Generative AIs, more-than-human authorship, and Husserl’s phenomenological ‘horizons’

Lesley Gourlay, Department of Culture, Communication and Media, University College London Institute of Education, UK, l.gourlay@ucl.ac.uk

Abstract
In recent years, developments in Artificial Intelligence have produced Large Language Models (LLMs) leading to a form of generative AIs (GAIs) trained on vast corpora of texts, capable of producing convincing predictive synthetic texts such as essays, reports, or other texts instantaneously. This development has profound implications for human entanglements with technologies in terms of how these AIs might constitute new forms of subjectivities, texts, and knowledge practices. This development has particular significance for higher education in terms of academic writing, assessment, and research. In terms of higher education, specific concerns have been raised about the implications for assessment, study practices, and the status of knowledge and learning in the context of these ‘writing machines’. Universities and government bodies have reacted in various ways, with some commentators calling for the sector to ‘embrace’ these generative AIs (GAIs) as merely the latest ‘tools’ available for study and research, while others seek to outlaw their use. The current academic research surrounding this phenomenon is underdeveloped with relatively few studies having been constructed, due to the rapid recent acceleration of the technology. It is also under-theorised; early research has reacted to the fast-changing landscape of GAIs by focusing on the technical and practical capacities. Meanwhile, supranational private providers such as OpenAI are influencing higher education internationally, with a complex range of effects which are as-yet unknown. This paper reviews the current state of the art in related bodies of research literature and proposes that the field could benefit from a wider variety of critical and theoretical perspectives. Drawing on the concept of the sociotechnical imaginary from science and technology studies, it considers how discourses and practices surrounding GAIs are evolving in society and education. It then considers the effects of authorship and the writing subject, with reference to the concept of more-than-human authorship. It then draws on recent research in the philosophy of technology, proposing Husserl’s outer and inner ‘horizons’ as a potential framework with which to consider the complex entanglements of human and nonhuman agency, as enrolled in more-than-human authorship and entangled with the presence of GAIs in the ‘lifeworld’ of contemporary higher education. It concludes by proposing future directions for work in this area, in order to gain better theoretical purchase on the phenomenon at the various levels set out above.

Keywords

Introduction
Large Language Models (LLMs) and generative Artificial Intelligence (GAI)s which draw on extremely large-scale bodies of texts by ‘scraping’ the internet have recently become available for widespread use, with OpenAI’s AI Chat Generative Pre-trained Transformer, or ChatGPT. This was launched in Nov 2022, and had been accessed by 1.6 billion users by Mar 2023, an increase of over 55% from the previous month (Ruby 2023). The introduction of this technology, which can ‘write’ plausible texts of all kinds and take part in convincingly human-like interaction with the user, has led to a complex set of questions and controversies in terms of the possible impacts on human society. The potentials are far-reaching in terms of gathering and presenting information instantaneously in a manner which would be impossible for a human individual, given the technology’s ability to draw together and synthesise text from a vast corpus. However, a range of concerns have been raised about potential risks to human society, work, and education – and even to the existential security of humanity itself, as commercial tech giants engage in a race to develop General Artificial Intelligence, without sufficient safeguards currently in place to provide protection to human existence. In terms of higher education, specific concerns have been raised about the implications for assessment, study practices, and the status of knowledge and learning in the context of these ‘writing machines’. Universities and government bodies have reacted in various ways, with some commentators calling for the sector to ‘embrace’ these generative AIs (GAIs) as merely the latest ‘tools’ available for study and research, while others seek to outlaw their use. The current academic state of the art surrounding
this phenomenon is underdeveloped with relatively few studies having been conducted, due to the rapid recent acceleration of the technology. It is also under-theorised; early research has reacted to the fast-changing landscape of GAIs by focusing on the technical and practical capacities. Meanwhile, supranational private providers such as OpenAI are influencing higher education internationally, with a complex range of effects which are as-yet unknown. Higher education institutions are forced to decide how to respond to these potentially revolutionary technologies within the wider contemporary and historical contexts in which they operate. Individual faculty staff and students are orientating themselves to the emergence and increasing uptake of GAIs, perhaps by adopting or tailoring their embodied knowledge and writing practices.

