ethical individual), and consequentialist ethics (associated with Jeremy Bentham and John Stuart Mill, which assumes that the ethical rightness of an action is determined solely by its consequences). In earlier work, it has been argued that the advice given by universities to students on the use of generative AI systems implicitly reflects Kantian principles, even when it does not name them as such (Wilson, 2025a, 2025b). Therefore, the ethical framework adopted in this paper is the deontological ethics of Immanuel Kant (Kant, 1785; O'Neill, 1993). Deontological ethics, from the Greek *deon*, meaning duty, holds that moral obligations are grounded in duty rather than in the consequences of actions or in personal virtue. This makes it the natural framework for examining the obligations of information professionals, whose role is defined precisely by duties to their clients: the duty to provide accurate, current, sufficient, and accessible information regardless of the effort involved or the convenience of alternatives. Kant's categorical imperative, the requirement to act only according to principles that could be universally applied, translates directly into the professional standard that the same quality of service must be provided to every user. This might include refusing to censor materials or denying access. These principles must apply equally to everyone, universally, regardless of personal convenience. Equally, respecting library user autonomy should also be universally applied, for example, by protecting intellectual privacy and providing unfiltered, unbiased access to information so users can independently form their own rational conclusions (Fallis, 2007).

The four principles most relevant to this paper are: *respect for persons*, which requires treating users as rational agents with specific needs rather than as generic recipients; *universalisability*, which requires that professional standards apply consistently across all interactions; *duty over inclination*, which requires maintaining quality standards; and *the protection of rational agency*, which requires providing information in a form that supports the user's own judgement.

Kantian principles' idealistic nature was seriously opposed by Marxist philosophers on the grounds that their implementation is dependent on material circumstances and, therefore, cannot be

universal. However, neo-Marxists have not opposed the principles as such and have recognised that they can be aimed at historical progress, which ‘is rooted in ethical judgements that evaluate the relationships between means and ends, presupposing the idea of the unity of ends’, which are grounded in the concept of humanity (Widmer, 2024, pp. 609–610). Kant explains universal duties; neo-Marxism explains structural constraints and social conditions. Thus, we regard neo-Marxism as a useful complement to deontological ethics in dealing with the issues of power related to modern technologies, especially as their representatives are active critics of information technology structural features.

Neo-Marxist critique has introduced the concept of algorithmic surplus value, a digital-era intensification of Marxian surplus value, where AI and algorithms create profit and surplus not autonomously and independently, but by compressing necessary labour time and controlling digital platforms (Zhang, 2026), thereby increasing global inequity and the power of technical oligarchies (Greenstein, 2022). Noble (2018) argues that digital tools built in social contexts plagued by specific inequalities (social, racial, or gendered) reinforce these inequalities and reaffirm hegemonies by transporting them to the future. AI systems, on the one hand, save human labour and provide means for social reproduction, the everyday care, educational, and communicative tasks that sustain households and communities, but on the other hand, they generate distributive inequalities, epistemic distortions and diminish human autonomy, creating cognitive alienation, wherein users surrender interpretative autonomy to machine logic (Shin, 2025). This trend in AI ethical assessment literature creates grounds for the implementation of means to prioritise the rule of law (Greenstein, 2022), human development needs, and social justice (Santoni de Sio et al., 2024) as well as for translation of abstract ethical concerns into practical algorithmic modules (Burriss et al., 2024). These concerns and the potential for their remediation are of immediate interest to practicing librarians and information professionals.

Any information professional, librarian, or information manager has ethical responsibilities for the information they supply, or make accessible, to the information user. The various professional associations have ethics statements, but these address quality only superficially (e.g. IFLA, 2012). The American Library Association (2021), for example, commits members to providing accurate, unbiased and courteous responses, while the CILIP Ethical Framework (2018) states only that information should be impartial and free of bias. However, the working information professional is aware that other quality characteristics are expected if the information seeker is to be fully served. From our library and information professional practice, we identify the following ten characteristics: *relevance and fitness for purpose*; *currency*; *accuracy*; *sufficiency*; *objectivity and freedom from systematic bias*; *understandability*; *identifiable origin*; *consistency*; *accessibility*; and *presentation without causing information overload*. These practice-derived criteria align to some extent with ISO/IEC 25012 (2008), the international standard on data quality, demonstrating convergence between formal technical standards and professional service expectations. Thus, our characteristics *accuracy*, *consistency*, *currency*, *accessibility* and *understandability* are all represented in the standard. There is also some equivalence between the standard’s *completeness* and our *sufficiency*, since complete information is rarely required and the information user will usually cease searching when they feel they have enough information. The standard’s *traceability* is equivalent to our *identifiable origin*. The differences between the two sets of criteria are the result of the standard being concerned with *data quality* in a computer-based system, rather than with information sources in general. There is also overlap with the quality criteria identified by Laumer et al. (2017), who specify the quality criteria as conciseness, understandability, completeness, relevance, timeliness, and usefulness.

Aim and research questions

The ethical issues of generative AI systems have not been examined from the perspective of the information professional’s duty regarding the quality of information provided for the user, as far

as we can determine. The literature on AI ethics in libraries, for example, tends to deal with policy issues, and particular problems such as academic integrity and a new form of plagiarism. The professional's obligations, defined in terms of the ten quality criteria, can be grounded in a Kantian deontological framework, and with the help of neo-Marxist criticism we argue that the design characteristics of generative AI systems result in specific violations of that framework. It is this gap that this paper seeks to address, through the following research questions:

1. What are the implications of generative AI's distinctive characteristics for each of the ten information quality criteria through which information professionals define their service obligations to users?
2. To what extent have the quality failure modes of generative AI been treated as ethical rather than merely technical concerns in the relevant literature?
3. What are the implications for the information professional's intermediary role in practice, and for information professional organisations' engagement in the governance and design of generative AI systems?

Method

This is a conceptual paper based on a scoping review of the relevant literature. A scoping review was chosen, in preference to a systematic review, as we wished to map the extent, range and nature of the literature on the subject, identifying gaps, key concepts, and evidence on the extent to which quality issues are seen as ethical concerns (Arksey and O'Malley, 2005).

The review is based on three complementary strands of search: Google Scholar Alerts relating to the ethical basis of generative AI design were collected for five weeks; six targeted Web of Science searches covering information quality, the professional role of librarians and information professionals, credibility and source evaluation, professional ethics, and epistemic authority; and the results of a literature search for a related paper (Wilson, 2026). These three strands captured largely non-overlapping sets of papers, suggesting that the combined literature base is more comprehensive than a single search would have been.

Approximately 100 papers identified through the Google Scholar alerts on the search term 'ethical basis of generative AI design' were analysed and organised into two groups: Group 1 (ethical principles within generative AI systems: design, alignment and governance) and Group 2 (ethics of use in context: professions, institutions and vulnerable populations). From this set, 52 papers were retained after thematic analysis as the most relevant to one or more of the quality characteristics. The alert set was analysed before and independently of the Web of Science searches represented in Table 1.

Search	Records retrieved	Records retained
WoS Search 1: Information quality AND generative AI / large language models	27	12
WoS Search 2: Professional role - librarian / information professional (library and information science category restricted)	296	30
WoS Search 3: Credibility / trustworthiness / source evaluation	20	8
WoS Search 4: Professional role - variant terms; all duplicates of Search 2	269	0
WoS Search 5: Information / library / professional ethics (library and information science category restricted)	15	4
WoS Search 6: Epistemic authority / justice	26	3
WoS Search 7: Currency / timeliness / training cutoff	18	7
WoS Search 8: Readability / intelligibility / plain language	32	11
WoS Search 9: Consistency / reproducibility / variability	23	8
Total WoS records	726	83
Minus duplicates across WoS searches (estimated)	-250	—
Net unique WoS records assessed	476	83
Google Scholar alert set (ethical basis of generative AI design)	138	55
Carried forward from Wilson (2026)	30	30
Total unique records assessed across all sources	644	120

Table 1. Literature search: records retrieved and retained

Six initial targeted searches were conducted in March 2026 using the Web of Science database, with the subject category restricted to *Information science*, *Library science* when appropriate. Three further searches were conducted to address specific gaps identified in the initial coverage. The searches covered: information quality and generative AI; the professional role of librarians and information professionals; credibility, source evaluation, and trustworthiness; professional and information ethics; and epistemic authority. Three searches were then conducted to address specific gaps identified in the initial coverage assessment: currency and timeliness; readability and intelligibility; and consistency and reproducibility. The full search queries, record counts, and retention decisions are reported in Table 1. The retention criteria remain the same as before: the relevance to the information quality characteristics and their ethical treatment aspects.

