



Open Library of Humanities



Part of the Ubiquity
Partner Network



Open Library of Humanities

Reading in Ruins: Exploring Posthumanist Narrative Studies

How to Cite: Elstermann, A 2020 Computer-Generated Text as a Posthuman Mode of Literature Production. *Open Library of Humanities*, 6(2): 31, pp. 1–23. DOI: <https://doi.org/10.16995/olh.627>

Published: 23 December 2020

Peer Review:

This article has been peer reviewed through the double-blind process of *Open Library of Humanities*, which is a journal published by the Open Library of Humanities.

Copyright:

© 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Open Access:

Open Library of Humanities is a peer-reviewed open access journal.

Digital Preservation:

The Open Library of Humanities and all its journals are digitally preserved in the CLOCKSS scholarly archive service.

READING IN RUINS: EXPLORING POSTHUMANIST NARRATIVE STUDIES

Computer-Generated Text as a Posthuman Mode of Literature Production

Annika Elstermann

University of Heidelberg, DE

annika.elstermann@as.uni-heidelberg.de

A central idea of posthumanism in a technological society is the actual transition of the human towards a post-human entity, the cyborg. This entanglement between humanity and technology can not only be found in – actual and fictional – cyborgs, but also in computer-generated text production. Through the close collaboration between human creativity and artificial intelligence, algorithmically facilitated writing is emerging as an art form that is proving promising for literary analysis in a posthuman context. This article will examine computer-generated fiction as a new, posthuman mode of text production and use poststructuralist and related theory – mainly Roland Barthes, Michel Foucault, and Susan Sontag – to explore the implications that such forms hold for the roles of authors, readers, and that of literary critics and scholars.

And it is in the light of our senses,
our capacities (rather than those of another age),
that the task of the critic must be assessed.

—Susan Sontag, 'Against Interpretation' (1966)

1. Introduction

In the 1960s, questions surrounding the relationship between author, text, and reader were at the core of the work of critics such as Roland Barthes, Michel Foucault, and Susan Sontag, and even though that decade occurred over half a century ago, poststructuralism and its associated movements still serve as reference points for many approaches to interpretation. While these concepts remain useful and applicable, a recontextualisation seems to be called for in some areas, for the sake of theory itself as well as for new forms of writing. Posthumanism offers a useful starting point for this endeavour, as it has some precursors in these slightly more traditional theories, but also bridges a gap towards the inclusion of non-traditional textual forms. Algorithmically facilitated text production in turn not only makes for an ideal practical reference point for posthuman theory, but also has an inherent affinity with poststructuralist literary experiments that began to deconstruct the idea of the author and of textual meaning.

As technological advances have facilitated greater strides in the field of machine learning, computer-generated text warrants closer attention both as a literary phenomenon and as a catalyst for a re-evaluation of traditional theory in a posthuman context. While it might, at one point, have been possible to put off computer-generated text as elaborate technological gadgetry, we can no longer afford to disregard an entire subset of contemporary cultural output merely by virtue – or perceived lack thereof – of its origin and the underlying process of creation. Rather, literary criticism has much to gain from recognising and examining computer-generated texts as potentially literary not only in spite of, but in some cases precisely because of, their peculiar creation processes. The argument at the core of this article is that the specifically posthuman characteristics of computer-generated texts in fact present an opportunity to fruitfully evaluate established key

concepts of literary theory from a new angle. After a brief outline of what might constitute computer-generated text, I will explore to what extent Barthes' 1967 essay 'The Death of the Author' can be applied to algorithmic text production – and how, in turn, the concept of computer-generated text might influence approaches to Barthes' proposition. This is supplemented and expanded using Foucault's 1969 text 'What is an Author?', in which he points towards difficulties regarding the author's disappearance and problematises attempts to substitute concepts such as writer, scribe, or work, and Sontag's 1966 essay 'Against Interpretation', where she highlights the importance of art as a sensory experience. Finally, a brief analysis of a computer-generated movie script contrasted against the interpretation presented by the filmed version will show how computer-generated texts can be interpreted, and how different readings can be derived from the same material.

Of course, critics in the 1960s did not write their theoretical works with computer-generated texts in mind. This is no disadvantage here – on the contrary. Different viewpoints can be highlighted, and new insights might yet be found by reading them through two new lenses: that of posthumanism, and that of computer-generated texts. By building on theories and schools of thought originally concerned with more traditional texts, we might also be able to find ways of dealing with these fundamentally new forms, rather than relegating them to the unapproachable sidelines of the avant-garde or non-literary.

2. Computer-Generated Texts: The Cyborg in Writing

Computer-generated text is a term that encompasses many different possible processes, some of which will be further elucidated in this article. Attention will be limited to those texts which are presented as poetry, narrative prose, or drama – the kinds of text, that is, that would, under other circumstances of creation, easily be accepted as potentially literary. This does not include automated translation, for example, as the mechanics behind that – both technical and theoretical – are quite different.