The sociotechnical imaginaries of generative AIs

Writing and authorship are phenomena which take place at the level of individual embodied human practice, but they are also human activities which shape societal and political life at a large scale in terms of how we communicate, work, organise society, and generate knowledge. An example of this is the influence of microblogging social media sites such as Twitter, whose short form and instantaneous publication to a large readership has had a profound effect on opinion formation, political movements, and the broadcast of news, in addition to how users engage with popular culture. Shifts in mediatic practices, as seen with the development of the printing press, have historically had widespread effects on social structures and practices, the professions, and day-to-day ways of communicating meanings and understanding knowledge (e.g. Friesen 2017). A fundamental shift in how texts are produced and distributed may also have disruptive effects on the values, priorities, ideologies, and fears of that society. In the case of synthetic texts produced by AIs, the implications extend to questions regarding the nature and boundaries of human agency, and speculation regarding the possibilities of non-human sentence of AIs. In 2021, Google’s LLM-based ChatBot generator Language Model for Dialogue Applications (LaMDA) was at the centre of a controversy when Google engineer Blake Lemoine raised ethical concerns about this new conversational AI, specifically, that it had become a sentient being, based on his interaction with the bot (Lemoine 2021, Tiku 2022). Claims that LaMDA is sentient have been refuted on the basis that they merely recombined existing human texts based on prompts (Griffiths 2022), with AIs being likened to ‘stochastic parrots’ playing a role by responding in terms of the probability of the next word in the text on the basis of textual corpora (Bender et al 2021). However, these technologies produce plausible dialogue and texts that have been shown to ‘feel human’ and as a result have been anthropomorphised by users, raising concerns about accuracy and safety (Berkovitz 2022). It is noteworthy that this controversy has also sparked an intense debate about the boundaries of the human and non-human in the public domain, in addition to the large amount of media attention and controversy afforded to the GAI ChatGPT. This includes positions ranging from utopian predictions of benign technology releasing humans from tedious and time-consuming research and writing, through to expressions of concern at the risks posed by these AIs. A further controversy erupted at Google, in which researchers raised the alarm about a range of risks posed by GAIs, including environmental damage (Rillig et al 2023), and the reinscription of discriminatory language and views in synthetic text generated by internet-based texts (Bender et al 2021). GAIs are not simply neutral and uncontroversial tools, but are disruptive, contentious, and polarising due to their uncanny nature (Costello 2023), which appears to disturb our notions of voice and what is human, their appearance of ‘sentience’, the ethical ambiguities and issues surrounding authorship, ethics, and trust. They are not politically neutral, but function as echoes or amplifiers of the values inscribed in the texts they are trained on. How the field of higher education orients itself to GAIs ethically, how they are viewed as entities, how governments, NGOs, and educational policymakers respond to synthetic texts or are positioned by them, are vital questions as these technologies develop. Science and technology studies scholars concern themselves with how technologies operate in society, what their effects are, and how societies enrol technologies to express values and ideologies. In Jasanoff and Kim’s influential framing, sociotechnical imaginaries are defined as ‘…collectively held, institutionally stabilized and publicly performed visions of desirable futures, animated by shared understandings of social life and social order attainable through and supportive of advances in science and technology’ (Jasanoff & Kim 2015, p. 6). The concept was originally applied in analyses of the futures-focused ideologies and large-scale infrastructural activities of the nation-state, and has subsequently been used to elucidate how technologies in a range of contexts are themselves products of particularly ideological positions held in complex and often contradictory ways across societies and large organisations (e.g. Barker & Jewitt 2022, Mager & Katzenbach 2021), and are also constitutive of these ideologies. The development of GAIs, their reception, and how they go on to develop may be seen as part of a wider sociotechnical imaginary concerning how we regard ourselves as humans, our relationships to digital technologies, and our orientations towards the future.
Generative AIs and more-than-human authorship