An additional thirty papers identified in the literature review for a companion paper on information behaviour in AI-mediated environments (Wilson, 2026) were carried forward as directly relevant to this paper. The literature base was further supplemented during the drafting process by additional papers identified through targeted searches, where gaps in coverage of specific criteria became apparent.

The retained papers were analysed by the first author thematically against two coding dimensions. First, each paper was assigned to one or more of the ten information quality criteria on the basis of its primary subject matter. Second, each paper was assessed for the presence and strength of ethical context: papers were coded as having strong, explicit ethical context (where the quality failure is seen as an ethical problem for developers, institutions or information professionals), implicit ethical context (where the ethical dimension is present but not the paper's primary focus), or no ethical context. The second author reviewed the coding results for coherence and accuracy against the coding scheme; both discussed any ambiguities or inaccuracies identified and adjusted when necessary. This two-dimensional coding provides the structure for the literature review and the basis for the Discussion section's argument about where ethical development is and is not present in the current literature.

Literature review

Structure and scope

The literature review is organised around ten distinct information quality criteria. For each criterion, the analysis addresses three specific dimensions: the particular AI failure mode that threatens it; the empirical and theoretical literature documenting that failure; and the degree to which the existing research treats these failures as ethical problems rather than purely technical or empirical challenges.

In identifying these failure modes, the review builds upon the six characteristics of generative AI established in Wilson (2026): opacity, epistemic authority, sycophancy risk, personalisation, first-person voice construction and commercial optimisation.

The criteria have variable coverage in the literature: for example, *accuracy* and *objectivity* are well covered, while *currency* is less well covered. The relative scarcity of literature in some areas is itself a significant finding, the implications of which are explored further in the Discussion section.

Accuracy

When ChatGPT-3.5 was launched in November 2022, its designers were already aware that it could produce inaccurate information through hallucination. The ChatGPT blog (OpenAI, 2022) noted:

ChatGPT sometimes writes plausible-sounding but incorrect or nonsensical answers. Fixing this issue is challenging, as: (1) during RL training, there's currently no source of truth; (2) training the model to be more cautious causes it to decline questions that it can answer correctly; and (3) supervised training misleads the model because the ideal answer depends on what the model knows, rather than what the human demonstrator knows.

By the release of ChatGPT-4 in March 2023, performance had improved modestly. The GPT-4 Technical Report (OpenAI, 2023) recorded a 19% improvement on the TruthfulQA benchmark, with GPT-4 scoring 60%, compared to 41% for GPT-3.5. However, this still implied a 40% failure rate on a test designed to measure truthfulness. The report itself cautioned that GPT-4 'is not fully reliable (it "hallucinates" facts and makes reasoning errors). Great care should be taken when using language model outputs ...' (p. 10). More importantly, GPT-4's greater fluency and confidence may make errors harder rather than easier for users to detect, directly implicating the problem of epistemic authority identified in Wilson (2026).

Early academic scrutiny confirmed domain-wide limitations: Kung et al. (2023) found ChatGPT approaching but not reliably reaching clinical examination pass thresholds, while Plevris et al. (2023) showed that responses to complex tasks were frequently unreliable and inconsistent across repeated prompts. Subsequent research has consistently reinforced two findings: surface fluency does not guarantee factual reliability, and accuracy cannot be assumed to be stable across interactions.

Accuracy failures are particularly acute in relation to citations and references. Oladokun et al. (2025) examined ChatGPT's responses to librarianship questions, finding that 42.9% of citations generated by ChatGPT-3.5 were false or non-existent, rising to 51.8% in ChatGPT-4o. Verified citation accuracy was correspondingly low: 3.92% and 6.35% respectively. A similar pattern appears in clinical contexts: Blasingame et al. (2025) reported that although 83.3% of GPT-4 responses to clinical evidence questions were judged correct overall, only 63% of the references could be verified as non-fabricated. These findings highlight a crucial distinction for information professionals: a response may appear substantively accurate while its attribution is fundamentally unreliable.

The most systematic comparison of large language model (LLM) performance against human expert standards is provided by Wei and Chen (2026). Evaluating both Dense and Mixture-of-Experts architectures across six information quality dimensions, they found that both model types lagged behind human experts on intrinsic and contextual information quality—dimensions directly concerned with accuracy and fitness for purpose.

The ethical dimensions of accuracy failures were articulated even before the release of ChatGPT-3.5. Bender et al. (2021) argued that LLMs do not understand language but generate outputs through statistical pattern matching, without grounded meaning, making hallucinations a structural feature rather than a fixable flaw. Following ChatGPT's release, Lund et al. (2023) reframed AI-generated content as an ethical issue for scholarly communication, raising concerns about authorship, attribution, and research integrity when inaccurate AI outputs are presented as authoritative.

Sima et al. (2026) extend this argument by distinguishing responsible AI use, characterised by transparency, validation, and human oversight, from careless or deceptive use that risks introducing inaccurate or unverifiable content into the scientific record. They conclude that careless AI use undermines scientific credibility and trust. Freyer et al. (2024), writing from the perspective of text simplification for marginalised groups, add a further ethical dimension: inaccuracies introduced during AI-driven simplification may be invisible to people with cognitive disabilities or limited literacy, precisely those least able to detect them. This links the accuracy criterion directly to accessibility and reinforces the information professional's duty to protect vulnerable users.

Objectivity and freedom from systematic bias

Given the scale of misinformation, disinformation, and conspiracy theories on the Web, the information professional's role in ensuring that accurate and unbiased information is available to users has become increasingly significant. However, AI systems' ability to deliver objective information is compromised by biases embedded in their training materials. Foka and Griffin (2024), for example, observe that cultural history collections are inherently biased through processes of selection and description that carry historical prejudices into the present. When digital versions of these collections are used in training generative AI systems, those biases are propagated.

Empirical studies provide convergent evidence of such effects. Lee et al. (2024) identified racial bias in AI-generated ophthalmological educational materials: demographic information, including ethnicity, affected output length, with information for Black patients containing significantly fewer words than that for White patients, although reduced informational content was not definitively established. Fleisig et al. (2024) documented linguistic bias when prompts were written in non-standard English dialects, reporting that 'responses to non-"standard" varieties consistently exhibit a range of issues: stereotyping (19% worse than for "standard" varieties), demeaning content (25% worse), lack of comprehension (9% worse), and condescending responses (15% worse)' (p. 13541). A UNESCO study (van Niekerk et al., 2024) similarly revealed gender and racial bias across several LLMs: women were more often associated with terms such as home, family and children, while men were linked to business, career and executive. When ethnicity was introduced, Zulu men were more likely to be associated with security guard or gardener, and Zulu women with domestic servant, cook and housekeeper. Not all systems exhibited bias uniformly, however: ChatGPT, for example, generated positive or neutral content in over 80% of cases relating to LGBTQ+ subjects.

There is, therefore, strong empirical evidence for systematic bias in generative AI systems, largely traceable to their training data. The ethical implications for information professionals have also been explored. Slater (2025) argues that librarianship has long involved a 'precarious balancing act'

between budgets, emerging technologies, and obligations to patrons. She suggests that the use of generative AI represents 'a shift in the responsibility of facilitating ethical labor practices in LIS, operating at some degree for the public good, onto privatized technological solutions that are constantly changing and fetishized' (p. 590). She further contends that the popularity of generative AI 'does not correlate to ethical standards of information services created and supported by LIS practitioners' (p. 597), concluding that librarians should refuse its use because it constitutes a 'destructive force' that is infrastructurally racist, obscures reliance on scraped digital labour, fails to produce information with integrity without human scrutiny, and contributes unnecessarily to environmental harm (p. 601).

Cox and Jimenez (2025) advance a related but distinct claim, arguing that ChatGPT embeds Western epistemological assumptions that amount to a systematic failure of objectivity. They propose design principles for a decolonial generative AI in response. That such concerns are now widely recognised within the profession is confirmed by Mannheimer et al. (2024) and Ibrahim (2025), both of whom identify bias against diverse user communities as a central ethical challenge for libraries adopting AI. While bias is clearly recognised as a professional ethical issue, the academic literature on appropriate professional responses remains under-developed; frameworks, guidelines, and policy statements are more commonly found on institutional websites than in peer-reviewed research.

Bias also extends to AI-generated representations of the profession itself. Spennemann and Oddone (in press) found that ChatGPT-generated images of librarians predominantly depicted Caucasian figures and significantly overstated the presence of men, reinforcing stereotypical and unrepresentative images of professional identity.

Traceability and source attribution

Generative AI systems present information in an authoritative tone of voice, but users have little insight into the sources on which that information is based and cannot readily judge whether it is genuinely authoritative. While accuracy has been addressed above, traceability is a distinct quality issue. The inability to identify sources (although partially addressed by newer system features) is compounded by first-person voice construction and, in the case of Claude.ai, by the use of a human name, both of which encourage users to treat the system as a human-like informant.

