On the Question of Authorship in Large Language Models

Carlin Soos* and Levon Haroutunian**

* UCLA - Department of Information Studies, Los Angeles, California - United States
** University of Washington – Department of Linguistics, Seattle, Washington - United States
* carlinsoos@gmail.com, **levon.haroutunian@gmail.com

Carlin Soos is a knowledge organization researcher interested in classification theory, domain analysis, and category development. Carlin received his PhD in Information Studies from UCLA in 2023; his dissertation explored how tacit knowledge, heuristic reasoning, and the formation of perceptual categories influences knowledge organization practices and social classification. He now supports academic programming and course management as part of the Office of the Provost at Monclair State University.

Levon Haroutunian is a Natural Language Processing engineer. They hold a Master's degree in Computational Linguistics from the University of Washington and a Bachelor's degree in Linguistics from UCLA. Their expertise lies primarily in semantic parsing, data-to-text generation, large language models, and the ethics of AI.

Soos, Carlin and Levon Haroutunian. 2024. "On the Question of Authorship in Large Language Models". *Knowledge Organization* 51 (2): 83-95, 54 references. DOI:10.5771/0943-7444-2024-2-83.

Abstract: Adoption of pre-trained large language models (LLMs) across an increasingly diverse range of tasks and domains poses significant problems for authorial attribution and other basic knowledge organization practices. Utilizing methods from value-sensitive design, this paper examines the theoretical, practical, and ethical issues introduced by LLMs and describes how their use challenges the supposedly firm boundaries separating specific works and creators. Focusing on the implications of LLM usage for higher education, we use hypothetical value scenarios and stakeholder analysis to weigh the pedagogical risks and benefits of LLM usage, assessing the consequences of their use on and beyond college campuses. While acknowledging the unique challenges presented by this emerging educational trend, we ultimately argue that the issues associated with these novel tools are indicative of preexisting limitations within standard entity-relationship models, not wholly new issues ushered in by the advent of a relatively young technology. We contend that LLM-generated texts largely exacerbate, rather than invent from scratch, the preexisting faults that have frequently posed problems to those seeking to determine, ascribe, and regulate authorship attributions. As the growing popularity of generative AI raises concerns about plagiarism, academic integrity, and intellectual property, we advocate for a reevaluation of reductive work-creator associations and encourage the adoption of more expansive authorial concepts.

Received: 02 October 2023; Revised: 17 January 2024; Accepted: 31 January 2024

Keywords: large language models; authorship; natural language processing; ChatGPT

† This article was selected as one of two best papers at the Ninth North American Symposium on Knowledge Organization, June 15-16, 2023, held online, organized by the ISKO Chapter for Canada and the United States.

1. Introduction

Problems concerning authorial attribution have long been at the heart of knowledge organization (KO) practice. Authority records, author sets, and authorized access points have been developed to standardize how works associated with particular individuals, groups, communities, and institutions are collated, named, and displayed, and these various approaches have fruitfully aided collection management, data curation, and information retrieval in a multitude of different settings. Yet KO scholars and practitioners have long warned of the conceptual and practical limitations associated with these industry-standard approaches, noting that the intellectual boundaries separating particular works and authors are not as clear-cut as title pages and bylines might suggest. Networks of influence, professional collaboration, cultural exchange, and interpersonal support generally complicate our ability to conclusively declare which ideas belong to whom, a reality often flatted in reasonable attempts to solve consequential information problems.

Nonetheless, these underlying limitations persist, issues that have been further exacerbated by the use of pre-trained large language models (LLMs) in many creative domains. Svenonius (2009) notes that authorship has become increasingly "diffused" since the time Charles Cutter published his *Rules for a Printed Dictionary Catalogue* (1875), a descriptor that accurately characterizes how LLMs are currently developed, accessed, and operated. The streamlined integration of these models into chatbots, writing tools, and search engines has further maximized usability while simultaneously obscuring the immense amount of labor and resources behind the technology, but these underlying rhizomatic qualities inevitably become an issue when questions of authorship are brought to the fore. In this paper, we will explore how this situation affects plagiarism allegations, focusing specifically on concerns raised in the field of higher education.

1.1. Context and Background

OpenAI's release of ChatGPT in November 2022 triggered near-immediate concerns across college campuses. Perhaps still on guard from a reported rise in cheating attributed to the remote-instruction phase of the COVID-19 pandemic (Jenkins et al. 2022; Dey 2022), administrators and faculty quickly began speculating about widespread chatbot misuse. An abundance of largely hyperbolic news coverage questioning ChatGPT's ability to "replace humans" (Lock 2022) no doubt exacerbated these anxieties, leading to concerns about "radical consequences for teaching and learning" (Dolan 2023). As is common with these types of technological innovations, the panic subsided almost as quickly as it emerged, leaving in its wake a lingering malaise and ambivalence. Although concerns persist about the use of generative AI for cheating, university talking points now strike a balance of offensive disciplinary policies and practical recommendations for productive classroom integration (e.g. University of Washington 2023; UCLA 2023; University of Wisconsin-Madison 2023). As the pedagogical value of ChatGPT and its competitors continues to be explored (Kasneci et al. 2023), educators are finding new ways to utilize these sophisticated language models without compromising the integrity of their teaching.

While some schools are attempting to outright ban all applications of generative AI, others view its use by students, staff, and faculty as inevitable; user-friendly interfaces can make ChatGPT-like tools too tempting to resist, and this allure is only heightened by the social, professional, and economic pressures looming over learners and teachers alike. Complicating matters further is the embedding of these models into preexisting information structures, such as the utilization of ChatGPT by Khan Academy and Quizlet (OpenAI n.d.) or Anthropic's partnerships with Slack and Zoom (Anthropic n.d.). As the line between "chatbot" and "research tool" is further blurred, determining where cheating starts and stops becomes increasingly more difficult. Viewed from this perspective, initial warnings about how ChatGPT will "upend longstanding concepts of plagiarism, authorship, ownership, and learning" are not entirely unfounded (McCarthy 2023). However, with these new challenges comes an opportunity to revisit institutional norms, question our preexisting assumptions about their conceptual validity, and advocate for updated values that match the current moment.