GAIs are textual entities, engaged in the processing and generation of meaning. Users engage with the technology for literacy purposes, which are socially and materially situated in specific contexts governed by complex norms, rules, and expectations, not only in terms of grammar and vocabulary, but also in terms of generic expectations of different forms of texts, and expectations or assumptions surrounding their human authorship (Swales 1990, Bazerman 2000, Devitt 2015). An example of this would be the academic essay, which follows a number of generic and stylistic conventions, with the expectation that it is authored by a single human individual enrolled in a credit-bearing course. ChatGPT has been found to be capable of producing plausible essays of pass standard. This point has garnered a great deal of attention in the context of higher education, in which many high-stakes assessments take place in the medium of essays written outside of invigilated exams (Sharples 2022a, 2022b). The potential for students to deploy an AI to produce an essay raises a host of questions about potential widespread disruption to this already troubled form of assessment, which has been vulnerable for many years to plagiarism in the form of internet-based ‘patchwriting’, and also by human authors employed by ‘essay mills’. Similar concerns have been raised surrounding other forms of writing and how AIs might destabilise the notion of trust around research and authorship. The technology has only very recently entered the mainstream, and how it affects practices in terms of their approaches to writing and texts, as well as how ‘more-than-human’ hybrid texts might emerge, will be a crucial question to consider.

Applied linguistics and literacy studies have a rich seam of theoretical and empirical work on how texts travel, mutate, combine, cross boundaries, and are hybridised (Kell 2015, Silverstein & Urban 1996). The emergent field of posthuman applied linguistics (Pennycook 2018a, 2018b) has explored how human and nonhuman actors interact to create meaning in complex ways. The question arises as to how universities should respond to this development (Lund & Wang 2023, Milano et al 2023, Sharples 2022a), with potentially far-reaching consequences if essay writing were to be abandoned or modified (Rudolph et al 2022). Some educationalists advocate that GAIs should be treated like any other tool, while others regard the use GAIs as a form of plagiarism (Perkins 2023, Eke 2023, King 2023, Fyle 2022). A third approach is to ‘embrace’ GAIs and adapt practices in order to incorporate the use of AIs into higher education learning and assessment (Kasnci et al 2023, Haleem et al 2022, Bozkurt et al 2023, Otsuki 2023). However, a range of challenges surround the status of commercially produced AIs owned by providers who have no obligation towards the practices and values of higher education institutions. Critical studies in digital education form a small but growing field in which the roles of platform capitalism, big data, algorithmic cultures, surveillance technologies, and learning analytics have been interrogated, informed by alternative theoretical positions regarding the sociomaterial nature of education processes (Fenwick et al 2011, Gourlay & Oliver 2018), human-nonhuman digital assemblages (Gourlay 2021), the ethics of AI in education (Holmes & Porayska-Pomsta 2022), critiques of digital assetization (Komljenovic 2022) and the technofinancial ‘futuring’ of edtech in higher education (Williamson & Komljenovic 2022). This rich field of enquiry offers the conceptual and methodological apparatus for detailed investigation of how GAIs are operating and having effects in universities.