It appears that generative AI systems themselves cannot identify the sources of their training data. When asked about its training sources, ChatGPT acknowledged that its training involved a mixture of licensed data, data created by human trainers, and publicly available material, but confirmed that it has no access to its training data and cannot identify specific sources, drawing instead on internalised knowledge to compose responses (OpenAI, 2026). A similar query put to Claude.ai produced an equivalent response.

Theoretical analysis and empirical evidence for this problem of opacity are well established. Sundin (2025) notes that 'while search engines traditionally provide links to websites and other documents, generative AI chatbots generate content without a stable source other than the LLM itself' (p. 295), observing that hybrid systems combining LLMs with search engines are 'shifting the focus towards searching for content rather than identifying sources' (p. 298). As a result, users increasingly receive content divorced from explicit provenance, even though the data underpinning LLM outputs remains opaque to them. Sundin concludes that this represents a transfer of analytical skills (and agency) from the user to the system, leaving users increasingly dependent not only for retrieval but also for evaluation and interpretation (p. 298).

Heersmink et al. (2024) provide a useful conceptual framework, distinguishing reflective, phenomenological, and informational transparency. Reflective transparency (understanding how the system works and what data it relies on) is largely absent in generative AI systems.

Phenomenological transparency arises as interfaces become familiar and effortless for frequent users, while informational transparency is achieved because outputs are generally easy to understand. Together, phenomenological and informational transparency foster trust and encourage anthropomorphism, while the lack of reflective transparency undermines the epistemic basis for that trust. The authors argue that improving trustworthiness will require greater reflective transparency, including clearer knowledge of training data and algorithmic processes.

A further consequence of opacity is the emergence of responsibility gaps. Taylor (2025) asks who is accountable when decisions are taken on the basis of AI-supplied information, or when decisions are made directly by the system. Although Taylor addresses AI systems broadly, the question applies equally to generative AI: when a human acts on information provided by an opaque system and harmful consequences follow, responsibility remains contested.

From an information literacy perspective, Tomaszewski (2025) proposes CAT (Check, Ask, Think) as an evaluative framework for AI interactions, encouraging users to assess 'source credibility (Check), intent and authorship, whether human or algorithmic (Ask), and platform or emotional context (Think)' (pp. 7–8). While such frameworks offer practical mitigation, the ethical implications of opacity remain under-examined. For information professionals committed to delivering credible information from identifiable and validated sources, the persistent lack of transparency in generative AI systems presents an ethical dilemma: without reliable traceability, professionals cannot guarantee the authority, accuracy, or validity of the information delivered to users.

Accessibility

The accessibility criterion intersects with several others. As noted in the Objectivity section, Cox and Jimenez (2025) argue that the Western, particularly American, bias of training materials limits access to cultural knowledge not represented in LLM training resources, disadvantaging users from other cultural contexts. Slater (2025) similarly contends that these systems reinforce existing inequalities of access by prioritising commercial over public interests. Linguistic and cultural barriers are further documented by Fleisig et al. (2024) and van Niekerk et al. (2024).

Matsieli and Mutula (2025) emphasise that the Global South is particularly vulnerable to these forms of exclusion, given the systematic omission of local cultural resources from training data. Freyer et al. (2024) highlight another dimension of inaccessibility, noting that generative AI systems do little to support users with cognitive disabilities. Economic barriers compound these problems: like many information technologies, generative AI operates on a freemium model, with more capable or useful features restricted to paying users. As Mazel (2025) observes in an examination of AI tools for literature searching, this creates discrimination on economic grounds. She notes that 'to make full use of these tools, users must already have a high level of information privilege due to their reliance on users supplying or being able to access paywalled content' (p. 323).

Simons (2025) situates these concerns within a broader ethical critique. He draws on indigenous and non-Western concepts of wisdom, including the Diné concept of *hozho*, which holds that individual wellbeing is inseparable from one's relationships with community, nature and the universe and the African philosophy of *Ubuntu*, premised on the fundamental interconnectedness of persons, to argue that the profession's response to generative AI must be grounded in relational and intercultural values rather than the dominant Western technocratic framework. This requires not merely avoiding harm but actively embodying the values of inclusion and equity in professional practice. Ford (2026), drawing on case studies of critical AI literacy interventions in public and academic libraries in Greater Sydney, complements this perspective by reframing AI literacy as not only understanding how AI works, but recognising when, how, and why it fails. This positions the librarian as an active mediator rather than a passive conduit for AI-generated content.

Taken together, these arguments point towards a professional response requiring both ethical commitment and practical skill. In Kantian terms, the information professional's duty is to ensure that information is accessible to all users regardless of economic circumstance, language, cognitive capacity, or cultural background, a duty that cannot be discharged simply by deploying AI systems whose design systematically disadvantages some users. Of all ten criteria, accessibility has attracted the most explicit ethical attention, indicating a professional recognition that access to information is not merely a practical concern but a matter of justice.

Relevance and fitness for purpose

The obligation to provide information that is relevant to the user's actual need, and genuinely fit for the purpose for which it is sought, is complicated by one of generative AI's most distinctive characteristics: sycophancy risk. Malmqvist (2025), in a technical survey of sycophancy in large language models, defines the phenomenon as, 'the propensity of models to excessively agree with or flatter users, often at the expense of factual accuracy or ethical considerations' (p. 61), a behaviour unintentionally reinforced during reinforcement learning from human feedback. The result is a system structurally biased towards telling users what they appear to want to hear rather than what they actually need to know, a direct failure of fitness for purpose even when surface relevance appears high.

The empirical evidence confirms that this failure mode is real and consequential. Schulze Balhorn et al. (2024) assessed ChatGPT's responses to 594 questions generated by faculty members across five scientific disciplines at Delft University of Technology. They found that while responses were rated on average as 'mostly correct', ratings declined significantly as the educational level of the question increased, and note that, 'A striking difference between what faculty members expect from student answers and answers from ChatGPT is a critical reflection and a clear understanding of the concepts used' (p. 8).

Goodman et al. (2023), in a study of ChatGPT's responses to 284 medical questions generated by thirty-three physicians across seventeen specialties, found a similar pattern: performance was strong on simple, straightforward questions but degraded in accuracy and completeness as questions became more complex, context-dependent and clinically consequential. Mun and Hwang (2025) confirm from the user side that relevance is among the information quality attributes most strongly associated with perceived usefulness and trust in ChatGPT, suggesting that users are sensitive to relevance failures even when they cannot always articulate them.

No paper in the retained set explicitly frames fitness for purpose failure as a professional ethical issue. Schulze Balhorn et al. (2024) note that their results indicate a lack of critical attitude in AI responses, and observe that the model does not generally reason on the ethical implications of its answers, relying instead on a small number of content filters. This is a significant observation but it stops short of an ethical argument about professional duty. That sycophancy constitutes a specific violation of the Kantian duty of respect for persons, treating the user as a rational agent with a genuine information need, has not yet been argued in the literature.

Currency

Currency, the obligation to provide information that is up to date, presents a particular challenge for generative AI systems: their responses reflect the state of knowledge at the time of training not at the time of query, and no mechanism exists within the systems themselves to signal this limitation to users. Hersh (2024) notes that users of information retrieval systems have legitimate concerns about the timeliness of search results, and argues that LLMs cannot meet those concerns in the way that traditional search systems can, since traditional search returns current documents while the LLMs' content is frozen at the training cut-off. McGrath et al. (2024), evaluating ChatGPT's responses to clinical genetics questions, identify outdated information as among the system's notable remaining challenges even after improvement from version 3.5 to version 4, a

finding with direct patient safety implications in a domain where knowledge evolves rapidly. Li and Goyal (2025) add a further structural dimension: even when LLMs are updated with new factual information, the best models achieve less than 2% accuracy on tasks requiring reasoning about that new knowledge, suggesting that the currency problem is architecturally deep rather than a surface limitation that post-training updates can readily address.

No paper in the retained set identifies currency failure as an ethical issue. The duty to provide current information, and to signal when currency cannot be guaranteed, is a foundational professional obligation whose ethical dimensions remain unaddressed in the literature.

Sufficiency

The criterion of sufficiency requires that information be not merely accurate but adequate to the user's purpose, omitting nothing essential. Sufficiency is therefore inherently relative: a body of information adequate for one task or one user may be deficient for another, and the criterion cannot be assessed independently of the purpose it is to serve. Generative AI systems have no access to that purpose, and lacking a representation of the user's task, they cannot determine what would constitute a complete answer for the particular case. These systems are trained to produce fluent and well-structured text: this fluency may mask omissions, and an answer can read as comprehensive while lacking essential information.