1.2. Paper Goals and Structure

This paper seeks to address the implications of LLMs for KO authorial attribution and student plagiarism claims. Following a review of our methods and theoretical framework, we provide a technical summary of LLMs and introduce relevant literature from the field of natural language processing (NLP). Next, we review theories of authorship within KO, focusing primarily on Soos and Leazer's concept of the "author-as-node" (2020). Building upon this network theory, we proceed to discuss the various authorship-related issues introduced by generative AI, describing through a hypothetical value scenario (Friedman et al. 2017) how the nature of pretrained models—as well as their creative outputs—complicate the supposedly firm boundaries separating specific works and creators. With these considerations in mind, we expand the author-as-node framework using the concept of "communicative intent" (Bender et al. 2021). To conclude, we reiterate and reaffirm previously acknowledged concerns about "the author" as a distinct categorical entity while maintaining the importance of idea attribution for personal development and community accountability.

2. Methodology and Theoretical Framework

The content generated by LLMs relies upon actions and inputs from a multitude of sources, a collection of stakeholders including, but not limited to, the end user who inputs a query; the engineers, programmers, and researchers who build the model; the designers who develop the front-end interface; the organizations, institutions, and companies funding the project; the people who synthesize and publish the training data; and the innumerable individuals responsible for the content and information represented in those immense datasets. Operating within complex rhizomatic assemblages (Deleuze and Guattari 1987), each of these stakeholders is inextricably influenced by an untold number of professional, social, technical, and cultural factors that cannot be entirely understood or documented. Attempting to note all of these factors will inevitably miss something, and any such list can be much longer or shorter depending on

where and at what level one draws the system's boundaries. For example, if attribution is given to developers and programmers of a certain LLM, should attribution also be given to the developers and programmers behind the base open-source code those people likely built from? Determining where the intellectual and technical labor responsible for an LLM starts and ends is deceptively difficult, a dilemma that is only further exacerbated as proprietary models become embedded within third-party programs and interfaces. As LLMs are used by tech companies to further develop and refine their products, it is becoming increasingly impossible to avoid these models entirely.

Approaching our work through this kind of assemblage thinking, this paper theorizes about the implications of LLM authorship on KO. Focusing on notions of plagiarism and academic integrity, we ultimately question how textual works generated entirely or in part by large language models differ from those created through more traditional publishing means. Utilizing goals and methods from value-sensitive design (Friedman and Kahn 2007; Winkler and Spiekermann 2018)—notably stakeholder analysis and value scenarios (Friedman et al. 2017)—we examine the primary technical components of LLMs and analyze key life-cycle phases of its generated content. In doing so, we do not attempt to offer a complete picture of the harms and benefits of LLM usage, nor do we provide an exhaustive list of the victims and benefactors attached to any particular enterprise. Rather, motivated by social discourse surrounding ChatGPT's place on college campuses, we offer one primary value scenario of how a student might use the program to complete a writing assignment; we then weigh the risks and benefits of their action, focusing primarily on concerns related to plagiarism and academic dishonesty. Throughout this work, we aim to offer a balanced perspective on the issue that acknowledges the valid pedagogical concerns raised by this technology while also emphasizing its similarity to other situations previously described in the KO literature. Given the abundance of thoughtful KO scholarship focused on the problems of authorship, we believe the discipline is uniquely positioned to address this timely issue.

3. Technical Overview of LLMs for KO

Language modeling is the task of computationally representing how humans use language. In practice, language models typically predict and generate a sequence of words given another sequence as context. These models are a useful component of nearly every kind of NLP system, from automatic speech recognition to machine translation to natural language generation.

Language models based on neural networks are far and away the most common types used today. The simplest type of neural language model is a feed-forward neural network (Bengio et al. 2003), which is composed of a number of layers containing sets of units typically referred to as "neurons." The first layer is an embedding layer, which converts the individual words of an input into vectors of numeric values. Every unit—or neuron—of this embedding vector is connected to every neuron in the next layer through a weight and a bias value. The first stage of computation applies those weights and biases to the initial vector; a nonlinear function (such as a sigmoid function) is then applied to the initial vector to determine the values of each neuron in the second layer. The neurons in the second layer are similarly connected to the neurons of the third layer, and so on. More complex types of neural models incorporate different types of connections between neurons, which are necessary to account for the sequential nature of text data. Weights and biases are generally referred to as "parameters," and a model's size is usually described by its number of parameters.

During training,[1] neural language models are optimized on token prediction tasks, where they must predict output text based on an input. Input text is first split into a sequence of tokens, which are typically words or sub-word pieces. These tokens are then mapped to corresponding vectors, which are then run through the matrices that comprise the model's parameters. The output of this computation is another sequence of vectors. This sequence can then be compared against the expected output, which is tokenized and mapped in the same manner. The result of this comparison is a loss score, which determines the degree to which the generated output differs from the expected output. Using the Chain Rule from multivariable calculus, the training routine updates the model's parameters to reduce the loss score; in other words, the parameters are modified to increase the similarity between the actual and expected outputs. This process repeats for every input/output pair in the training data. Typically, training concludes after many full passes (called "epochs") over the training data.

Neural language model training results in what is referred to as a parametric memory: language models distill and "memorize" their training data in their parameters to produce output that aligns with the observed data patterns. This parametric memory is the sum total of the "knowledge" that a language model has. After training, a purely generative language model has no access to its training data and cannot access any additional external information.

Once trained, an LLM produces new text by ingesting an input: splitting given text into a sequence of tokens, mapping those tokens into corresponding vectors, and running those vectors through its parameters. The model produces output by iteratively predicting the vector of the next token in the sequence. In other words, it predicts the most likely continuation of the input, based on the information stored in its parametric memory.

2.1. Scaling Up: The Birth of Large Language Models

In 2017, the advent of a specialized type of neural network, called a Transformer, gave rise to a new era in language modeling (Vaswani et al.). One of the first examples of a Large Language Model is BERT, a Transformer-based language model that advanced the state of the art on many common NLP benchmark tasks (Devlin et al. 2019).

The shift to Transformer-based language models marked an increase in both the size of models and the data used to train them. BERT has approximately 110 million parameters—which is relatively massive compared to its contemporaries—and a similarly large training corpus: English Wikipedia, which included 2.5 billion words in the version the authors used; and BookCorpus (Zhu et al. 2015), which contained around 800 million words pulled from approximately 11,000 unpublished books scraped from Smashwords. Following BERT was a flood of pre-trained language models, with notable examples including ERNIE (Zhang et al. 2019), GPT-2 (Radford et al. 2019), XLNet (Yang et al. 2019), BART (Lewis et al. 2020), T5 (Raffel et al. 2020), and GPT-3 (Brown et al. 2020).