Generative AIs, embodied practices and posthuman texts

The act of writing cannot be understood as a disembodied information transfer or the production of text, just as any technology or device used for writing cannot be regarded as a mere instrument or tool (Vlieghe 2016), whether the writer is using paper and pen, or a digital device. Instead, any form of writing, from longhand using graphomotoric control, to the motor space and visual space of keyboards and screens (Mangen & Velay 2010) enrols the body in interaction with a physically tangible device. An example of this is the orientation and movement of the body, hands, eyes, and device when using a laptop, as opposed to texting using a mobile phone, or writing using paper and pen. As such, the materiality of communication (Lenoir 1998) is not neutral but is agentive and constitutive of human subjectivities as writing bodies. Work in media theory has revealed the closely intertwined relationships between technologies of inscription and the material, cultural and ideological forms of education (McLuhan 1994, Friesen & Cressman 2010, Friesen 2017) and the generation of ‘alphabetized bodies’ (Kell 1998). Stiegler, working in the philosophy of technology, holds that our embodied subjectivities are formed through our prosthetic interrelationships with technologies (Stiegler 1998). Vieghe (2016) and calls for a technosomatic (Richardson 2010) account of how practices of reading and writing impact our bodies and the social and cultural effects of that entanglement, arguing that this leads to ‘a grammar’ of responding to the world via gestural routines. Typing differs from the productive gesture of handwriting as it can be seen instead as a pointing gesture (Mangen & Velay 2010), in which the keyboard writer becomes a consumer, with different technological arrangements opening up different spaces of experience in which we come to inhabit a particular
relation to script (Vlieghe 2016). In this regard, one of the pressing questions for research in this area centres on how GAs will change human practices of writing and generating texts, and how these effects will be felt in the everyday practices of education, writing, study and meaning making.

Textual practices in education and beyond have always been intertwined with technologies of inscription, from the earliest recorded human practices of writing using clay tablets (Friesen 2017), to the emergence of digital technologies. As such, texts have always emerged from and formed part of complex assemblages of human and non-human actors, including artefacts, devices, institutions, generic conventions, and specific embodied and temporal practices. Theorists working in social semiotics have in recent decades accounted for the multimodal nature of communication (Kress & van Leeuwen 1996), and more recently the complex materiality of textual communication has been theorised in posthuman applied linguistics (Pennycook 2018a, 2018b). The latter field of scholarship has revealed the sociomaterial nature of agency in meaning making, challenging humanistic assumptions about how texts are produced. The advent of GAs requires further theoretical extension to provide insights into how these ‘more-than-human’ machine–human texts emerge, how we might begin to understand the nature of authorship in the AI age, and how to understand the rapidly evolving and uncanny nature of these texts (Costello 2023). These AIs radically disrupt notions of authorship, intellectual property, the boundaries of institutional and individual responsibility, and ultimately raise profound ontological questions about how we understand human sentence, semiosis and intentionality, and how we respond to the AIs’ plausible appearance.

Phenomenology and academic writing

The way that we experience writing has been explored in detail from a first-person perspective using the approach of phenomenology. Phenomenology has been defined as follows: ‘Phenomenology is the study of structures of consciousness from the first-person point of view. The central structure of an experience is its intentionality, its being directed towards something, as it is an experience of or about some object.’ (Stanford Encyclopaedia of Philosophy 2013). It came to prominence as a branch of philosophy in the work of Husserl (1970), Heidegger (1978a, 1978b) and Merleau-Ponty (2012), among other thinkers. The focus is on how we experience things, ‘… notably the significance of objects, events, tools, the flow of time, the self, and others, as these things arise and are experienced in our ‘lifeworld’.’ (Stanford Encyclopaedia of Philosophy 2013). An analysis using this perspective reveals in detail the nature of first-person experiences, focusing on aspects such as time, space, where we place our attention, the actions we take with our bodies, eyes and movements, the way we interact with others, and our linguistic activity. There is also a focus on what conditions of possibility are brought into being by things.