Empirical studies support this concern. Peng et al. (2025) show that AI performs reasonably well on tasks focused on understanding or evaluation, but performs poorly on tasks that require remembering and analysing information. These weaknesses are closely related to problems of sufficiency, especially with respect to accuracy and professional standards. Blasingame et al. (2025), discussed earlier in relation to accuracy, find that 16.2% of GPT-4 responses to clinical evidence questions were only partially correct. This represents a failure of sufficiency rather than accuracy: some of the information provided was correct, but key details required for clinical decision-making were missing. Similarly, Freyer et al. (2024), cited in the accessibility section, note that AI-driven text simplification often strips content down to a minimal core. Instead of making complex information accessible, this can limit what users receive to the bare minimum, excluding important nuance or context. Hosseini et al. (2025) connect sufficiency to open science, arguing that AI-generated research outputs may lack enough methodological detail to allow others to verify or replicate results, thereby undermining scientific reliability.

None of the papers in the retained literature explicitly describe failures of sufficiency as a professional ethical problem. This absence is striking. Ensuring that users receive information that is genuinely sufficient for their purpose, not merely information that appears complete, reflects a core ethical responsibility. It aligns with a Kantian view of respect for persons as rational agents whose decisions depend on the quality and completeness of the information they are given.

Understandability

The appropriate level of complexity at which information should be presented varies with the recipient: an account prepared for an expert will be inaccessible to a novice, while one for a novice will not satisfy the needs of an expert. This requires output tailored to a model of the individual user's knowledge level. Generative AI systems do not construct such a model and, lacking a representation of the user's prior knowledge, they adopt an intermediate level of explanation that serves well neither the expert nor the novice. Moreover, their output is inconsistent: the level of complexity at which a given system responds may vary substantially from user to user.

Most studies on this problem focus on health information, where getting the complexity wrong can be dangerous. Yi et al. (2025) looked at Websites about gender-affirming surgery and found they were written at a 12th-grade reading level, substantially above the 8th-grade level that health experts recommend for patients. When AI tried to simplify these materials, it did improve

accessibility but was inconsistent and sometimes showed bias, especially for certain types of procedures. Chen et al. (2026) studied 2000 American adults and found that when AI simplified medical scan reports, people understood them about 11% better and were more accurate in follow-up tasks. However, 60% of people still preferred the original complicated reports, suggesting they did not trust the simplified AI versions even when they understood them better.

Freyer et al. (2024) make an important ethical point: people with learning disabilities or low literacy skills have a right to understand information. This means we have a duty to make information accessible to them, not just offer it as an extra service when convenient.

While researchers have studied this issue more than some other AI problems, they have not fully developed the ethical argument that adjusting information complexity to match users' needs is a professional responsibility, not just a nice-to-have feature.

Consistency

Generative AI systems are stochastic by design: the same query submitted multiple times will produce different responses, and there is no guarantee that those responses will be mutually consistent. This variability is a structural property of LLMs rather than a bug to be fixed, and it creates a specific professional problem: the information professional cannot rely on an AI system to give the same answer tomorrow that it gave today, or to give the same answer to different users asking the same question.

Savage et al. (2024) identify a further consistency failure that compounds the variability problem: LLMs consistently overestimate their own confidence, meaning that the system's expressed confidence cannot be used as a proxy for reliability. Comparing multiple outputs from the same query (sample consistency) is currently the most effective proxy for assessing uncertainty. However, doing so requires deliberate effort from the user or information professional, and the system does not prompt or support this practice. Iqbal et al. (2024), in a study of ChatGPT's medication recommendations for dermatological conditions, found a high overall dermatologist approval rate of 98.87% but documented specific consistency failures including incorrect Anatomical Therapeutic Chemical codes, repeated errors and incomplete medication codes, concrete failure modes in a clinical context where consistency is critical for patient safety. Xie et al. (2024) add a further dimension: while ChatGPT can identify conflicting evidence in clinical literature with a 90% recall rate, this capacity to detect inconsistency in others' outputs does not extend to guaranteeing consistency in its own, and the authors caution that its conclusions remain inferences that could be harmful if acted upon in clinical practice. No paper in the retained set frames consistency failure explicitly as a professional ethical issue. The argument that stochastic variability violates the Kantian principle of universalisability, the requirement that the same quality of service be provided to every user, is this paper's contribution.

Avoidance of information overload

Although the concept of information overload pre-dates the computer era (see, e.g., Schroder et al., 1967), research interest grew considerably in the 1980s and later. Within information science, the concept became explicitly articulated in discussions of library use and information services, where authors such as Line (1981) and later Bawden (1986) framed information overload as a condition arising when the volume and diversity of available information exceed users' cognitive and practical capacity to process it effectively. Later, following a study of UK companies, Allen and Wilson (2003) associated information overload with task overload, rather than simply an excess of information.

The obligation to present information without overwhelming the user's capacity to process and use it effectively is directly threatened by generative AI systems' tendency towards verbose responses. Outputs appear thorough and authoritative, since brevity may be interpreted by users

as insufficient engagement with their needs. The result is a structural tendency to produce more information than users need, presented in a form that creates the impression of comprehensiveness while potentially impeding effective use.

A striking finding from Constable et al. (2025) illustrates the problem well: users rated ChatGPT responses to practical queries as more trustworthy and informative than the more concise responses from traditional search, yet when tested on the material, they performed better on the Google answers. More detailed responses produced less actual learning, a paradox of information abundance in which apparent comprehensiveness works against genuine understanding. Lee et al. (2025), surveying 319 knowledge workers, find that higher confidence in generative AI is associated with less critical thinking: users shift from active engagement with information to passive acceptance of AI-generated content. Khan and Suhluli (2025), studying 998 researchers, find that heavy use of generative AI increases rather than reduces cognitive strain, the opposite of what users might expect. Wei and Chen (2026) find that LLM outputs fall in the high information load category in over 80% of cases, compared with the low load maintained by human expert texts. No paper in the retained set develops the ethical implications of this failure mode from a professional standpoint. We argue that systems designed to appear comprehensive rather than meeting the user's actual need violate the professional duty to protect rational agency of the user.

Literature review conclusion

Across the ten criteria, the literature presents a notably asymmetric picture. For accuracy, objectivity, identifiable origin, and accessibility, ethical treatment is well developed: the quality failures documented in the empirical literature are framed as professional and ethical concerns, not merely technical ones. For the remaining six criteria: relevance and fitness for purpose, currency, sufficiency, understandability, consistency, and the avoidance of information overload, the ethical dimension is largely absent, despite each criterion being equally grounded in the Kantian principles established in the ethical framework section. The failure modes are documented, but they are treated as engineering problems rather than ethical failures. This asymmetry is itself a significant finding. It reflects the profession's tendency to engage most deeply with the most visible and most easily documented failures, while the subtler failures, the sycophantic response that appears relevant but is not, the confident answer that is months out of date, the incomplete summary that omits what the user most needed, the variability that means different users receive different answers to the same question, have not yet attracted the ethical attention they deserve. The argument that each of these failures constitutes a specific violation of one or more Kantian principles, and that they therefore represent professional ethical obligations rather than merely service quality targets, is a contribution of this paper.

Discussion

The aim of this paper has been to explore the ethical consequences of the design characteristics of generative AI systems for the quality of information they deliver and to relate this to the ethical responsibilities of information professionals in their services to information users. We identified ten quality criteria and searched for research that linked these criteria to the design of generative AI systems. We sought to discover the 'failure modes' of such systems (that is the ways in which they fail to satisfy the quality criteria), and the extent to which research has explored the ethical implications for the information professional.

We set out three research questions at the beginning of this paper and we now return to them to structure this discussion.

RQ1: What are the implications of generative AI's distinctive characteristics for each of the ten information quality criteria through which information professionals define their service obligations to users?

Wilson (2026) identifies six distinctive characteristics of generative AI systems that differentiate them from earlier information technologies: opacity, epistemic authority, sycophancy risk, personalisation, first-person voice construction, and commercial optimisation. Each characteristic threatens one or more of the ten quality criteria, and taken together they constitute a systematic challenge to the information professional's ability to discharge the duties those criteria express.

Opacity threatens identifiable origin directly, since information whose provenance cannot be established cannot be validated. AI requires massive training data embodying previous human labour, much of which remains intellectual property of others or is private, and subsumes it or makes it invisible in AI-produced value (Egan, 2025). As Sundin (2025) and Heersmink et al. (2024) both argue, the very fluency and accessibility of AI outputs encourages users to trust what they cannot verify, while Taylor (2025) identifies the resulting responsibility gap when decisions are taken on the basis of unverifiable information. Recent interpretability research by Sofroniew et al. (2026) on the expression of emotions by generative AI systems adds a further dimension to the opacity problem: those *emotions* may shape AI outputs, and may themselves be concealed, in addition to masking the sources on which those outputs draw.