The creation of GPT-3 in 2020 marked the apex of increases to model size; it has a whopping 175 billion parameters, almost 1,600 times larger than BERT. GPT-3's gigantic scale came along with, of course, a gigantic training set, containing English Wikipedia and BookCorpus along with the CommonCrawl dataset, which is a web crawl dataset consisting of the text from billions of web pages. Like BERT before it, GPT-3 showed impressive performance gains on a variety of NLP tasks.

GPT-3 also marked the beginning of a new LLM paradigm. Previous LLMs were usually not directly applied to specific tasks of interest. Instead, researchers would download a pre-trained language model like BERT and train its parameters further on a smaller set of task-specific data—a process known as fine-tuning. Because GPT-3 was released closed-source, its users could not simply download the model and train it further. However, GPT-3 achieved impressive performance *without* being fine-tuned, through a method called in-context learning (ICL; Brown et al. 2020). To apply ICL, a user supplies a "prompt" to the model that includes a handful of in-context demonstrations (e.g. a few examples of English sentences paired with their French translations) along with their input to the model (e.g. a new English sentence), and the model is expected to produce output in the format represented by the demonstrations (e.g. the French translation of the input). In this way, GPT-3 functions as a general-purpose LLM: it is intended to be used on a wide variety of tasks, with no need (or option) for customization.

2.2. Data

For the reasons described above, massive corpora are necessary to create large language models. Neural language models get their power from their parametric memory, and their parametric memory comes from the data the models ingest during training. Unfortunately, the sheer size of these corpora means that researchers who create or use them cannot be fully aware of what they contain (Paullada et al. 2020). The opacity of many of these large datasets is due to what Bender et al. (2021) call "documentation debt," which is "a situation where the datasets are both undocumented and too large to document post hoc" (615). Numerous audits of large machine learning datasets have found that they contain non-trivial amounts of unwanted content (Dodge et al. 2021), copyright violations (Bandy and Vincent 2021), sexually explicit material (Birhane et al. 2021), and hate speech (Gehman et al. 2020). For a more detailed critique of practices surrounding the collection and use of machine learning datasets, see Paullada et al. (2020).

CommonCrawl exemplifies the problem of documentation debt. The dataset is an effort by The Common Crawl Foundation to "[democratize] access to web information by producing and maintaining an open repository of web crawl data" (Common Crawl n.d.). As of April 2023, CommonCrawl contains 3.1 billion web pages (Nagel 2023). An analysis by Luccioni and Viviano (2021) found that around 5% of the web pages included in CommonCrawl contain hate speech and slurs. There have been many efforts to filter CommonCrawl (most notably C4; Raffel et al. 2020), including by Brown et al. (2020) during their creation of GPT-3. However, it is virtually impossible to comprehensively filter or audit a dataset of this scale. Additionally, as Bender et al. (2021) point out, the nature of Internet data means that datasets like CommonCrawl necessarily overrepresent the voices of young, male Internet users in developed countries at the expense of other cultures, worldviews, and experiences.

With language models as large and complex as GPT-3, it can be difficult to conceptualize the links between the data it was trained on and the output it produces. However, an understanding of the training data used to train an LLM should be in the foreground of any attempt to determine the authorship of its output.

2.3. Where we are now: ChatGPT

Most of OpenAI's current state-of-the-art models are direct descendants of GPT-3[2]—or, more specifically, of InstructGPT (Ouyang et al. 2022). What differentiates InstructGPT from the initial version of GPT-3 is mainly two new phases of training called instruction tuning and reinforcement learning from human feedback (RLHF).

Instruction tuning is an extension of pre-training in which the model is trained on a dataset containing pairs of instructions (such as writing prompts or math problems) and answers. The motivation behind this training is to align the model with its intended downstream application: users will input a prompt and define a target output format with the expectation that the model will generate a response conforming to their specifications (Ouyang et al. 2022). Still according to the authors, to train InstructGPT, OpenAI collected a set of instructions and answers from human labelers, including users of GPT-3 and paid contractors. The resulting dataset has not been released. RLHF follows both pre-training and instruction tuning, and this phase may even continue once the model is deployed (as is the case with ChatGPT (OpenAI 2022). During this phase of training, human judges are presented with multiple model outputs for the same prompt and asked to rank them in order of quality (OpenAI 2022). Once successfully trained on this feedback, the model is more likely to produce output similar to the higher-rated examples.

OpenAI has stated that its motivation for using RLHF is to "make artificial general intelligence (AGI) aligned with human values and follow human intent" (Leike et al. 2022). In practice, Ouyang et al. (2022, 10) accomplish this by "[having] labelers evaluate whether an output is inappropriate in the context of a customer assistant, denigrates a protected class, or contains sexual or violent content".

2.4. Human Language Production and LLM Text Generation

ChatGPT and its ilk are undoubtedly impressive technological feats. After a brief interaction with OpenAI's chatbot, many users are surprised by its apparent mastery of the English language. However, Bender et al. (2021) provide a cautionary reminder for interpreting LLM-generated text: "coherence [is] in the eye of the beholder" (616). LLMs might *appear* to understand human language and produce meaningful output in response, but that meaning is actually created by their human interlocutors, not the LLMs themselves (Bender and Koller 2020).

Human communication "takes place between individuals who share common ground and are mutually aware of that sharing (and its extent), who have communicative intents which they use language to convey, and who model each others' mental states as they communicate" (Bender et al. 2021, 616). Language models, having no experience of the world beyond the tokens in their training data, do not share common ground with their human users and lack both communicative intents and mental states. When a person reads text generated by an LLM, it may seem as though there is thought or affect behind the response. This is not the case, and the illusion of communication comes from our own human linguistic capabilities: "our perception of natural language text, regardless of how it was generated, is mediated by our own linguistic competence and our predisposition to interpret communicative acts as conveying coherent meaning and intent, whether or not they do" (Bender et al. 2021, 616).

4. Works and Authorship Theory in KO

LLMs process trillions of forms and learn to recognize statistically significant patterns in their usage, but this is not the same thing as understanding their meaning (Saussure 1959). Just as a copy of *Wuthering Heights* is a representation of Emily Brontë's work and not the work itself, the forms used to pre-train an LLM are not intrinsically meaningful. This distinction affects how we describe the functionality of generative language models and impacts how we create, store, and access documentation through a knowledge organization system (KOS).