Most projects of text generation are based on a corpus of works composed by human writers, with an algorithm either randomly reassembling items in the corpus, or analysing probabilities of syntax and vocabulary within the corpus to predict and

generate a likely combination of words and phrases according to a programmer's specifications. While what we call computer-generated fiction has been around for some time now – Scott French used his programme 'Hal' to generate the romance novel *Just This Once* in 1991 (*Chicago Tribune*, 1991), and early experiments predate even this by decades – it has been gaining considerable traction over the course of the past decade, and has by now also entered into the public sphere. In December 2017, a number of mainstream news outlets reported on a three-page story titled 'Harry Potter and the Portrait of What Looked Like a Large Pile of Ash', produced using a predictive text engine coded by Botnik Studios (2018a) and trained on the text of the seven Harry Potter books (Botnik Studios, 2017). Popular reporting focussed on the novelty aspect of the creation process with a certain degree of sensationalism. Some of the headlines were: 'Bot tries to write Harry Potter book – and fails in magic ways' (Flood, 2017, *Guardian*); 'Harry Potter chapter written by bots is magically terrible' (Cooper, 2017, *CNET*); 'Here's What Happens If You Let a Whimsical Robot Write the Next Harry Potter Chapter' (Bruner, 2017, *Time Magazine*); 'AI Attempts To Write Harry Potter And It Goes Hilariously Wrong' (McCall, 2017, *IFLScience*); and 'There is a new chapter in Harry Potter's story – and it was written by artificial intelligence' (Maggio, 2017, *Business Insider*).

As these headlines already indicate, the chapter, for want of a better word, makes for a slightly weird read, and is funny in an absurdist way. While no discernible logical plot emerges, the text is still humanly readable, every sentence is grammatically sound, and J.K. Rowling's style is recognisable throughout in phrases and imagery. The way in which the creative process is framed in these headlines further indicates that a technological singularity is nigh, and that it is taking the shape of a robot sitting at a keyboard somewhere, typing out *Harry Potter* fanfiction on its own. Phrases such as 'written by bots', 'whimsical robot', and 'written by artificial intelligence' imply that the main portion of the text was in fact produced independently by a machine. The verbs 'tries', 'attempts', and 'let' strongly suggest that this machine has agency and an intrinsic wish to produce this text. The article on *Business Insider* for instance continues with this description of the creative process: 'An artificial

intelligence tool read all of the 'Harry Potter' books and automatically generated a new, self-written chapter out of what it learned. The output text was mostly raw and incomprehensible, so a few writers intervened to make it understandable' (Maggio, 2017).

In fact, the human element in the creation of this short chapter was far more significant than post-creation editing. In response to a question on Twitter, Nat Towsen, a writer for Botnik, explained:

It's not automated! We have a team of writers who all use the Botnik predictive text keyboard. We trained keyboards on all 7 books and had a big writing jam. Then I took the best pieces of copy, arranged them into a narrative, and wrote some copy to fill in the gaps. (Towsen, 2017)

Having a predictive text keyboard that was trained on those seven books means that an algorithm identified patterns in J.K. Rowling's novels – frequently occurring names, phrases, word clusters – and presented the human Botnik writers with a choice of eighteen words, each of which would be likely to follow the current one in J.K. Rowling's writing style. The team of writers that Towsen mentions picked one of these eighteen words and were then again presented with a choice for the next probable word to follow after that.¹ This is in essence a practical realisation of Fredric Jameson's definition of pastiche as 'the imitation of a peculiar or unique, idiosyncratic style, the wearing of a linguistic mask, speech in a dead language' (Jameson, 1991: 17). At the very least, the process of composition – an author's style borrowed by other writers via an algorithm – in this case explains the syntactic coherence of the final product.

The disparity between the actual process and popular reporting already highlights a central question in the field of computer-generated texts: who wrote this? Certainly

¹ A variety of differently-trained predictive keyboards, including the ones for *Harry Potter* narration and dialogue used in this example, are publicly available via <https://botnik.org/apps/writer/> (Botnik Studios, 2018b).

not a solitary, artistic machine, but at the same time, no solitary, artistic human author, either. In other words, how human, nonhuman, or posthuman are these processes of text creation?

Considering how algorithmic text production is being framed in popular reporting essentially as the triumph of the machine over human creativity, it would be easy to assume that there is no space for humanity in this field. In fact, however, human influence is essential to the output of text-generating programmes, and human agency finds footholds at many different points in the process, as has already been made evident in the example of the *Harry Potter* chapter mentioned above. At the most principal level, no algorithm can come into existence without a human programmer, an actual writer, somewhere at its root. Human influence is thus literally hard-coded into the text creation. Depending on the methods employed, it occurs elsewhere, too. No computer programme can generate words out of nothingness. Whether the programme in question is a complex recursive neural network or simply a tool that combines words into random strings, these words need to be pulled from some source, and that source corpus will always ultimately consist of words and phrases thought and written by humans. At the other end of the production process, human intervention is also possible, and frequently pragmatically necessary. Text might be formatted for print or electronic publication. Prior to that, it can be selected, cut, rearranged, or otherwise amended. Even a seemingly trivial revision, such as adding punctuation, can influence perceived meaning.

In the case of *Just This Once*, for example, programmer Scott French came up with the original idea, wrote a programme to analyse romance writer Jacqueline Susann's style, established rules regarding plot structure and characters, made decisions at frequent intervals during the text generation (Boudreau, 1993: 1), and significantly edited the final output. French has stated that he wrote about 10% of the text himself, and the algorithm about 25%, and that the remaining 65% of the output was a 'joint venture between the two' (*Chicago Tribune*, 1991). Of course, even the 25% of actual computer-generated text was determined by French's coding and, crucially, by patterns and probabilities identified through the analysis of Susann's works. Indirectly, the words contained in the programme's output are

hers. The difficulty of separating human and non-human creator even at this early stage of computer-generated fiction is evidenced by a legal argument that took place following the publication of the novel (Lohr, 1993: 1) – copyright law certainly does not appear to be equipped to handle the question of who owns the rights to a dead author's writing style (Boudreau, 1993: 2).