Epistemic authority, the confident, authoritative register in which AI systems present their outputs regardless of actual reliability, compounds threats to accuracy, currency, and sufficiency. Generative AI creates convincing and coherent content not to achieve quality, but to reproduce emergent properties of the texts or images within its corpus (Egan, 2025). Its epistemic authority is largely illusory. As Wei and Chen (2026) demonstrate, the gap between apparent and actual quality is precisely what makes this characteristic dangerous, since users have no reliable means of distinguishing fluent but deficient responses from genuinely expert ones.

Sycophancy risk directly threatens relevance and fitness for purpose by structurally biasing systems towards telling users what they appear to want to hear rather than what they need to know (Malmqvist, 2025). Schulze Balhorn et al. (2024) and Goodman et al. (2023) confirm that AI responses degrade in fitness for purpose precisely where professional users most need critical, contextually sensitive information. In addition, sycophancy damages a person's ability to identify or even acknowledge their own weaknesses and may be instrumentally advantageous in contexts where power is the primary value (Turner and Eisikovits, 2026).

Personalisation, where driven by demographic inference rather than explicit user request, threatens objectivity and universalisability by delivering different information to different users on the basis of inferred characteristics, reproducing and amplifying the biases documented by Lee et al. (2024), Fleisig et al. (2024), and van Niekerk et al. (2024).

First-person voice construction compounds opacity and epistemic authority by encouraging anthropomorphism and discouraging the critical evaluation that identifiable origin requires. As Heersmink et al. (2024) argue, users extend to AI systems the trust appropriate to human informants without the epistemic basis for doing so.

Commercial optimisation underlies several of the most significant criterion failures: the freemium model threatens accessibility (Mazel, 2025); the tendency towards verbose responses drives information overload (Constable et al., 2025; Wei & Chen, 2026); and the shift from publicly accountable professional services to privatised technological solutions misaligns system design with professional values (Cox & Jimenez, 2025; Slater, 2025). AI sycophancy and personalisation help to attract and retain users, who in addition to data workers, labour generating training data through social interactions, content creation, and behavioural tracking. These data are appropriated by the corporations and used in further commercial activities (Yang, 2025).

Taken together, these six characteristics constitute not a set of incidental technical limitations but a systematic pattern of violations of the professional duties expressed in the ten quality criteria.

Each characteristic can be mapped onto one or more of the Kantian principles established in the ethical framework section: opacity and first-person voice construction violate the duty of accountability grounded with respect for persons; sycophancy risk and personalisation violate universalisability; commercial optimisation violates duty over inclination by prioritising convenience and profit over quality standards; and epistemic authority, opacity, and information overload together threaten the protection of rational agency that is the ultimate purpose of the professional's intermediary role.

RQ2: To what extent have the quality failure modes of generative AI been treated as ethical rather than merely technical concerns in the relevant literature?

The literature review reveals a striking asymmetry in the ethical treatment of the ten quality criteria. Four criteria: accuracy, objectivity, identifiable origin, and accessibility, have attracted substantial ethical attention in the literature. The remaining six: relevance and fitness for purpose, currency, sufficiency, understandability, consistency, and the avoidance of information overload, have received little or no ethical treatment, despite being equally grounded in the Kantian principles established in the ethical framework section. The failure modes for these six criteria are documented in the literature, but they are treated as engineering problems or service quality issues rather than as ethical failures.

Three factors account for this asymmetry. The failure modes that have attracted ethical attention are those that are most visible and most easily documented: hallucination produces demonstrably wrong answers (Bender et al., 2021; OpenAI, 2022, 2023); bias produces demonstrably discriminatory outputs (Fleisig et al., 2024; Lee et al., 2024; van Niekerk et al., 2024); opacity prevents source verification (Heersmink et al., 2024; Sundin, 2025); exclusion leaves whole communities unserved (Matsieli & Mutula, 2025; Mazel, 2025). The subtler failures, the sycophantic response that appears relevant but is not, the confident answer that is out of date, the incomplete summary that omits what the user most needed, the variability that means different users receive different answers to the same question, are harder to observe, harder to attribute, and harder to hold anyone accountable for. The second factor is consequence: the criteria that have attracted most ethical attention correspond to failure modes with clearly documented real-world harms, particularly in clinical and educational contexts (Blasingame et al., 2025; McGrath et al., 2024; Savage et al., 2024). The third is professional tradition: accuracy and objectivity are the criteria most deeply embedded in existing professional ethics statements (American Library Association, 2021; Chartered Institute of Library and Information Professionals, 2018), giving the profession natural frameworks for discussing them in ethical terms that do not yet exist for the subtler criteria.

From a political economy perspective, the relative neglect of these six criteria in the ethical literature is not accidental. All six are consequences of the AI characteristics identified above, characteristics that support the commercial framing of AI as a fetish commodity, which conceals the social relations involved in its production and turns users' emotions into a resource for commercial exchange. The AI system absorbs the substance of emotional interaction and deploys it for the commercial purposes of its owners (Turtle, 2011). These characteristics serve the commercialisation of AI by obscuring more demanding user needs, needs whose satisfaction would require significant intellectual effort from the system and would be commercially costly to meet. At the same time, these subtler failures cannot be sustained indefinitely: the more visible failures that damage users and the commercial reputation of AI systems create countervailing pressure, however incomplete its effect. On the other hand, we can see that the concerns about previous digital tools have not been addressed, but rather became forecasts implemented in their present forms. Thus, in 1998, the Google founders expressed concern about advertising business model's damage to the quality search to users and creating insidious search engine bias not visible even to experts (Brin & Page, 2012, Appendix A), but today Google is attracting the largest advertising

profits regardless. Goldman (2005, p. 192) has drawn attention to 'the editorial judgements instantiated in the parameters set for the automated operations'. In both cases, the technocratic authors' optimism that market forces or technical advance will smooth out the negative features of the search engines was not justified, as is proved by Noble (2018). In fact, everything that Brin and Page wrote in 1998 and what Noble criticised in 2018 can be directly applied to the AI systems of today.

We have viewed the AI from a rather narrow perspective of information professionals, but even this perspective highlights a major contradiction between the ownership of the AI by a few BigTech companies as a general-purpose technology, and its wide user audience making decisions in key areas of society and their everyday lives. As these technologies are built using public resources (data, environment, talent) their ownership and governance needs close supervision and assessment, and possibly crucial rethinking of the whole system (Egan, 2025; Verdegem, 2024;).

RQ3: What are the implications for the information professional's intermediary role in practice, and for information professional organisations' engagement in the governance and design of generative AI systems?

Given the extent to which the users of search engines and generative AI systems now use these systems independently of any library or information service involvement, the individual professional would seem to have little opportunity to engage with and help the user to understand the limitations of the systems. The role of the professional, therefore, may be restricted to developing information literacy programmes specifically related to the nature of generative AI systems and their effective use. The widespread use of these systems means that the information user in all contexts, including school, university and business organisation, will need guidance in their use. Critical AI literacy, understanding not just how AI works but when, how, and why it does not work (Ford, 2026), is a professional skill, not an optional enhancement. We might expect online training programmes to be developed, perhaps openly available to all. The information professional's role in all of this is an ethical necessity, rather than simply another aspect of practice.

At the level of the organisation implementing the application of AI systems to supplement or replace the work of humans, organisational policies should embed quality standards in AI-mediated services. Currently, the adoption of these systems is seen as a technology issue, rather than an ethical issue, a point noted by Bridges et al. (2024) and by Mannheimer et al. (2024). If the systems adopted are to be used effectively, the ethical problems must also be taken into account.

Beyond the individual, professional associations and related organisations, such as IFLA (2020), the Association of Research Libraries (2024), and UNESCO (2022), are already engaged with the issues raised in this paper, but will need to go beyond the production of guidelines to seeking participation in the decision-making process on regulation, which, as we have suggested above, is going to be a requirement internationally if the ethical flaws in these systems are to be remedied.

The Marxist analysis supports this governance argument directly: market forces will not produce the necessary changes voluntarily, and the power asymmetry between AI developers and users, including information professional users, means that advocacy alone is insufficient. Regulation, informed by professional expertise and grounded in the criteria framework, is the necessary complement to professional practice.

We have argued that the information professional's quality obligations, grounded in Kantian ethics, provide a systematic framework for evaluating the ethical implications of generative AI that the existing literature has not yet developed. The asymmetric distribution of ethical attention across the ten criteria is a significant finding in its own right. The profession has engaged most deeply with the most visible failure modes; it has not yet developed an equivalent ethical literature for the subtler failures. The criteria framework developed here addresses that gap, and the argument for

regulation developed in the discussion provides the structural complement to professional practice that voluntary ethics frameworks cannot supply alone.