4.1. The Conceptual Structure of a Work

Smiraglia explains that "works are core narratives in every part of human experience—from sacred texts to legal foundations to iconic structures to iconic novels" (2019, 311). While we tend to engage with these "mentefacts" (Gnoli 2018), or mental constructs, through physical artifacts, "a work is abstract at every level, from its creator's conception of it, to its reception and inherence by its consumers" (Smiraglia 2019, 310). From an information retrieval perspective, these conceptual problems are typically circumvented by forming records around specific items. Hypothetically speaking, identifying the title of a bibliographic object is a straightforward activity; a brief glimpse at a book cover or title page is usually enough to accomplish the task. From there, assigning the author should be similarly easy.

While nice in theory, there are at least two factors that complicate the description of linear author-work association.

1. Different manifestations of a particular work can exhibit significant deviations from the original expression.
2. Since works are abstract concepts, determining the boundary where one ends and another begins is fundamentally a matter of perception.

To the first point, take *Wuthering Heights*. According to *Resource Description and Access* (RDA) guidelines, all versions of the novel are to be collocated under the same nominal authorized access point (AAP) associated with the original manuscript: Emily Brontë. Editions published in 1848 and 1948 will likely have different covers and exhibit cosmetic editorial differences, but, by and large, few would

deny both are versions of the same work. But how much can a particular representation be changed before it is no longer *Wuthering Heights*? For example, under the entity-relationship model at the core of the *Functional Requirements for Bibliographic Records* (FRBR), translations of a work should be primarily associated with the original author. This means a Hebrew edition of the text will be attributed to Brontë even though, at the time she penned her novel, the language had yet to be revitalized for general use.

On the one hand, this Hebrew translation will hopefully preserve the abstract work concept intended by Brontë; as such, her creative labor deserves recognition. On the other hand, using her name as the primary AAP "inevitably devalues the role of the translator and ignores the creative license and labor required in the translation process" (Soos and Leazer 2020, 486). Translating is not a one-to-one process in which individual words are simply swapped for identical ones of another language. A talented translator will exhibit fluency in the source and target languages, possess a deep knowledge of the particular work, and utilize various linguistic tools to articulate its essence. So, while the goal is to maintain both the semantic and affective qualities evoked by Brontë, a translator's unique choices can severely alter a reader's experience.

To the second point, as abstract concepts, works are subject to the same factors that impact all perceptive activities. While writing her book, Brontë both purposefully and implicitly built on the things she had previously read, the people she knew, and the social context in which she lived to create something new. When a reader engages with *Wuthering Heights,* their understanding of her work is influenced by similarly personal factors—and, having now interacted with the novel, it is difficult to know how her ideas might impact their own creative production. In some kind of authorial butterfly effect, if Jim Steinman had never read *Wuthering Heights,* he might not have been inspired to compose "It's All Coming Back to Me Now" (popularized by Celine Dion). Still, regardless of the fact that he has explicitly cited Brontë's story as the primary inspiration behind the song, the two are unanimously viewed as distinct works.

Within the FRBR model, the concept of a "super work" (Svenonius 2009, 38) seeks to situate derivative works, like Steinman's, as "ideational nodes within the set" (Smiraglia 2019, 313). An influential entity like *Wuthering Heights* can be viewed as a primary connective node within an instantiation (Smiraglia 2007, 182) or textual identity network (Leazer and Furner 1999), but this core progenitor is intentionally positioned adjacent to, rather than fused with, the works it inspired. Yet even within these more robust webs of relationships, there still exists the problem of determining where one thing ends and another begins. Smiraglia arguably resolved this conceptual issue when he defined a work as "a deliberately created informing entity intended for communication" (2019, 308), with "deliberately" being the key term. This prioritization of a creator's intentionality is supported in other disciplines, where artistic genres like the readymade and the parody use creative ideation and motivation to distinguish influence from theft.

4.2. Influence and Intention

Quests for individuality and authenticity can be equally liberatory as they are oppressive. While there is undeniable value in personal expression, pressures that tie a person's economic, professional, or social worth to the originality of their creative output force them to view their peers as competition rather than collaborators. In *The Anxiety of Influence*, Bloom argues that writers are both limited and motivated by this desire to distinguish themselves from their predecessors.

> For the poet is condemned to learn his profoundest yearnings through an awareness of other selves. The poem is within him, yet he experiences the shame and splendor of being found by poems—great poems—outside him. To lose freedom in this center is never to forgive, and to learn the dread of threatened autonomy forever (Bloom 1997, 26).

In an act of *kenosis,* the author seeks "discontinuity with the precursor" (14), a response that paradoxically concedes power to the other's influence. Moving away from something is as much a response as moving towards, and in rejecting the progenitor work a writer simply reaffirms their place within the creative continuum.

Although Bloom constructed his theory around poetic networks, the anxiety of influence transcends genre and medium to gesture towards a broader humanistic desire for self-actualization. While this tendency is not inherently bad, the judgment of a work based on its intellectual purity sets a standard of originality almost impossible to achieve. Authors think and create surrounded by the works of others, not within sterile incubators free from outside influence. So when the ultimate test of intellectual autonomy rests upon someone's ability to produce innovative work—poetic or otherwise—completely removed from that of others, anxiety is a reasonable response to an unachievable expectation.

Building from Bloom and Foucault (1977), Soos and Leazer suggest that the author "as a lone and entirely detached figure simply does not exist," arguing instead that "the complex nature of intellectual and creative production makes it impossible to draw a clear and distinct boundary around a particular work and attribute it to one unique individual" (2020, 487). Rather than viewing authors as "owners" of an idea, they suggest that an "author-as-node"

approach better preserves the inherently collaborative nature of creative production. Just as work-based instantiation networks connect individual items through a unifying progenitor node, this model positions an author as a singular entity within a sea of influential relationships.

That being said, even Bloom rejected the claim that "no one ever had or ever will have a self of his or her own" as nothing more than an "unamiable fiction" (1997, xlvi). Yes, works are created within complex intellectual ecosystems, but, as individuals, the people that produce them have unique perspectives and talents worthy of recognition. To borrow Smiraglia's word, they have intentionality.