The science fiction film *Sunspring* (2016), a more recent example, may boast a more refined algorithm, but involves a similar amount of human influence, though some of it in different manifestations. The recurrent neural network written by Ross Goodwin was trained on a corpus of science fiction movie scripts to generate a screenplay of its own (Newitz, 2016): again, a source corpus that is very obviously and essentially human in its origin. As a short film entered into Sci-Fi London's 48-Hour Film Challenge, the script also needed to be turned into an actual movie with a director, actors, a set designer, and so on. While in this case, there was little to no editorial intervention in the actual printed text, the realisation of the script involved nothing but human interpretation and decision-making.

Not all computer-generated fiction is turned into movies, of course, and twenty-five years after *Just This Once*, publishing dynamics have shifted significantly, so that rigorous human editing is by no means necessary in order to publish a text for a global readership online. Even without this post-text-generation intervention, though, human agency always remains embedded in the code itself. Any algorithm still needs to be written in the first place, and it requires a corpus of texts from which it can draw semantic and syntactic patterns.

If even an algorithm needs to be constructed and utilised, like any piece of technology, then what differentiates a computer programme designed to generate text from any other tool we use for writing? If human involvement is so essential to the machine's output, why call it *posthuman* at all?

Conversely, an argument could be attempted that any act of writing inherently occurs from a hybridity of human and machine, or instrument.² Whether that

² See also Hayles, who argues that "almost all contemporary literature is already digital" (Hayles, 2008: 159) due to the influence that computers have on the process, from writing, to cover design, to printing and publishing.

instrument is a pen, a typewriter, or a keyboard connected to a word processor, the moment in which the act of writing is realised temporarily merges human and machine into a functional unit akin to a cyborg. And, of course, the materiality of text creation matters, just as that of its final presentation and reception does. Whether a writer is working on a wax tablet, a scroll, a page in a codex, a typewriter, or a computer screen will likely play some role when it comes to the ultimate output (see Butler, 2011: 6–10).

Yet with all the influence that the use of any of these tools has on the process of creation, no literary critic would likely postulate that all writing is inherently posthuman. What, then, separates human from posthuman text production? While writing is generally facilitated by tools and may be influenced in its form and even content by the material reality of those tools, every word of textual output still originates in a human writer's thoughts: and not just individual words, but phrases and their combination into meaning. A text-generating algorithm meanwhile takes human input and modifies it – according to human programming and direction, but at a calculating capacity that generally exceeds individual human ability. The human author alone could not feasibly have analysed the same corpus, identified hundreds of patterns, and applied them to a new creation of texts. In short: the precise algorithmic output is essentially unpredictable even for the programmer.

Closer to the centre of this spectrum, however, the waters become decidedly murkier. The word processor that underlines – or even automatically corrects – mistakes in spelling, grammar, and punctuation, that identifies and highlights questionable style, and suggests more fitting words from a built-in thesaurus, for instance, manipulates human output according to a set of pre-programmed parameters that are not necessarily transparent to the writer. This is exacerbated to an even greater degree with any smartphone autocorrect feature – which can in equal parts help or hinder communication – or, more pertinently, predictive texting, where an algorithm programmed into a phone analyses a user's communicative patterns and, during text composition, suggests which word or phrase they might

want to use next.³ Is a message composed this way less human than one written with a pen? To an extent, yes. But is it non-human? Clear lines of demarcation cannot be drawn at these points of intersection, which is what makes these texts decidedly posthuman: the output cannot be definitively separated into man and machine, and at many points, neither can the process.⁴

3. Interpreting Computer-Generated Texts

There might still be some residual unease about texts produced by computers and the idea of writing as a cyborg activity. After all, there is something uncanny about a machine carrying out such a fundamentally human activity, and about not being able to tell what was written by a machine and what was written by a person. With the cyborg as 'a hybrid of machine and organism' (Haraway, 1991: 149), the *humanism* of posthumanism might appear dangerously on the fringes. Crucially, in this context, this inseparability raises the question of whether – and how – texts can be approached in which authorship is not necessarily unclear, but certainly ungraspable in the posthuman entanglement of corpus, programmer, editor, and algorithm. How can literary critics and scholars work with such texts as potential, and potentially meaningful literature?

One obvious approach to this would be to align computer-generated texts with Roland Barthes' 'Death of the Author'. If the authority over that which a text signifies is removed from the author – the writer – and seen as embedded in the text and thus up to the reader, then it should hardly matter whether the text in question was created by a person or a machine or a hybrid of the two. If the origin is discarded, then what we are left with remains, after all, a text, and taking a text for what is on the page can, pragmatically, work just as well for a computer-generated text as it

³ This is not an entirely new idea originating in the iPhone generation of technology. T9 was invented in the 1990s for mobile phones with three-by-four keyboards, and also took word frequency and some user patterns into account as it adapted over time on each individual phone (Grover, King and Kushler, 1998).

⁴ This does not indicate a *protohuman* android (a machine that is almost human and learning to write), but is *posthuman* in that it has its origins in something specifically human, and is based on and transcends ideas of humanism.

does for any other. This approach, if it can be employed, has the great advantage of working categorically, since the words on the page are always the one component that we have for certain, no matter how much or little insight we have into the psyche of their creator or the process of their creation.