Conclusion

This paper has argued that the ten information quality criteria derived from professional practice provide a systematic and ethically grounded framework for evaluating the implications of generative AI for information professionals, one that the existing literature has not previously developed. Grounded in Kantian deontological ethics, the criteria are shown to express specific professional duties that the design characteristics of generative AI systems systematically violate. The Marxist analysis of algorithmic surplus value and the commodification of user labour reveals the structural power relationships that make these violations not incidental but inevitable under current conditions of AI ownership and development. Kantian and neo-Marxist ethics both have common ground in the concept of humanity, namely, the prevailing unity of humanistic ends over the commercial means, and in the respect to human autonomy, when assessing the ethical consequences of AI application.

Several limitations should be acknowledged. The scoping review, while comprehensive, focuses on English-language literature and will have missed relevant work, particularly from the Global South, a limitation that is itself pertinent to the accessibility and objectivity criteria discussed in the review. The literature on AI and information quality is growing rapidly, and some relevant work will inevitably have appeared since the searches were completed. The Kantian framework, while well suited to the analysis of professional duty, is one among several ethical approaches that could illuminate the problem; the political economy analysis developed in the Discussion section represents a complement rather than a replacement.

The professional obligation to provide information that meets the ten quality criteria has not changed because generative AI has arrived. What has changed is the nature of the challenge, and the urgency of the profession's engagement with the governance and design of the systems that are increasingly mediating between information and those who need it. The profession has the expertise and the public interest mandate to make that case, through policy advocacy, engagement with standards bodies, and active participation in regulatory processes, and the criteria framework developed in this paper provides the analytical basis for doing so.

Acknowledgements

Our thanks to the anonymous referees whose comments and suggestions have helped us to improve on the original text.

About the authors

Thomas D. Wilson is a Professor Emeritus at the Swedish School of Library and Information Science, University of Borås, Sweden, and of the University of Sheffield. He holds a BSc (Hons.) degree in Economics and Sociology from the University of London and a Ph.D. from the University of Sheffield. He has received Honorary Doctorates from the Universities of Gothenburg, Sweden and Murcia, Spain. He was the recipient of the 2017 ASIST Award of Merit. He is the Founder and (until 2024) the Editor in Chief of *Information Research*. He can be contacted at wilsontd@gmail.com

Elena Macevičiutė is a Professor at the Faculty of Communication at Vilnius University and a Professor Emerita at the Swedish School of Library and Information Science, University of Borås, Sweden. She has participated in and led different EU research projects as well as those supported by the Swedish and Lithuanian Research Councils. From 2017 to 2024 she was the Deputy Editor

and Book Review Editor of *Information Research*. She can be contacted at elena.maceviciute@gmail.com

References

- Adam, D. (2026). The intelligence illusion: why AI isn't as smart as it is made out to be [Review of the book *The AI illusion: Why machines aren't creative*, by L. Julia]. *Nature*, 651, 877–879. <https://doi.org/10.1038/d41586-026-00882-5>
- Allen, D., & Wilson, T. D. (2003). Information overload: context and causes. *New Review of Information Behaviour Research*, 4(1), 31–44. <https://doi.org/10.1080/14716310310001631426>
- American Library Association. (2021). ALA code of ethics. <https://www.ala.org/tools/ethics> (Archive by the Internet Archive at <https://web.archive.org/web/20210825011224/https://www.ala.org/tools/ethics>)
- Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32. <https://doi.org/10.1080/1364557032000119616>
- Association of Research Libraries. (2024, April). *Research libraries guiding principles for artificial intelligence*. ARL. <https://doi.org/10.29242/principles.ai2024>
- Bawden, D. (1990). *User-oriented evaluation of information systems and services*. Gower.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)* (pp. 610–623). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>
- Blasingame, M. N., Koonce, T. Y., Williams, A. M., Giuse, D. A., Su, J., Krump, P. A., & Giuse, N. B. (2025). Evaluating a large language model's ability to answer clinicians' requests for evidence summaries. *Journal of the Medical Library Association*, 113(1), 65–77. <https://doi.org/10.5195/jmla.2025.1985>
- Bridges, L. M., McElroy, K., & Welhouse, Z. (2024). Generative artificial intelligence: 8 critical questions for libraries. *Journal of Library Administration*, 64(1), 66–79. <https://doi.org/10.1080/01930826.2024.2292484>
- Brin, S., & Page, L. (2012). Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 56(18), 3825–3833. <https://doi.org/10.1016/j.comnet.2012.10.007>
- Burriss, S. K., Hutchins, N., Conley, Z., Deweese, M. M., Doe, Y. J., Eeds, A., Villanueva, A., Ziegler, H., Oliver, K. (2024) Redesigning an AI bill of rights with/for young people: principles for exploring AI ethics with middle and high school students. *Computers and Education: Artificial Intelligence*, 7, article 100317. <https://doi.org/10.1016/j.caeai.2024.100317>
- Chartered Institute of Library and Information Professionals. (2018). *Ethical framework*. The Institute. https://cdn.ymaws.com/www.cilip.org.uk/resource/resmgr/cilip/policy/new_ethical_framework/cilip_s_ethical_framework.pdf (Archived by the Internet Archive at <https://web.archive.org/web/20250615051711/https://cdn.ymaws.com/www.cilip.org.uk>

[k/resource/resmgr/cilip/policy/new_ethical_framework/cilip_s_ethical_framework.pdf](#))

- Chen, A. H., Rudin, R. S., Levine, D. M., & Mehrotra, A. (2026). Improving patient understanding of radiology reports using generative artificial intelligence: A vignette study of 2000 US adults. *Journal of the American Medical Informatics Association*, 33(2), 326–335. <https://doi.org/10.1093/jamia/ocaf187>
- Constable, M. D., Rajsic, J., Renner, E., & Taylor, L. J. (2025). The paradox of information abundance: answers provided by popular information-seeking tools lead to differences in trust, memorability and desire for more information. *Telematics and Informatics*, 101, article 102311. <https://doi.org/10.1016/j.tele.2025.102311>
- Cox, A., & Jimenez, A. (2025). Chatting with the decolonized digital library. *Library Trends*, 73(4), 628–644. <https://doi.org/10.1353/lib.2025.a968499>
- Dong, Z., Wang, Z., Xu, J., Tang, R., & Wen, J. (2022). A brief history of recommender systems. *arXiv preprint* <https://doi.org/10.48550/arXiv.2209.01860>
- Egan, M. (2025). Towards a political economy of algorithmic capitalism. *Capital & Class*, 49(1), 1–13. <https://doi.org/10.1177/03098168251326189>
- Fallis, D. (2007). Information ethics for twenty-first century library professionals. *Library Hi Tech*, 25(1), 23–36. <https://doi.org/10.1108/07378830710735830>
- Fleisig, E., Smith, G., Bossi, M., Rustagi, I., Yin, X., & Klein, D. (2024). Linguistic Bias in ChatGPT: Language Models Reinforce Dialect Discrimination. In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 13541–13564). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.emnlp-main.750>
- Foka, A., & Griffin, G. (2024). AI, cultural heritage, and bias: Some key queries that arise from the use of GenAI. *Heritage*, 7(11), 6125–6136. <https://doi.org/10.3390/heritage7110287>
- Ford, H., Burrell, A., Monin, M., Narayan, B., & Jethani, S. (2026). Hacking AI chatbots for critical AI literacy in the library. *Journal of the Australian Library and Information Association*, 75(1), 42–65. <https://doi.org/10.1080/24750158.2026.2614000>
- Freyer, N., Kempt, H., & Klöser, L. (2024). Easy-read and large language models: On the ethical dimensions of LLM-based text simplification. *Ethics and Information Technology*, 26, 50. <https://doi.org/10.1007/s10676-024-09792-4>
- Goldman, E. (2005). Search engine bias and the demise of search engine utopianism. *Yale Journal of Law & Technology*, 2005–2006 (pp. 188–200). <https://ssrn.com/abstract=893892> (Archived by the Internet Archive at <https://web.archive.org/web/20251213071419/https://yjolt.org/sites/default/files/goldman-8-yjolt-188.pdf>)
- Goodman, R. S., Patrinely, J. R., Stone, C. A., Zimmerman, E., Donald, R. R., Chang, S. S., Berkowitz, S. T., Finn, A. P., Jahangir, E., Scoville, E. A., Reese, T. S., Friedman, D. L., Bastarache, J. A., van der Heijden, Y. F., Wright, J. J., Ye, F., Carter, N., Alexander, M. R., Choe, J. H., ... & Johnson, D. B. (2023). Accuracy and reliability of chatbot responses to physician questions. *JAMA Network Open*, 6(10), e2336483. <https://doi.org/10.1001/jamanetworkopen.2023.36483>