Any KO theory of authorship inevitably reaches a seemingly contradictory impasse: people are unique individuals with unique ideas and unique intentions—and, at the same time, they are complicated stimuli sponges soaking in the world around them. Authors are influenced by those who came before them, the people who inspire them, and the communities that care for them, but each offers an essential quality that only they can supply. While nuanced discourse can simultaneously hold the importance of relationality and individuality (Littletree et al. 2020), notions of authorship conveyed through standard ontological frameworks generally fail to capture this duality. FRBR extends authorship beyond individual persons to include families and corporate bodies, and the replacement of "author" with "contributor" in RDA perhaps better gestures to the expansive nature of work creation. However, use of standardized AAPs in author attribution still detaches a person, family, or corporate body from their broader context. In doing so, we are essentially suggesting that influence is secondary to the intention it inspires.

Although epistemically valuable, these influence networks are often too complex and messy to visually represent through a basic KOS. At the end of the day, a student probably just wants to find *Wuthering Heights* in the university stacks and finish their assignment, and they will likely do so by searching for "Emily Brontë," not "Jim Steinman." Authorial networks might help the user contextualize Brontë's work, but this is not typically the primary goal of most collection catalogs. Yet while presenting authors as "owners" of a work is a reasonable choice given user-warranted practices, doing so defends particular ontological commitments that hide the social, cultural, economic, and professional "complexities that affect the production of new objects and ideas" (Soos and Leazer 2020, 486). The consequences of these decisions extend far beyond any one user's search query.

5. The Authorship of LLM Content

Most universities have some kind of academic integrity policy. Cheating and other forms of intellectual dishonesty are of primary concern, with plagiarism being one of the most vehemently condemned. Learning to find, interpret, and cite sources are core skills needed for academic success, and plagiarism—a spectrum of actions that ranges from an uncited paraphrase to the wholesale appropriation of another student's writing—is largely viewed as antithetical to the ethos of the academy.

Plagiarism occurs when "somebody presents the work of others (data, words or theories) as if they were his/her own and without proper acknowledgment" (Wager and Kleinert 2012, 167). Under the authorship concepts defended by RDA and FRBR, avoiding accusations of plagiarism appears to be a straightforward task: you only need to indicate when you are referring to another person's work and never suggest their ideas are your own. Simple enough. We can debate the conceptual boundaries of works and authors, but, using the attribution protocols generally accepted across higher education, plagiarism is most often framed as an entirely avoidable issue[3].

The broader adoption of generative AI has revealed the limitations of this approach. Following the relatively quick adoption of ChatGPT by students and staff, many institutions formally declared the use of pre-trained language models to produce or enhance one's work to be a violation of academic integrity. Based on the above definition, asking ChatGPT to write your *Wuthering Heights* essay seems to be a clear-cut case of plagiarism; the student did not produce the content and is presenting it "as if they were his/her own and without proper acknowledgment." But who—or what—is being plagiarized here?

5.1. Communicative Intent and Work Creation

OpenAI has done a wonderful job of developing an application that appears to possess so-called "general intelligence." But, as previously noted, while ChatGPT's "human-like" responses can be quite convincing, the chatbot does not understand what it is saying, at least not in the typical sense in which people use those words. It also does not answer user queries in an intentional act of communication—again, at least not in the way implied by such a claim.

This lack of "communicative intent" (Bender et al. 2021) marks the fundamental distinction between the way humans and LLMs utilize language. Within the context of Smiraglia's definition, an inability to experience or express intentionality essentially disqualifies ChatGPT from being able to produce a work. So, although a language model is capable of producing information, it cannot produce a work. Absent a work, there is no victim of plagiarism.

5.2. User Queries and Feedback

But although the model itself may be incapable of intentional action, there are myriad other associated parties who are. The most obvious is the accused student.

For all intents and purposes, there is nothing technically preventing this person from being named the creator of the *Wuthering Heights* essay. Entering a query into ChatGPT, copying the text into a new document, adding their name, and submitting the file are all intentional acts focused on recording and expressing a particular viewpoint. Sure, the student did not fabricate a majority of the text, but the essay was deliberately created using their actions, knowledge, and capabilities. As such, the onus arguably rests with them.

5.3. Training Data

This appears to be a victimless crime until one considers the broader context. The plethora of data used to train an LLM directly supports the parameters it uses to generate new responses. ChatGPT may be incapable of "understanding," but the millions of authors responsible for its immense training set probably are. Although they did not personally write the exact words used in this exact essay, the collective can be viewed as a "family or corporate body" responsible for this immense network of data. Following this logic, one could argue that the generated essay paraphrases this corpus of material, making the members of this family/corporate body targets of plagiarism.

5.4. Model Creation

Well, it's an answer. But, as Dehouche argues, an accusation of plagiarism "appears rather inadequate when the 'others' in question consist in an astronomical number of authors, whose work was combined and reformulated in unique ways" (2021, 21). While those individuals intentionally created the material that was used to train ChatGPT, and while they offer a wonderful metaphor for describing how textual identity networks function, OpenAI's staff was actually the one that developed the GPT model that made the *Wuthering Heights* essay possible.

In an interesting turn, OpenAI can now either be viewed as the victim of plagiarism (by the student) or its perpetrator (towards the dataset family/corporate body). Both the code used to create ChatGPT and the parametric memory defined during its training are proprietary works intentionally created by those at OpenAI.[4] As the student failed to cite either, that can be viewed as an act of plagiarism. If the code and memory are prioritized, plagiarism accusations could theoretically be avoided by simply citing either the chatbot or its makers as a source. (Determining how to grade such research, however, essentially leads to the same problem.) At the same time, ChatGPT's parametric memory was constructed from billions of other works that cannot be cited. Whether that memory constitutes a work on its own, or whether it is simply an extension of the works aggregated in its training data, is another matter entirely.

A summary of the pros and cons related to various authorship ascriptions for LLM-generated content is presented in Table 1.

6. Plagiarism Revisited

All creative acts are forms of collaboration. New ideas and works develop within a broader social context that directly and indirectly contributes to their production, and any single author is but one node in a vast network of influence. The ambiguous boundaries between specific authors and works are further eroded by the innately diffused nature of LLMs.

Our failure to accommodate this generative content within preexisting notions of plagiarism reveals the conceptual limitations of an author-as-owner approach and highlights the importance of networked attribution. "Plagiarism" is a semantic category that allows for varying degrees of membership. Its prototypical examples—for example, paying another person to write your English paper—support the existence of linear work-author relationships and reaffirm the validity of the class. However, the "internal structure" of this category (Rosch 1975) is much more stratified than standard usage of the term suggests. We suggest that the ambiguous nature of LLM-generated works just presents a more obvious challenge to the seemingly stable concept.