In recent years postphenomenological enquiry has been developed to guide descriptions and enquiries into the nature of the experience of technologies in particular (e.g. Ihde 1990, 2002, Verbeek 2015a, 2015b, 2023). Further work has focused on digital technologies with a posthuman lens, such as Adams and Thompson’s (2016) Researching Posthuman Worlds, in which they developed a series of heuristics to be used to investigate the nature of experience of digital technologies, by an approach they characterise as ‘interviewing objects’; the methodology I applied in Gourlay (2021) to investigate the effects of digital technology on textual practices, writing and meaning-making in higher education. My investigations revealed the profound influence that digital technology has on the day-to-day practices of study and academic writing. GAs and their effects may also be investigated in this manner, as Adams has recently demonstrated in a study where she ‘interviews’ AI in education using these heuristic questions to investigate teachers’ experiences of using GAs (Adams 2023). Marin (2021) analyses the study practices of the university in terms of what she terms mediatistic displacement, in which texts move between various modes of speech and text, taking a phenomenological approach which focuses on sensory configurations and the mediality of embodied gestures, drawing on the work of Flusser (2011, 2014). Her analysis of interviews with students reveals in detail what she calls the gestures of academic writing as study practice. She identifies three categories: dis-assembling, assembling, and interlacing. The first gesture of dis-assembling involves ‘smashing’ texts that have been read as part of a literature review, using quotes, or paraphrasing elements, using a range of actions such as wiring digitally, handwriting, highlighting, placing sticky notes, circling text, typing notes, and so on. The second and third gestures of assembling and interlacing, she argues reveal that it is not a linear process of writing a text from the start to the finish, but more like ‘assembling a collage’ (Marin 2021, p. 47). Marin (2022) raises the question of the effect the effortlessness of generating texts using AIs might have on study practices, as students become disconnected from the process of reading, researching, and writing. It is crucial to recognise that GAs fundamentally alter the nature of the experience of academic writing in terms of the gestures they bring forth, and therefore how the embodied authoring self and knowledge practices unfold.
Husserl’s horizons and technological intentionality

Working in the philosophy of technology, Mykhailov and Liberati (2022) use concepts from phenomenology and postphenomenology to analyse the nature of programs developed by the higher order computer programming language C++, and the unsupervised learning technique used in machine learning, the ‘Generative Adversarial Model’. They refer to Husserl’s (1939, 1966) suggestion that objects passively ‘relate’ to us when we look at them. For them, this apparently small step in Husserlian phenomenology is significant, ‘…since it shows how the object is active in the intentional relation binding subject and object’ (Mykhailov & Liberati 2022, p.843). In this regard, the object is seen as constituting itself independently of the observing subject. However, its agency goes further in that ‘By constituting itself, the object directly shapes human behaviours, human goals, and preferences, as if it were an active entity.’ (Mykhailov & Liberati 2022, p.843). As they put it:

The fact that objects might ‘act’ towards the human subjects opens a new perspective on the ‘human-technology’ relationship. 5 Objects are not ‘neutral’ because they are not dead and inert things around the subject. The object is not a passive entity that depends on the subject, but it is an active ‘thing’ that can dynamically participate in the world. Thus, by simply looking at the passive synthesis in phenomenology, it is possible to frame the object’s presence as something ‘alive’ that directly acts on the subject. Consequently, according to this perspective, objects have a specific ‘intentionality’ pointing to the subject.

(Mykhailov & Liberati 2022, p.844).

They go on to consider the nature of the object in terms of Husserl’s concept of its horizons (Husserl 1939, 1950). They focus on the ‘outer horizon’ of an object, which consists of aspects of the background which constitute the object, ‘…a connection binding the object to what is around it, which is produced and structured by the object itself.’ (Mykhailov & Liberati 2022, p.845). The ‘inner horizon’ of an object refers to aspects of the object which are not perceived by the subject:

‘…according to Husserlian phenomenology, this ‘horizon’ makes the objects transcendent, because the subject can never perceive all the aspects of an object at once… the object is ‘infinite’ because it has more to offer than what is manifest in front of the perceiving subject… Thus, thanks to all these elements that are part of the inner horizon, it is possible to think of the object as an infinite resource of different praxes since it constantly shows different elements and changes through time. The subject does not produce this infiniteness, but the object in its constitution generates it.’

(Mykhailov & Liberati 2022, p.845).