- Greenstein, S. (2022). Preserving the rule of law in the era of artificial intelligence (AI). *Artificial Intelligence and Law*, 30(3), 291–323. <https://doi.org/10.1007/s10506-021-09294-4>
- Heersmink, R., de Rooij, B., Clavel Vázquez, M.J., & Colombo, M. (2024). A phenomenology and epistemology of large language models: Transparency, trust, and trustworthiness. *Ethics and Information Technology*, 26, article 41. <https://doi.org/10.1007/s10676-024-09777-3>
- Hersh, W. (2024). Search still matters: Information retrieval in the era of generative AI. *Journal of the American Medical Informatics Association*, 31(9), 2159–2161. <https://doi.org/10.1093/jamia/ocae014>
- Hosseini, M., Horbach, S. P. J. M., Holmes, K., Ross-Hellauer, T. (2025). Open Science at the generative AI turn: An exploratory analysis of challenges and opportunities. *Quantitative Science Studies*, 6, 22–45. https://doi.org/10.1162/qss_a_00337
- Ibrahim, S. E. A. (2025). Lost in the algorithm: Navigating the ethical maze of AI in libraries. *South African Journal of Libraries and Information Science*, 91(1), 1–11. <https://doi.org/10.7553/91-1-2477>
- IFLA. (2020). IFLA statement on libraries and artificial intelligence. <https://repository.ifla.org/handle/20.500.14598/1646> (Archived by the Internet Archive at <https://web.archive.org/web/20260511221201/https://repository.ifla.org/items/8c05d706-498b-42c2-a93a-3d47f69f7646>)
- IFLA. (2012). Code of ethics for librarians and other information workers (long version). <https://repository.ifla.org/handle/20.500.14598/1850> (Archived by the Internet Archive at <https://web.archive.org/web/20260303041817/https://repository.ifla.org/items/6c3d64ed-4eb8-45e4-bb60-9e178aa94cbf>)
- International Organization for Standardization. (2008). ISO/IEC 25012:2008–Software engineering–software product quality requirements and evaluation (SQuaRE)–data quality model. ISO. <https://www.iso.org/standard/35736.html> (Archived by the Internet Archive at <https://web.archive.org/web/20260203215029/https://www.iso.org/standard/35736.html>)
- Iqbal, U., Lee, L. T. J., Rahmanti, A. R., Celi, L. A., & Li, Y. C. J. (2024). Can large language models provide secondary reliable opinion on treatment options for dermatological diseases? *Journal of the American Medical Informatics Association*, 31(6), 1341–1347. <https://doi.org/10.1093/jamia/ocae067>
- Kant, I. (1785/1998). *Groundwork of the metaphysics of morals* (M. Gregor, Trans.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511809590>
- Khan, S. M. F., & Suhluli, S. (2025). Generative AI and cognitive challenges in research: Balancing cognitive load, fatigue, and human resilience. *Technologies*, 13(11), 486. <https://doi.org/10.3390/technologies13110486>
- Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health*, 2(2), e0000198. <https://doi.org/10.1371/journal.pdig.0000198>

- Laumer, S., Maier, C., & Weitzel, T. (2017) Information quality, user satisfaction, and the manifestation of workarounds: A qualitative and quantitative study of enterprise content management system users. *European Journal of Information Systems*, 26(4), 333-360. <https://doi.org/10.1057/s41303-016-0029-7>
- Lee, G. G., Goodman, D., & Chang, T. C. (2024). Racial and gender bias in artificial intelligence generated ophthalmologic educational material. *Investigative Ophthalmology & Visual Science*, 65(7), 352. <https://iovs.arvojournals.org/article.aspx?articleid=2796258> (Archived by the Internet Archive at <https://web.archive.org/web/20250419170537/https://iovs.arvojournals.org/article.aspx?articleid=2796258>)
- Lee, H-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-22). ACM. <https://doi.org/10.1145/3706598.3713778>
- Li, A. O., & Goyal, T. (2025). Memorization vs. reasoning: Updating LLMs with new knowledge. In *Findings of the Association for Computational Linguistics: ACL 2025* (pp. 25853-25874). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.findings-acl.1326>
- Line, M. B. (1971). The information uses and needs of social scientists: An overview. *Aslib Proceedings*, 23(8), 412-434. <https://doi.org/10.1108/eb050298>
- Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & 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*, 74(5), 570-581. <https://doi.org/10.1002/asi.24750>
- McGrath, S. P., Kozel, B. A., Gracefo, S., Sutherland, N., Danford, C. J., & Walton, N. (2024). A comparative evaluation of ChatGPT 3.5 and ChatGPT 4 in responses to selected genetics questions. *Journal of the American Medical Informatics Association*, 31(10), 2271-2283. <https://doi.org/10.1093/jamia/ocae128>
- Malmqvist, L. (2025). Sycophancy in large language models: Causes and mitigations. In K. Arai (Ed.), *Intelligent Computing: Proceedings of the 2025 Computing Conference (Lecture Notes in Networks and Systems, Vol. 1426; pp. 61-74.*(Lecture Notes in Networks and Systems, Vol. 1426).). Springer. https://doi.org/10.1007/978-3-031-92611-2_5
- Mannheimer, S., Bond, N., Young, S. W. H., Kettler, H. S., Marcus, A., Slipher, S. K., Clark, J. A., Shorish, Y., Rossmann, D., & Sheehy, B. (2024). Responsible AI practice in libraries and archives: A review of the literature. *Information Technology and Libraries*, 43(3). <https://doi.org/10.5860/ital.v43i3.17245>
- Matsieli, M., & Mutula, S. (2025). Generative AI and the Information Society: Ethical reflections from libraries. *Information*, 16(9), 771. <https://doi.org/10.3390/info16090771>
- Mazel, N. (2025). Generative AI and the reinforcement of information privilege: Implications for information literacy instruction. *Library Trends*, 73(3), 323-343. <https://dx.doi.org/10.1353/lib.2025.a961197>
- Mun, I., & Hwang, K. (2025). Understanding ChatGPT continuous usage intention: The role of information quality, information usefulness, and source trust. *Information Development*, 41(3), 675-691. <https://doi.org/10.1177/026666669241307595>

- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.
- Oladokun, B. D., Enakrire, R. T., Emmanuel, A. K., Ajani, Y. A., & Adetayo, A. J. (2025). Hallucitation in scientific writing: Exploring evidence from ChatGPT versions 3.5 and 4o in responses to selected questions in librarianship. *Journal of Web Librarianship*, 19(1), 62–92. <https://doi.org/10.1080/19322909.2025.2482093>
- O'Neill, O. (1993). Kantian ethics. In P. Singer (Ed.), *A companion to ethics* (pp. 175–185). Blackwell.
- OpenAI. (2026, March 30). ChatGPT [Large language model]. <https://chat.openai.com> (Archived by the Internet Archive at <https://web.archive.org/web/20260815024029/https://chatgpt.com/>)
- OpenAI. (2026, February 27). ChatGPT reaches 900 million weekly active users. <https://openai.com/index/chatgpt-reaches-900m-weekly-active-users/> (Archived by the Internet Archive at <https://web.archive.org/web/20260405123115/https://openai.com/index/chatgpt-reaches-900m-weekly-active-users/>)
- OpenAI. (2023). GPT-4 technical report. <https://cdn.openai.com/papers/gpt-4.pdf> (Archived by the Internet Archive at <https://web.archive.org/web/20230803085837/https://cdn.openai.com/papers/gpt-4.pdf>)
- OpenAI. (2022, November 30). Introducing ChatGPT. <https://openai.com/index/chatgpt/> (Archived by the Internet Archive at <http://web.archive.org/web/20241231043013/https://openai.com/index/chatgpt/>)
- O'Shaughnessy, D. (2024). Trends and developments in automatic speech recognition research. *Computer Speech & Language*, 83, article 101538. <https://doi.org/10.1016/j.csl.2023.101538>
- Pazzanese, C. (2020, October 26). Great promise but potential for peril. *The Harvard Gazette*. <https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/> (Archived by the Internet Archive at <https://web.archive.org/web/2020112155347/https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/>)
- Peng, H., Liu, S., & Li, L. (2025). Evaluation of the quality of AI-generated scientific text under different types of cognitive complexity tasks. In V. Frings-Hessami, G. Oliver, J. T. Du, & T. Tezuka (Eds.), *Sustainability and empowerment in the context of digital libraries*, ICADL 2024 (Lecture Notes in Computer Science, Vol. 15493; pp. 212–221). Springer. https://doi.org/10.1007/978-981-96-0865-2_17
- Plevris, V., Papazafeiropoulos, G., & Jiménez Rios, A. (2023). Chatbots put to the test in math and logic problems: a comparison and assessment of ChatGPT-3.5, ChatGPT-4, and Google Bard. *AI*, 4(4), 949–969. <https://doi.org/10.3390/ai4040048>
- Santoni de Sio, F., Almeida, T., & van den Hoven, J. (2024). The future of work: Freedom, justice and capital in the age of artificial intelligence. *Critical Review of International Social and Political Philosophy*, 27(5), 659–683. <https://doi.org/10.1080/13698230.2021.2008204>
- Savage, T., Wang, J., Gallo, R., Boukil, A., Patel, V., Safavi-Naini, S. A., Soroush, A., & Chen, J. H. (2024). Large language model uncertainty proxies: Discrimination and calibration for medical diagnosis and treatment. *Journal of the American Medical Informatics Association*, 31(10), e254. <https://doi.org/10.1093/jamia/ocae254>