While the subject is fodder for an interesting philosophical discussion, we think debating whether ChatGPT's *Wuthering Heights* essay is an example of plagiarism—or a component of a bigger plagiarism racket—largely avoids and obscures the more pressing issue at the core of the exercise. When real humans are being obviously plagiarized, holding the culprit responsible is often viewed as a way of rectifying the harm caused to this other party. But when the "other" is unidentifiable, what harm is being caused? Why are so many people upset by the thought of a student getting an "A" on an essay produced by an LLM?

Plagiarism is perhaps best viewed as an attempt to standardize a prescriptive claim about intellectual morality. In higher education, "plagiarism evokes deeply held emotions related to deviance, credibility, and what it means to be outside the norm" (Rooksby quoted in McCarthy 2023, 4). So even when a particular victim may be difficult to identify, submitting an essay you did not write undermines the core tenets of an academic meritocracy: you should be assessed based on what you know and how well you can articulate that knowledge. Yet this protective barrier around "what you know" is deceptively precarious. Removed from the author-as-owner paradigm, the phenomenon is nearly impossible to enforce.

To be clear, we do not present this argument in defense of student misconduct or wish to refute the importance of

	LLM Model or Code	**Training Data**	**End User**
Pros of Author Attribution	– Authorship can be relatively easy to name with proprietary models (i.e. those developed and distributed by notable tech companies, such as OpenAI). – May conform to legal arguments concerning private intellectual property.	– Acknowledges the extensive amount of labor and resources needed to develop LLM models. – Draw a connection between LLM outputs and inputs. – Illustrates the general importance of recognizing how ideas are exchanged and built upon in work creation.	– Most straightforward and easiest to identify. – When the user is an individual, this kind of attribution mirrors common authorship assumptions. – Prioritizes the kind of intentionality noted in influential KO work theories.
Cons of Author Attribution	– Both closed- and open-source models build upon preexisting research and code; can be difficult to determine when a new model or piece of code is distinct enough to constitute a "new" entity/work.	– Immense size of training dataset limits our ability to determine whose content is being utilized. – Aggregate nature of parametric memory makes it impossible to reverse engineer output to determine its originating source(s).	– Fails to account for external labor and resources required to generate output. – The level of a user's effort feels unbalanced with their level of recognition.
Implications for Plagiarism	– Provides legal and financial protection to a for-profit company that is often benefiting from open-source information and unpaid labor.	– Although potential harm can be vaguely acknowledged, the material harm to individuals cannot be coherently articulated or corrected.	– Feels largely antithetical to the spirit of academic integrity and plagiarism policies.

Table 1: Pros and cons

intellectual honesty. Quite the contrary. Simply asking people not to "steal" someone's "property" (i.e. their works and ideas) is a low bar that prevents us from having deeper discussions about what it means to think and live in relationship with others. We should demand more of those within a learning community, and reconsidering our views of solitary authors with wholly distinct ideas provides an opportunity to explicitly acknowledge our reliance on one another. When work production is reframed as a community activity rather than the mark of independent genius, the harm of plagiarism is no longer reduced to a localized interpersonal event.

7. Conclusion: Reframing Accountability

From an educational assessment perspective, excessive use of ChatGPT by a student potentially negates the learning goals a particular assignment was designed to address. This is concerning and worthy of our attention. Yet when instructors or administrators claim that this LLM-sourced content has been "plagiarized," the actual harm caused by the student's action is paradoxically obscured by the debates such claims tend to provoke. In these situations, rather than considering how the people's actions potentially compromised individual responsibility, learning, or community accountability, these debates inevitably focus on the authorial capacity of the language model. As works created using LLMs require management within existing information retrieval systems and KOS, this is a problem that needs to be addressed; KO professionals must establish practical guidelines for how these works are identified and placed in relationship to other works. That being said, depending on how one defines an author (See Table 1), the student may not have plagiarized anything—conversely, depending on the particular stakeholders being prioritized, the intellectual property, content, and ideas of many different people may have been unethically appropriated. Again, this is a difficult and important conversation worth having. However, we do not necessarily believe this is the primary issue educators are attempting to address when they claim a student has committed plagiarism.

We suggest that the problem being raised in these situations is not about plagiarism but the related concept of accountability. Within academia, as in other domains, authorship is used to "confer credit" for a job well done, but the connection between individuals and ideas additionally ensures "authors understand their role in taking responsibility

and being accountable for what is published" (ICMJE 2023). The opaque webs of influence central to LLM writing tools complicate our ability to assign responsibility and, consequently, challenge what it means to be held accountable to both oneself and one's community. In such situations, whether or not plagiarism has been committed is ultimately beside the point. Accountability might include a commitment to challenging one's self or avoiding intellectual shortcuts—it may also include not using a particular research tool or method the community has deemed off-limits. Whatever the details may ultimately be, the first step to accountability is ensuring all parties know their obligations. Utilizing ambiguous terminology such as "author" or "plagiarism" to describe a new technology that frankly confuses many people is not setting the community up for success. If accountability is the goal, alternative terminology will likely prove more valuable.

Because many educational institutions already have existing plagiarism clauses included in their academic integrity policies, it makes sense that people would utilize this language to describe a new problem. Yet overreliance on this preexisting term runs the risk of simply complicating matters further by encouraging cyclical, pedantic, and legalistic arguments. Instead of contorting existing terms—such as plagiarism—to accommodate an entirely new situation, there will likely be more success if alternative, more descriptive language is utilized. Given the many troubles with defining what an "author" is under the most standard of publishing circumstances, the notoriously ambiguous term and the related sin of plagiarism are likely to cause more problems than they resolve. At a moment when many people are simply confused about what LLMs even are, the addition of more confusing variables will likely make matters worse.

ChatGPT is a powerful tool with many promising pedagogical uses—at the same time, it can also facilitate nonlearning and perpetuate harmful educational practices. The integration of LLM tools into different domains will continue to reveal possible benefits and consequences, and our affective responses to these applications are perhaps best taken as opportunities to reevaluate the social values authorship and idea ownership have come to represent. As the discussions prompted by these new technologies lead us to reflect on our values, we are provided with a meaningful opportunity to reaffirm those congruent with our aspirations and let go of what no longer serves us.