However, for most literary scholars, a degree of discomfort regarding the direct application of Barthes here likely still persists. One challenge might be that Barthes' argument is at once a pro- and an anti-humanist one – dethroning the author for a democratisation of interpretation, while simultaneously depersonalising the entire reading process. In 'The Death of the Author', this is not an insurmountable contradiction, but rather a tension that the text utilises as an argument. In the context of posthumanism, however, disregarding the immediate origin of a text might be one step too far away from the human element.

Another, more tangible argument is embedded in Barthes' text itself. That writing does not have 'an ultimate meaning' (54), that is, a single authorial truth to be deciphered, does not suggest that it has no meaning. Rather, the empowerment of the reader implies that a text has the potential for any number of meanings, of which none is by default more correct than another. While we, with Barthes, refuse to submit to a singular authority and accept that an ultimate meaning, a 'true' interpretation can be drawn from a text, we still implicitly assume that this potential for any number of meanings is inherent to that text. It is not that a text has no meaning, it is that potentially, it has all meanings.

In contemporary literary criticism meanwhile, even after 'The Death of the Author', or maybe because of it, scholars might still take it for granted that any of a multitude of meanings can be found in a text because the unknowable writer, or the construct of the implied author, placed it there, however unconsciously. This is a distinction that becomes palpable when we actually remove the individual author from a text and attempt to submit computer-generated texts to the same criteria for critical analysis as humanly written ones – suddenly, the abandonment of individual human originality is not so pragmatically simple anymore.

In his lecture 'What is an Author?', Michel Foucault (1979) argues that the attempt to 'bypass the individuality of the writer or his status as an author to concentrate

on a work [...] has merely transposed the empirical characteristics of an author to a transcendental anonymity' (16–17). In other words – and applied to this context – the Death of the Author has essentially raised the spectre of an author. There might not yet be an 'empty space left by the author's disappearance' (Foucault, 1979: 17) at all, as critics can still comfortably have the author at the back of their minds, as the individual construct of an implied author who might just be congruent with the real person, allowing us to rely on the fact – which may be an illusion – that someone originally ensured a certain degree of textual coherence (see Foucault, 1979: 22).⁵

A consequence of 'The Death of the Author' thus might have been not authorial atheism (in that there is no Author-God), but rather authorial henotheism – recognising that different and potentially rivalling versions of a presumed author and the associated interpreted meaning can coexist with equal validity. The idea of the author can still persist alongside Barthes' plurality of meaning – in that any interpretation could potentially be true; we cannot know which it is, so we suspend the knowing, and do not exclude others, as they might also be true.

The potential multiplicity of meanings of texts composed by individual human writers is in that sense rooted in the unknowable and therefore also potentially infinitely variable mind of the writer, which ultimately – and with Barthes, most importantly – finds its kaleidoscopic representation in the infinitely varied minds of a multiplicity of readers. So how can we, as human readers, assume the same potential for meanings in a text not directly thought and written by another human, but generated as the output of an algorithm? One instinctive reaction might be dismissal: if there is no original authorial truth, whether we want to identify it or not, then there cannot be a multitude of meanings; if there is nothing for the kaleidoscope to refract, then there is simply nothing.

⁵ It bears noting that both Barthes and Foucault use as their point of departure the idea of the individual author and thus exclude a production setting which has since emerged in large movie and video game studios especially: that of the collaborative writing and editing of a script. While this type of writing (which is likely motivated by profit maximisation and thus oriented along audience expectations) would be interesting to examine in a similar context, I will limit my focus here to the general distinction between human writing and that produced by a machine/cyborg.

The machine, like the author according to Barthes (1989), cannot express a specific inner self. But unlike the author, it has no conscious or unconscious impulse to do so, either. For this reason, algorithmic text production cannot be likened to Surrealist 'automatic writing' as Barthes describes it (51), since there might still be a deeply psychoanalytical component to be identified in this not non-conscious, but unconscious writing. If anything, then collective writing, which according to Barthes 'helped desacralize the image of the Author' (51), might be a fitting parallel or precursor, albeit on a very different scale. However, desacralising, dethroning, or disregarding the author is not the same as not having an author, which becomes very clear when aligning 'The Death of the Author' with a form of text production that very pragmatically challenges the traditional concept of an author. Barthes remains caught between neo-humanist notions and a posthuman longing for a focus not on the individual, human person, but on the content of writing, an independence of text from creator and context. Since computer-generated text is itself produced in such a posthuman way, and since the Death of the Author, as has been shown, is so markedly different from the absence of the author, Barthes' approach, while useful, does not appear to be sufficient for these purposes on its own.

Do we need an author (or an equivalent), or just a way of dealing with authorless texts? Foucault, discussing the void left by the author's disappearance (17), substitutes a number of different author-functions. Of these, only the most pragmatic one – legal liability (Foucault, 1979: 20) – actually requires a flesh-and-blood person, and of course this is the function that is furthest removed from the general reader. All other author-functions that Foucault outlines, including the potential for internal textual coherence, filiation, and authentication, are tied rather to a construct in the individual reader's mind (Foucault, 1979: 21) and might be highly subjective. This is underscored by the fact that many of these author-functions as Foucault describes them would be challenged if an author chose to write part of their work pseudonymously.