Their subsequent analysis goes on to look at unsupervised machine learning in the context of Generative Adversarial Models. They point out that computers are not normally regarded as ‘dead’ objects, but as active and interactive entities. They focus on the ‘outer horizon’ of computer programs developed with C++, pointing out that the other technological objects are required in order to write executable code, alongside material artefacts such as screens, keyboards and so on. Human and non-human agents work together in various ways, with the latter stages of the process taking place without human interaction. In terms of Husserlian phenomenology, the ‘outer horizon’ allows the object in the form of the program to autonomously link itself to other programs. In this regard, the program is able to independently produce its own intentionality. Husserl’s ‘inner horizon’ allows a theorisation of how unsupervised machine learning allows the system to change its parameters by itself, acting independently of the programmer’s intentions. The Generative Adversarial Model is so-called because it consists of two networks, a generator and a discriminator which interact with each other. As they put it, in such a case: ‘The object is transcendent since the object’s inner horizon cannot be grasped by the subject in its totality all at once’ (Mykhailov & Liberati 2022, p.845); and they characterise it as ‘infinite’ in terms of its digital system.

Mykhailov and Liberati point out that the field of study of technological intentionality lacks the use of in-depth phenomenological analysis, which they claim can show how objects are ‘alive’, with the outer horizon showing their active role, and the inner horizon revealing their ‘inner life’. As they put it, ‘The object constitutes itself through the passive synthesis, it proposes itself as infinite to the subject thanks to the richness of its inner horizon, and it connects itself to other objects around thanks to its outer horizon.’ (Mykhailov & Liberati 2022, p.851).

They propose this framework to better understand the technological intentionality of computing systems. The same analysis could be brought to bear on Generative AIs such as ChatGPT in education, allowing us greater theoretical purchase on these two aspects of their ontology – both how they act on objects and humans around

them in the world, and also how they are composed internally in terms of their nature, potential, and intentionalinity. Such an analysis would allow for a fresh consideration of the both the ‘real-world’ challenges posed by generative AI in education as set out above, its effects of human authorship, epistemologies and ontologies, and a perhaps a more nuanced insight into the ontology the AIs themselves.

Conclusions and future directions

The current field of research and literature on AIs in higher education has several areas in urgent need of development. Arguably, the majority of work on AIs in education has its origin in the field of educational technology, and tends to focus on the claimed potentials of AIs for education, adopting a stance of promoting these technologies, with scant emphasis on critical engagement with the broader effects on society, the sociopolitical dimensions of the for-profit private providers driving these development within a context of ‘platform capitalism’, the risks to society, and the effects on educational sectors, institutions and the groups and individuals who study or work in them. As such, the field can be said to be under-theorised, fragmented and in need of coherence and depth of analysis with a wider and more ambitious scope, drawing on a richer set of disciplinary perspectives, theories, and methodologies. Key questions are raised regarding how human subjects orient themselves to AIs, the degree to which they are perceived as in some sense sentient or agentive, and, crucially, the impacts of these on how knowledge is perceived, how it is generated, and what are the reverberations for knowledge practices in higher education. The insights and perspectives of phenomenology as discussed above may offer a means by which as a field we can consider the experience of students and faculty as authors alongside and entangled with generative AI in the ‘lifeworld’ of contemporary higher education.

References


Milano, S., McGrane and Leonelli, S. 2023. Large language models challenge the future of higher education. *Nature Machine Intelligence* [https://doi.org/10.1038/s42256-023-00644-2](https://doi.org/10.1038/s42256-023-00644-2)


Otsuki, G. J. 2023. OK computer: To prevent students cheating with AI text-generators, we should bring them into the classroom. *The Conversation* 23 Jan, [https://theconversation.com/ok-computer-to-prevent-students-cheating-with-ai-text-generators-we-should-bring-them-into-the-classroom-129905](https://theconversation.com/ok-computer-to-prevent-students-cheating-with-ai-text-generators-we-should-bring-them-into-the-classroom-129905)


Rillig, M., Agerstrand, M., Bi, M., Gould, K. and Sauerland, U. 2023. Risks and benefits of large language models: phenomenological and postphenomenological perspectives. *Environmental Science and Technology* 57(9), 3463-3466 [https://doi.org/10.1021/acs.est.3c01106](https://doi.org/10.1021/acs.est.3c01106)


This research was funded by the Leverhulme Trust, Major Research Fellowship MRF-20201-135