Endnotes

1. This description is a simplified explanation of a typical training routine for a neural language model. For a more detailed overview, see Jurafsky and Martin (2023).
2. The exception to this is GPT-4, which is generally assumed to have far more parameters than the GPT-3 class. The exact number of parameters has not been released, but OpenAI CEO Sam Altman has strongly suggested that it has fewer than 100 trillion (Vincent 2023).
3. Current events challenging the publication history of high-profile university faculty and administrators (Hartocollis 2024) also show how claims of plagiarism can be used for professional, personal, and political means.
4. Though, as we describe in Section 2.3, the datasets used to train ChatGPT and the sources of its parametric memory are largely not the sole property of OpenAI.

References

Anthropic. N.d. "Meet Claude." https://www.anthropic.com/product.

Bandy, John, and Nicholas Vincent. 2021. "Addressing 'Documentation Debt' in Machine Learning: A Retrospective Datasheet for BookCorpus." In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, edited by J. Vanschoren and S. Yeung.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Bi" In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)* 3-10 March 2021 Virtual Event Canadá. New York: ACM, 610-23.

Bender, Emily M. and Alexander Koller. 2020. "Climbing Towards NLU: On Meaning, Form, and Understanding in the Age of Data." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, edited by Dan Jurafsky, Joyce Chai, Natalie Schluter and Joel Tetreault, 5185–98. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.463.

Bengio, Yoshua, Réjean Ducharme, Pascal Vincent, and Christian Janvin. 2003. "A Neural Probabilistic Language Model." *Journal of Machine Learning Research* 3: 1137-1155.

Birhane, Abeba, Vinay Uday Prabhu, and Emmanuel Kahembwe.. "Multimodal datasets: misogyny, pornography, and malignant stereotypes." Preprint, submitted October 2021. https://arxiv.org/abs/2110.01963.

Bloom, Harold. 1997. *The Anxiety of Influence*. 2nd ed. Oxford: Oxford University Press.

Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, T. J. Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeff Wu, Clemens Winter, Christopher

Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. "Language Models are Few-Shot Learners." *Advances in Neural Information Processing Systems* 33:1-25. https://papers.nips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

Common Crawl. n.d. "About." https://commoncrawl.org/about/.

Cutter, Charles A. 1876. *Rules for a Printed Dictionary Catalogue.* U.S. Bureau of Education, Special Report on Public Libraries, Part II. Washington, D.C.: U.S. Government Printing Office.

Dehouche, Nassim. 2021. "Plagiarism in the Age of Massive Generative Pre-trained Transformers (GPT-3)." *Ethics In Science And Environmental Politics* 21: 17-23. https://doi.org/10.3354/esep00195.

Deleuze, Gilles and Félix Guattari. 1987. *A Thousand Plateaus: Capitalism and Schizophrenia.* Translation and Foreword by Brian Massumi. Minneapolis: University of Minnesota Press.

Dey, Sneha. 2021. "Reports of Cheating at Colleges Soar During the Pandemic." *NPR*, August 27, 2021. https://www.npr.org/2021/08/27/1031255390/reports-of-cheating-at-colleges-soar-during-the-pandemic.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. 2019." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies June 2019 Minneapolis, Minnesota*, edited by Jill Burstein, Christy Doran and Thamar Solorio. Association for Computational Linguistics, 4171–86. doi: 10.18653/v1/N19-1423.

Dodge, Jesse, Maarten Sap, Ana Marasović, William Agnew, Gabriel Ilharco, Dirk Groeneveld, Margaret Mitchell, and Matt Gardner. "Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus." 2021. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, November Online and Punta Cana, Dominican Republic*, edited by Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih. Association for Computational Linguistics, 1286–1305. https://doir.org/10.18653/v1/2021.emnlp-main.98.

Dolan, Jill. 2023. "Guidance on AI/ChatGPT: Memo to All Teaching Faculty - January 25, 2023." The McGraw Center for Teaching and Learning, Princeton University. https://mcgraw.princeton.edu/guidance-aichatgpt.

Foucault, Michel. 1977. "What Is an Author?" In *Language, Counter-Memory, Practice: Selected Essays and Interviews*, edited by Donald F. Bouchard, 113-38. Ithaca: Cornell University Press.

Friedman, Batya, David G. Hendry, and Alan Borning. 2017. "A Survey of Value Sensitive Design Methods." *Foundations and Trends in Human-Computer Interaction* 11(2): 63–125. https://doi.org/10.1561/1100000015.

Friedman, Batya and Peter H. Kahn. 2007. "Human Values, Ethics, and Design." In *The Human-Computer Interaction Handbook*, edited by Julie A. Jacko, 1267–92. Boca Raton: CRC Press.

Gehman, Samuel, Suchin Gururangan, Maarten Sap, Yejin Choi, and Noah A. Smith. 2020. "RealToxicityPrompts: Evaluating Neural Toxic Degeneration in Language Models". 2020. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, edited by Trevor Cohn, Yulan He, and Yang Liu, 3356–69. Online: Association for Computational Linguistics, 2020. https://doi.org/10.18653/v1/2020.findings-emnlp.301.

Gnoli, Claudio. 2018. "Mentefacts as a Missing Level in Theory of Information Science." *Journal of Documentation* 74, no. 6: 1226-1242. https://doi.org/10.1108/JD-04-2018-0054.

Hartocollis, Anemona. 2024. "The Next Battle in Higher Ed May Strike at Its Soul: Scholarship." *The New York Times*, January 14, 2024. https://www.nytimes.com/2024/01/14/us/plagiarism-harvard-claudine-gay-neri-oxman.html.

International Committee of Medical Journal Editors (ICMJE). 2023. "Defining the Role of Authors and Contributors: Why Authorship Matters." https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html.

Jenkins, Baylee D., Jonathan M. Golding, Alexis M. Le Grand, Mary M. Levi, and Andrea M. Pals. 2022. "When Opportunity Knocks: College Students' Cheating Amid the COVID-19 Pandemic". *Teaching of Psychology* 50, no. 4: 407-19. https://doi.org/10.1177/00986283211055906

Jurafsky, Dan and James H. Martin. 2023. "Neural Networks and Neural Language Models." In *Speech and Language Processing*, 3rd edition. Preprint, submitted January 2023. https://web.stanford.edu/~jurafsky/slp3/.