The concept of, essentially, a placebo-author would of course make for an easy way to circumvent the novelty of computer-generated texts. We can easily assign an identifier in the form of a name to any algorithm; the AI that produced *Sunspring*

has already named itself Benjamin (interview quoted in Newitz, 2016). In fact, personifying computer programmes and attributing agency to them seems to be the approach much of mainstream media is taking at this point. This would allow us to continue with literary criticism as is – and likely relegate computer-generated text back to its avant-garde fringe, barring the occasional brave venture, during which the machine could take the place of the Author-God.

However, any attempt to put the algorithm itself into the guise of human intentionality and agency would entirely minimise those human aspects that are actually present in the production and the reception of these texts. The *deus-quod-machina* is not a posthuman, but a decidedly anti-human concept and, as the Author-God, would limit interpretation rather than allow the exploration of meanings.

A more fruitful path towards a meaningful interpretation that takes into account the peculiarities of computer-generated texts rather than ignoring them for convenience's sake might, somewhat ironically, be found via Susan Sontag's essay 'Against Interpretation'. Precisely because algorithmic text production is decidedly not what Sontag would have had in mind, her proposed approach lends itself ideally to being recontextualised and is well worth closer inspection through the lens of posthumanism – because it emphasises the humanity of art.

Sontag's point of departure is that a work of art contains meaning on a multitude of different cognitive and sensory levels. This is by no means necessarily tied to the intention of the author or creator; in fact, Sontag specifically comments that 'a few of the films of Bergman [...] still triumph over the pretentious intentions of their director' (Sontag, 1967: 11). Interpretation, then, in the style of Marxist and Freudian analysis, 'digs 'behind' [...] that] manifest content [...] to find the true meaning – the latent content – beneath' (Sontag, 1967: 7). The concept of a 'true meaning', as with Barthes' critique, always looks towards the author as the origin of that meaning, backwards rather than forwards, to Bergman's 'pretentious intentions' rather than the potential sublimity of his films.

In the context of computer-generated fiction, this style of interpretation might be equivalent to analysing textual output in relation to the algorithm and the corpus behind the creation of the text – digging around in datasets in an attempt to probe

the digital counterpart to the author's unconscious. This makes for a perfectly valid approach: a practical expression of distant reading, in a way. Oscar Sharp, the director in charge of turning the computer-generated script of *Sunspring* into a short film, has called the result of the AI's training and text generation an 'amazing funhouse mirror to hold up to various bodies of cultural content and reflect what they are' (Newitz, 2016). Simply in terms of statistical representation, the script to *Sunspring* might reveal that anxiety over not knowing is a major theme in the films included in the corpus that the algorithm was trained on. It also indicates that those science fiction films are more about human interaction and introspection than about starships and space exploration. Yet, this is more quantitative corpus analysis than literary criticism – as an approach to interpretation, it is just as limiting as overtly looking for authorial intent. And of course, it is safe in the same way, too. Data, like the singular authorial meaning, implies truth and a mental escape hatch to pass responsibility over an interpretation on to someone or something else – essentially one of Foucault's author-functions.

While Sontag bemoans the 'armies of interpreters' that descend 'like leeches' particularly on literary works (8), curiously, computer-generated texts so far do not seem to attract such hosts of literary critics at all. Since they usually do not present a clear, coherent surface meaning, and since the lack of an author means that there is by definition no greater authority to contradict an interpretation, computer-generated texts should be ideally suited to any amount of interpretation. Why, then, do they still seem to be largely shunned by readers, especially academic ones, when it comes to accepting them as literature?

Sontag might see in this a successful 'flight from interpretation' into 'non-art' (10); algorithmic text production could in that sense be linked to abstract painting, which in Sontag's view has no content, and 'since there is no content, there can be no interpretation' (10). This is an important analogy and a very contestable stance. For one, if Sontag holds such abstract works with no apparent content to be uninterpretable, she is underestimating the stubbornness of determined critics. But more importantly, it is precisely this non-apparentness of content and meaning that should elicit interpretation as a means of elucidating what is not superficially, instantly visible.

Another possible explanation for the lack of interpretations of computer-generated texts could be the category of 'programmatic avant-gardism – which has meant, mostly, experiments with form at the expense of content' (Sontag, 1967: 11). This certainly applies to an extent, though not in the sense of experimenting with literary form; algorithms are far less capable of breaking patterns than humans (see Lau et al., 2018: 9). That computer-generated texts represent an experiment with the form of text production, however, is undeniable.

Yet this alone should not stall attempts at interpretation. If anything, an authorless text should reasonably facilitate interpretation. As Barthes puts it, 'to write is to reach, through a preliminary impersonality [...] that point where not 'I' but only language functions, 'performs' [...] suppressing the author in favour of writing (and thereby restoring, as we shall see, the reader's place)' (Barthes, 1989: 50). If there is no author to be suppressed to begin with, then 'the reader's place' should be easily accessible.

A further criticism that Sontag launches against interpretation is that it 'tames art' and 'makes [it] manageable, comfortable' (8). Arguably, this is entirely inverted when it comes to computer-generated texts. In the absence of an author, actual, presumed, implied, or claimed, interpretation can become highly uncomfortable. Rather than interpretation 'expressing their lack of response to what is there' on the page or screen, readers would actually need to do just the opposite of that – express their response directly or indirectly – in order to formulate an adequate interpretation of a computer-generated text.