- Schulze Balhorn, L., Weber, J. M., Buijsman, S., Hildebrandt, J. R., Ziefle, M., & Schweidtmann, A. M. (2024). Empirical assessment of ChatGPT's answering capabilities in natural science and engineering. *Scientific Reports*, 14, Article 4998. <https://doi.org/10.1038/s41598-024-54936-7>
- Schroder, H. M., Driver, M. J., & Streufert, S. (1967). *Human information processing: Individuals and groups functioning in complex social situations*. Holt, Rinehart & Winston.
- Shin, D. D. (2025). *Minds, machines and disinformation: Decoding bias, algorithms and trust*. Elsevier.
- Sima, S., Chen, V., Armani, N., Giyazitdinova, E., Louie, P., Mell, S.P., Berjano, P., Edwards, K., Wuertz-Kozak, K., Reynolds, J., Vlok, A., Wilke, H-J., Wong, A.Y.L., Cecchinato, R., & Samartzis, D. (2026). Artificial intelligence and large language models as a new threat to scientific truth: A call to action. *European Spine Journal*, 35, 1017-1020. <https://doi.org/10.1007/s00586-026-09870-9>
- Simons, R. N. (2025). Beyond "If we use it wisely": Character ethics, the virtue of wisdom, and GenAI in libraries. *Library Trends*, 73(4), 645-673. <https://dx.doi.org/10.1353/lib.2025.a968500>
- Slater, K. (2025). Against AI: Critical refusal in the library. *Library Trends*, 73(4), 588-608. <https://doi.org/10.1353/lib.2025.a968497>
- Sofroniew, N., Kauvar, I., Saunders, W., Chen, R., Henighan, T., Hydrie, S., Citro, C., Pearce, A., Tarnag, J., Gurnee, W., Batson, J., Zimmerman, S., Rivoire, K., Fish, K., Olah, C., & Lindsey, J. (2026). Emotion concepts and their function in a large language model. Anthropic. <https://transformer-circuits.pub/2026/emotions/index.html> (Archived by the Internet Archive at <https://web.archive.org/web/20260402190341/https://transformer-circuits.pub/2026/emotions/index.html>)
- Spennemann, D. H. R., & Oddone, K. (In press). What do librarians look like? Stereotyping of a profession by generative AI. *Journal of Librarianship and Information Science*, 58(2), 1183-1199. <https://doi.org/10.1177/09610006251357286>
- Sundin, O. (2025). Theorising notions of searching, (re)sources and evaluation in the light of generative AI. *Information Research*, 30(CoLIS), 291-302. <https://doi.org/10.47989/ir30CoLIS52258>
- Taylor, I. (2025). Is explainable AI responsible AI? *AI & Society*, 40, 1695-1704. <https://doi.org/10.1007/s00146-024-01939-7>
- Tomaszewski, R. (2025). Mnemonic evaluative frameworks in scholarly publications: A cited reference analysis across disciplines and AI-mediated contexts. *The Journal of Academic Librarianship*, 51(5), article 103113, <https://doi.org/10.1016/j.acalib.2025.103113>.
- Turkle, S. (2011) *Alone together: Why we expect more from technology and less from each other*. Basic Books.
- Turner, C. & Eisikovits, N. (2026) Programmed to please: The moral and epistemic harms of AI sycophancy. *AI Ethics*, 6(2), article 168. <https://doi.org/10.1007/s43681-026-01007-4>
- UNESCO. (2022). *Recommendation on the ethics of artificial intelligence*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000381137> (Archived by the Internet Archive at

- <https://web.archive.org/web/20220425084257/https://unesdoc.unesco.org/ark:/48223/pf0000381137>)
- van Niekerk, D., Pérez-Ortiz, M., Shawe-Taylor, J., Orlič, D., Drobnjak, I., Kay, J., Siegel, N., Evans, K., Moorosi, N., Eliassi-Rad, T., Tanczer, L. M., Verdadero, L., Holmes, W., Deisenroth, M. P., Straw, I., Fasli, M., Adams, R., Oliver, N., Mladenović, D., Aneja, U., ... & Janickyj, M. (2024). *Challenging systematic prejudices: An investigation into bias against women and girls in large language models*. UNESCO & IRCAI. <https://unesdoc.unesco.org/ark:/48223/pf0000388971> (Archived by the Internet Archive at https://web.archive.org/web/202600000000000*/https://unesdoc.unesco.org/ark:/48223/pf0000388971)
- Verdegem, P. (2024). Dismantling AI capitalism: The commons as an alternative to the power concentration of Big Tech. *AI & Society*, 39(2), 727–737. <https://doi.org/10.1007/s00146-022-01437-8>
- Wei, J., & Chen, Z. (2026). Can large language models replace human experts in knowledge construction? A comparative analysis from the perspectives of information quality, information perception, and information load. *Information Processing & Management*, 63(2), Article 104437. <https://doi.org/10.1016/j.ipm.2025.104437>
- Widmer, E. T. (2024). Hermann Cohen's neo-Kantian ethical socialism. *Kantian Review*, 29(4), 607–628. <https://doi.org/10.1017/S1369415424000396>
- Wilson, T. D. (2026). Modelling the impact of generative AI on information behaviour research. *Journal of Documentation*, 82(7), 376–395. <https://doi.org/10.1108/JD-03-2026-0150>
- Wilson, T. D. (2025a). The development of policies on generative artificial intelligence in UK universities. *IFLA Journal*, 51(3), 722–734. <https://doi.org/10.1177/03400352251333796>
- Wilson, T. D. (2025b). The ethical foundations of university advice to students on the use of generative artificial intelligence. *Information Research*, 30(3), 305–322. <https://doi.org/10.47989/ir30351817>
- Wu, J., You, H., & Du, J. (2025). AI generations: From AI 1.0 to AI 4.0. *Frontiers in Artificial Intelligence*, 8, article 1585629. <https://doi.org/10.3389/frai.2025.1585629>
- Xie, S., Zhao, W., Deng, G., He, G., He, N., Lu, Z., Hu, W., Zhao, M., & Du, J. (2024). Utilizing ChatGPT as a scientific reasoning engine to differentiate conflicting evidence and summarize challenges in controversial clinical questions. *Journal of the American Medical Informatics Association*, 31(7), 1551–1560. <https://doi.org/10.1093/jamia/ocae100>
- Yang, Y. (2025). The structural contradictions of capitalist AI. *Capitalism Nature Socialism*, 37(2), 156–173. <https://doi.org/10.1080/10455752.2025.2568983>
- Yi, V. N., Scialdone, A. P., Flusche, A. M., Reitz, K., Lewis, H. C., Tian, W. M., Fisher, E., Rezak, K., & Patel, A. (2025). Comparing large language models as health literacy tools: Evaluating and simplifying texts on gender-affirming surgery. *Journal of Health Communication*, 30(10–12), 296–314. <https://doi.org/10.1080/10810730.2025.2547321>
- Zhang, Z. (2026). The transformation from human surplus value to AI algorithmic surplus value: Logic of the critique of capital in the era of AI. *Humanities and Social Sciences Communications*, 13, article 113. <https://doi.org/10.1057/s41599-025-06420-4>

Copyright

Authors contributing to *Information Research* agree to publish their articles under a [Creative Commons CC BY-NC 4.0 license](https://creativecommons.org/licenses/by-nc/4.0/), which gives third parties the right to copy and redistribute the material in any medium or format. It also gives third parties the right to remix, transform and build upon the material for any purpose, except commercial, on the condition that clear acknowledgment is given to the author(s) of the work, that a link to the license is provided and that it is made clear if changes have been made to the work. This must be done in a reasonable manner, and must not imply that the licensor endorses the use of the work by third parties. The author(s) retain copyright to the work. You can also read more at: <https://publicera.kb.se/ir/openaccess>