Kasneci, Enkelejda, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, Stepha Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jürgen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, Jochen Weller, Jochen Kuhn, and Gjergji Kasneci. 2023. "ChatGPT for good? On Opportunities and Challenges of Large Language Models for Education." *Learning and Individual Differences* 103. https://doi.org/10.1016/j.lindif.2023.102274

Leazer, Gregory H. and Jonathan Furner. 1999. "Topological Indices of Textual Identity Networks." In *Proceedings*

of the 62nd Annual Meeting of the American Society for Information Science, 1999, edited by L. Woods, 345-58. Medford, NJ: Information Today.

Leike, Jan, John Schulman, and Jeffrey Wu. 2022. "Our approach to alignment research."

OpenAI, August 24, 2022. https://openai.com/blog/our-approach-to-alignment-research.

Lewis, Mike, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2019. "BART: Denoising Sequence-to-Sequence Pre-Training for Natural Language Generation, Translation, and Comprehension." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 7871–80. Online: Association for Computational Linguistics, 2019. https://10.18653/v1/2020.acl-main.703.

Littletree, Sandra, Miranda Belarde-Lewis, and Marisa Duarte. 2020. "Centering Relationality: A Conceptual Model to Advance Indigenous Knowledge Organization Practices." *Knowledge Organization* 47, no.5: 410-426. https://doi.org/10.5771/0943-7444-2020-5-410.

Lock, Samantha. 2022. "What is AI Chatbot Phenomenon ChatGPT and Could it Replace Humans?" *The Guardian*. https://www.theguardian.com/technology/2022/dec/05/what-is-ai-chatbot-phenomenon-chatgpt-and-could-it-replace-humans.

McCarthy, Claudine. 2023. "ChatGPT Use Could Change Views on Academic Misconduct." *Dean & Provost* 24(10): 1-4. https://doi.org/10.1002/dap.31202

Nagel, Sebastian. 2023. "March/April 2023 Crawl Archive Now Available." Common Crawl. https://commoncrawl.org/2023/04/mar-apr-2023-crawl-archive-now-available/.

OpenAI. n.d. "Enterprise." https://openai.com/enterprise.

OpenAI. 2022. "Introducing ChatGPT." https://openai.com/blog/chatgpt.

Ouyang, Long, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke E. Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Francis Christiano, Jan Leike, and Ryan J. Lowe. 2022. "Training Language Models to Follow Instructions With Human Feedback." Preprint, submitted March 2022. https://arxiv.org/abs/2203.02155.

Paullada, Amandalynne, Inioluwa Deborah Raji, Emily M. Bender, Emily L. Denton and A. Hanna. 2020. "Data and its (Dis)contents: A Survey of Dataset Development and Use In Machine Learning Research." Patterns 2, no. 11. https://doi.org/10.1016/j.patter.2021.100336

Radford, Alec, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. "Language Models are Unsupervised Multitask Learners." Technical report, OpenAI.

Raffel, Colin, Noam M. Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer." Preprint, submitted July 2020. https://arxiv.org/abs/1910.10683.

Rosch, Eleanor. 1975. "Cognitive Representations of Semantic Categories." *Journal of Experimental Psychology: General* 104, no. 3: 192-233.

Saussure, Ferdinand de. 1959. *Course in General Linguistics*. Translated by Wade Baskin. New York: The Philosophical Society.

Smiraglia, Richard P. 2007. "The 'Works' Phenomenon and Best Selling Books." *Cataloging & Classification Quarterly* 44, nos.3-4: 179-195. https://doi.org/10.1300/J104v44n03_02.

Smiraglia, Richard P. 2019. "Work." *Knowledge Organization* 46, no. 4: 308-319.https://doi.org/10.5771/0943-7444-2019-4-308.

Soos, Carlin and Gregory H. Leazer. 2020. "Presentations of Authorship in Knowledge Organization" *Knowledge Organization* 47, no. 6: 486-500. https://doi.org/10.5771/0943-7444-2020-6-486.

Svenonius, Elaine. 2009. *The Intellectual Foundations of Information Organization*. Cambridge, MA: MIT Press.

University of California, Los Angeles (UCLA). 2023. "ChatGPT and AI: Starting Points for Discussion." Online Teaching & Learning. https://online.ucla.edu/chatgpt-ai/.

University of Washington. 2023. "ChatGPT and other AI-based tools." Center for Teaching and Learning. https://teaching.washington.edu/topics/preparing-to-teach/academic-integrity/chatgpt/.

University of Wisconsin-Madison. 2023. "Considerations for Using AI in the Classroom." L&S Instructional Design Collaborative. https://idc.ls.wisc.edu/guides/using-artificial-intelligence-in-the-classroom/.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Ł. Ukasz Kaiser, and Illia Polosukhin. 2017. "Attention Is All You Need." *Advances in Neural Information Processing Systems* 30:1-11. https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Vincent, James. 2023. "OpenAI CEO Sam Altman on GPT-4: 'People Are Begging to Be Disappointed and They Will Be.'" *The Verge*, January 18, 2023. https://www.theverge.com/23560328/openai-gpt-4-rumor-release-date-sam-altman-interview.

Wager Elizabeth, and Sabine Kleinert .2012. "Cooperation Between Research Institutions and Journals On Research Integrity Cases: Guidance From The Committee On Publication Ethics (COPE)." *ACTA Informatica*

Medica 20, no.3:136-40. https://doi.org/10.5455/aim.2012.20.136-140.

Winkler, Till and Sarah Spiekermann. 2021. "Twenty Years of Value Sensitive Design: A Review of Methodological Practices in VSD Projects." *Ethics and Information Technology* 23:17–21. https://doi.org/10.1007/s10676-018-9476-2.

Yang, Zhilin, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2019. "XLNet: Generalized Autoregressive Pretraining for Language Understanding." In *Proceedings of the 33rd International Conference on Neural Information Processing Systems* 8-*14 December 2019 Vancouver, Canadá*, edited by Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc and Emily B. Fox. Red Hook, NY.

Zhang, Zhengyan, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, and Qun Liu. 2009. "ERNIE: Enhanced Language Representation with Informative Entities." In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* 8 July-2 August 2009 Florence, Italy. Association for Computational Linguistics, 1441–51. https://doi.org/10.18653/v1/P19-1139.

Zhu, Yukun, Ryan Kiros, Richard S. Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba and Sanja Fidler. 2015. "Aligning Books and Movies: Towards Story-Like Visual Explanations by Watching Movies and Reading Books." *2015 IEEE International Conference on Computer Vision (ICCV)*: 19-27.