It is at this point that Sontag's criticism and proposed approach and the possible interpretation of computer-generated texts align. One of the issues with applying 'The Death of the Author' to computer-generated texts is that Barthes, while proclaiming the birth of the reader, also depersonalises that same reader as 'a man without history, without biography, without psychology' (Barthes, 1989: 54); and if we assume that there is no human writer and no truly human reader, then this becomes not a post- but an anti-human sentiment. Because computer-generated text is such a posthuman mode of literature production, interpretation actually requires a more consciously human element on the reader's part. This does not necessitate sacrificing scientific objectivity for the sake of a purely sensory description, but includes the

need for a reasoned, that is verifiable and transparent, subjectivity, lest it become a Rorschach-test of the reader's psyche, and a greater burden of responsibility of the reader over his or her own interpretation.

These posthuman aspects are not only significant for the mode of text production, but also seem to come into play with regards to content. In 2018, a group of computer scientists, Jey Han Lau, Trevor Cohn, Timothy Baldwin, Julian Brooke, and an English professor, Adam Hammond, published a paper on their poetry-generating Artificial Intelligence. They trained the system on over 2,600 sonnets (Lau et al., 2018: 2) to establish models for language, meter, and rhyming words, and then had it generate quatrains of iambic pentameter according to their specified rhyme scheme (Lau et al., 2018: 6). Though the generation procedure itself was, of course, improved over time, there was no human editing of the textual output itself. In the subsequent evaluation of the output (Lau et al., 2018: 8), while volunteer laymen were frequently unable to distinguish between human and algorithmically produced poetry,⁶ English literature professor Adam Hammond still found notable differences. Though the form of a sonnet – rhyme scheme and meter – can be followed precisely by the algorithm, one of the problems that Hammond cited with the computer-generated lines was their 'lower emotional impact and readability' (Lau et al., 2018: 9). From this, the researchers draw the conclusion that 'future research should look beyond forms, towards the substance of good poetry' (Lau et al., 2018: 9).

Substituting perhaps 'human' or 'human-like poetry' for the value-judgment of 'good poetry' here, the outcome remains the same, and remains clear: emotional impact matters when it comes to reading and interpreting texts. As Sontag puts it, '[i]nterpretation takes the sensory experience of the work of art for granted, and proceeds from there. This cannot be taken for granted, now. [...] What is important now is to recover our senses. We must learn to see more, to hear more, to feel more' (13–14). However, we cannot entirely dismiss content or cognitive investigation

⁶ Further testing indicated that non-experts went primarily by rhyme scheme rather than meter or content (Lau et al., 2018: 9); from this, it can be surmised that they expect a lower degree of readability from poetry to begin with.

either. Unlike the idea of art that Sontag describes, computer-generated text still needs to be 'assimilate[d ...] into Thought [... and] into Culture' (Sontag, 1967: 13), to strengthen the human components of these posthuman texts on the recipient's side through a combination of emotional resonance and rational analysis. Reflection on the sensory experience can be part of the process of interpretation, as a basis for or interwoven with scholarly explication and cognitive understanding: not just 'show how it is what it is', but also 'what it means' (Sontag, 1967: 14), or rather, what it can mean, since one interpretation is not by itself superior to another.

4. Analysing a Computer-Generated Text

How different such coexisting interpretations of the same text can be might be best illustrated by a brief analysis of *Sunspring* contrasted with the interpretation presented by the filmed version. The movie casts three instead of four actors, assigning the lines of the closing dialogue of T to Elisabeth Gray, the same actress who plays H2. By casting H2 as a woman and making H much smaller and decidedly more awkward than C, a sense of resentment from H regarding the suggested romantic relationship between H2 and C emerges. That effect – as well as the implication of the relationship – is mainly rooted in nonverbal and non-scripted communication and interaction. The opening line of selling blood also takes on more significant weight as the bag that H pulls from his backpack in the end before he starts to cry is a filled blood bag, though clearly delivered too late.

While reading the script without the visual assistance of the filmed interpretation, a strong sense of confusion and disconnect emerges. Characters don't always seem to respond to one another, sentences are not always semantically coherent – such as 'The principle is completely constructed for the same time' (Benjamin and Goodwin, 2016: 2) – two characters are named H (separated into H and H2 by the filming crew), and there is no apparently identifiable plot.

This feeling of confusion and displacement need not be a deterrent, however, but rather can serve as a form which supports content. A closer inspection of the script reveals H as the central characters. The other characters never seem to speak to one another; C and H2 only ever speak to H, while T's lines might either be a separate

monologue, or spoken by the man on the roof that H goes to protect (4) and thus also addressed to H. Not only do they not directly speak to one another, they might not even be able to interact. When H asks 'Then what?', H2 says 'There's no answer'. C frowns, though it is unclear whether that is in reaction to H's question or H2's response, and replies 'We're going to see the money' (1). While H and H2 have their extended dialogue, C does not speak at all (2–3). It is not until H gets angry, starts to shake, and states that '[i]t may never be forgiven' (3) that C speaks again. This, in conjunction with the fact that both characters are named H, indicates that they might, in fact, not be separate characters at all, but aspects of the same character, a split personality which would also include T.

With this in mind, a sense of a repressed traumatic experience emerges. The script is full of references to regret and retrospection, such as '[...] shut up. I was the one who was going to be a hundred years old' (2), 'I think I could have been my life. [...] It may never be forgiven, but that is just too bad' (3), or 'I was the one that got on this rock with a child and then I left the other two' (3). Even stronger than that is the theme of not knowing and not understanding, with phrases like 'I don't know anything about any of this' (1), 'What do you mean?' (2), 'I don't know what you're talking about' (2), and 'I just have to ask you to explain to me what you say' (3) appearing repeatedly on every page except for the fourth and final one.

The absence of uncertainty on the final page is also revealing. After the disappearance of H2, H speaks his longest lines, and seems much surer and more determined than before. This leads to an extended stage direction, during which H appears to consider suicide: 'He cuts the shotgun from the edge of the room and puts it in his mouth' (4). Before this act can be carried out, however, H looks through 'a black hole in the floor leading to the man on the roof' (4), who might likewise be an incarnation of H pondering suicide in a different way. This causes H to abandon the shotgun and join the man – himself – on the roof to protect him. An interpretation which presupposes multiple versions of H also explains the seemingly contradictory stage direction at this point: 'He is standing in the stars and sitting on the floor' (4). If H was H2 all along, then having two of him now is coherent with the rest of the script.

T – who might be another part of H or might be speaking about H – transmits a latent sense of inadequacy particularly about a third person ‘he’: ‘He was like a baby [...] He couldn’t come any more [...] he was weak [...] He was a little late [...] I was much better than he did’ (4). This resonates with some of H’s earlier utterances: ‘In a future with mass unemployment, young people are forced to sell blood. That’s the first thing I can do’ (1) and, later, addressing H2: ‘You don’t have to be a doctor’ (2). In this scenario, unsettlingly, a likely role for C is that of a surgeon about to take blood or potentially other body parts from H for money. It is C who says that ‘[w]e’re going to see the money’ (1) and ‘I think you can still be back on the table’ (3), which could also refer to an operating table.

We can further gather from the dialogue that H ‘got on this rock with a child’ and left two others behind. H thus might be a young father who went to look for his fortune on a different planet, possibly even leaving some members of his family behind, but now, facing mass unemployment, fears his inadequacy and his inability to provide for his child. This crisis causes him to retreat into his inner self and look unsuccessfully for answers in a dialogue with another aspect of himself. The frustration of this conversation finally appears to lead to clarity. While the only way out, to be ‘free of the world’ (3) at first appears to be suicide, H ultimately finds a way to save and protect himself. The crisis is not fully averted, but finally a degree of reflection and stability seems to be restored. Here, a very posthuman sense of identity diffusion and displacement is utilised to explore a deeply human anxiety. In this interpretation, the text lends itself to a Freudian reading.

Some of these themes – those of inadequacy, for instance – were also highlighted in the interpretation realised by the filmed version, while others of course are very divergent. Neither of these two interpretations is more valid than the other. Even if we wanted to, there would be no one to ask. The algorithm might be prompted with a question, though that would likely require further interpretation – and, of course, considering the mechanics of the programme, the value and veracity even of a very direct answer would be more intuitively questionable than an author’s response to a similar question in an interview. In this context, there is no single truth anymore,

which ultimately frees the text up for interpretation beyond the concept presented in 'The Death of the Author'. As long as they are comprehensibly argued, different analyses can coexist with equal validity without the looming spectre of the departed author.

5. Conclusion

Ongoing developments in machine learning and algorithmic text production continue to confirm the growing significance of these new forms. With its increased capabilities, GPT-3, a deep learning language model published in 2020, has led to a proliferation of computer-generated material and has raised media attention once more. Human intervention, such as editing or selecting the final output, is still present but decreasing, and algorithmically generated texts are ever less distinguishable from human writing. Meanwhile, the technology is so accessible, and the results are so readable, that computer-generated text is decidedly not niche anymore. In addition to their growing cultural impact, these forms point towards very interesting, real and practical problems in the application of literary theory. Who is the author of a text, and what function do they fulfil for readers? What role does interpretation play, and how can we approach it? At what point does meaning enter a text?

Computer-generated text is fundamentally posthuman, and by virtue of the humanity contained in this posthumanness, and not least because of its human readers, it has the potential for being literature. If literary criticism can accept that, we can productively adapt existing theories and approaches to work in this new context. This is necessary, because as posthuman texts, they essentially require interpretation; this mode of production is by no means a 'flight from interpretation' (Sontag, 1967: 10), but rather a form that we can only truly grasp when we begin to interpret it.

A way forward might be to recognise that interpretation has always included a considerable amount of meaning-making on the side of the interpreter, that is, the reader. And just as form and content are inseparable, cognitive interpretation cannot be isolated from subjective reception, no matter how strictly rational we try to make it. These are factors that Sontag, and to an extent Barthes, identified and pointed out fifty years ago, but that new modes such as computer-generated texts are now forcing us to recognise and take into account. This article has not provided concrete

answers, but hopefully has raised questions which literary critics can continue to explore as textual production and reception evolve, to reflect on how we read and interpret not just computer-generated texts, but literature in general.

Acknowledgements

Many thanks to the editors of this volume, Julia Hoydis and Roman Bartosch, as well as the reviewers for their helpful and very attentive comments; Stefan Herbrechter for a constructive conversation in the early days; and Heiko Jakubzik for listening to and debating this idea with me from the very start, and for inspiring me to look down a number of roads otherwise not taken.

Competing Interests

The author has no competing interests to declare.

References

- Barthes, R** 1989 The Death of the Author. In: *The Rustle of Language*. Berkeley: Univ. of California Press. pp. 49–55.
- Benjamin and Goodwin, R** 2016 *Sunspring*. Available at http://www.thereforefilms.com/uploads/6/5/1/0/6510220/sunspring_final.pdf [Last accessed 4 January 2019].
- Botnik Studios** 2017 We used predictive keyboards trained on all seven books to ghostwrite this spellbinding new Harry Potter chapter [*Twitter*]. 12 December. Available at <https://twitter.com/botnikstudios/status/940627812259696643> [Last accessed 23 September 2018].
- Botnik Studios** 2018a Harry Potter and the Portrait of what Looked Like a Large Pile of Ash. *Botnik*. Available at <http://botnik.org/content/harry-potter.html> [Last accessed 23 September 2018].
- Botnik Studios** 2018b *Predictive Writer – HP Narration*. Available at <http://botnik.org/apps/writer/?source=d08198a9a936f791b7ffe144a2e9b1e3,0e155979285771266d520c44607722a1> [Last accessed 23 September 2018].
- Boudreau, J** 1993 A Romance Novel With Byte: Author Teams Ups With Computer to Write Book in Steamy Style of Jacqueline Susann. *Los Angeles Times*, 11 August. Available at http://articles.latimes.com/1993-08-11/news/vw-22645_1_jacqueline-susann [Last accessed 4 January 2019].

- Bruner, R** 2017 Here's What Happens If You Let a Whimsical Robot Write the Next Harry Potter Chapter. *Time Magazine*, 13 December. Available at <http://time.com/5062514/harry-potter-robot/> [Last accessed 23 September 2018].
- Butler, S** 2011 *The Matter of the Page*. Madison: University of Wisconsin Press.
- Chicago Tribune** 1991 Computer Turns Page on Rules of Writing Fiction, 18 March. p. 3.
- Cooper, G F** 2017 Harry Potter chapter written by bots is magically terrible. *CNET*, 12 December. Available at <https://www.cnet.com/news/harry-potter-new-chapter-predictive-text-botnik/> [Last accessed 4 January 2019].
- Flood, A** 2017 "He began to eat Hermione's family": bot tries to write Harry Potter book – and fails in magic ways. *The Guardian*, 13 December. Available at <https://www.theguardian.com/books/booksblog/2017/dec/13/harry-potter-botnik-jk-rowling> [Last accessed 4 January 2019].
- Foucault, M** 1979 What is an Author?. Translated by D. F. Bouchard and Sherry Simon. *Screen*, 20(1): 13–34. DOI: <https://doi.org/10.1093/screen/20.1.13>
- Grover, D L, King, M T, and Kushler, C A** 1998 'Reduced Keyboard Disambiguating Computer'. Available at https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=19981006&DB=&locale=en_EP&CC=US&NR=5818437A&KC=A&ND=1 [Last accessed 5 January 2019].
- Haraway, D J** 1991 A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century. In: *Simians, cyborgs, and women: the reinvention of nature*. New York: Routledge. pp. 149–181.
- Hayles, N K** 2008 *Electronic literature: new horizons for the literary*. Notre Dame: University of Notre Dame.
- Jameson, F** 1991 *Postmodernism, or, The Cultural Logic of Late Capitalism*. Durham: Duke University Press. DOI: <https://doi.org/10.1215/9780822378419>
- Lau, J H et al.** 2018 Deep-speare: A Joint Neural Model of Poetic Language, Meter and Rhyme, *arXiv:1807.03491 [cs]*. Available at <http://arxiv.org/abs/1807.03491> [Last accessed 4 January 2019].
- Lohr, S** 1993 The Media Business: Encountering The Digital Age – An occasional look at computers in everyday life; Potboiler Springs From Computer's Loins. *The New York Times*, 2 July. p. 1.

- Maggio, E** 2017 There is a new chapter in Harry Potter's story – and it was written by artificial intelligence. *Business Insider*, 13 December. Available at <https://www.businessinsider.de/there-is-a-new-chapter-in-harry-potters-story-and-it-was-written-by-artificial-intelligence-2017-12> [Last accessed 4 January 2019].
- McCall, R** 2017 AI Attempts To Write Harry Potter And It Goes hilariously Wrong. *IFLScience*, 14 December. Available at <https://www.iflscience.com/technology/ai-attempts-to-write-harry-potter-and-it-goes-hilariously-wrong/> [Last accessed 4 January 2019].
- Newitz, A** 2016 Movie written by algorithm turns out to be hilarious and intense. *Ars Technica*, 9 June. Available at <https://arstechnica.com/gaming/2016/06/an-ai-wrote-this-movie-and-its-strangely-moving/> [Last accessed 6 January 2019].
- Sharp, O** 2016 *Sunspring*. Available at <https://www.youtube.com/watch?v=LY7x2lhqjmc> [Last accessed 7 January 2019].
- Sontag, S** 1967 Against Interpretation. In: *Against Interpretation and Other Essays*. London: Eyre & Spottiswoode. pp. 3–14.
- Towsen, N** 2017. It's not automated! We have a team of writers who all use the Botnik predictive text keyboard. We trained keyboards on all 7 books and had a big writing jam. Then I took the best pieces of copy, arranged them into a narrative, and wrote some copy to fill in the gaps [*Twitter*]. 12 December. Available at <https://twitter.com/NatTowsen/status/940652654925111296> [Last accessed 23 September 2018].

How to cite this article: Elstermann, A 2020 Computer-Generated Text as a Posthuman Mode of Literature Production. *Open Library of Humanities*, 6(2): 31, pp.1–23. DOI: <https://doi.org/10.16995/olh.627>

Published: 23 December 2020

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Open Library of Humanities is a peer-reviewed open access journal published by Open Library of Humanities.

