

Investigating attitudes of professional writers to GPT text generation AI based creative support tools

Geoff Davis and Mick Grierson

UAL Creative Computing Institute, 45 Peckham Rd, Camberwell, London SE5 8UF.

Email: g.davis1020191@arts.ac.uk

Director of Studies: Professor Mick Grierson

Abstract:	4
Introduction.....	5
Research questions.....	8
Co-created or hybrid human-AI systems	9
Stakeholders and Ethics	14
Historical background of computers and literature.....	16
Personal background.....	21
Creativity Support Tool research	22
Creativity Support Tool use in the study	23
Contemporary and future uses of text generators	23
Materials and Methods.....	26
Sentiment Analysis	27
Method.....	27
Design of the experimental writing tool - initial investigations	28
The Feedback Answers	28
Experimental Ethics	29
Final design.....	30
Design Principles	30
Three Writing Experiments.....	32
Results - Processing, Statistics, Charting and Presentation	33
1 Respondents	33

2 Study analysis	41
3 Feedback questions after the experiments	50
4 Creative and Non-Creative Occupations	56
5 Sentiment Analysis	62
6 Feedback questions after the three experiments	83
7 Summary of findings.....	100
Regarding the generator	104
Aesthetics.....	107
Conclusion	109
Declaration of Interest Statement.....	113
References.....	113
Appendices.....	128
Further reading on creative writing and text generation	128
Credits	131
Feedback comments.....	132
Experimental Design Overview	159
Experiment Online Design.....	164

Abstract:

This research addresses creative and ethical issues for professional and amateur writers when making AI hybrid or co-created works. We present a mixed-methods study, exploring how these writers used generated texts in a new online creative writing support tool. The study consisted of three creative writing tasks of increasing complexity. Each started with generated text which was then further edited to create a finished work. Feedback on their creative process was gathered as ratings and personal responses, alongside analytics and on-screen activity monitoring.

Invited participants were professional and serious amateur writers, including journalists, academics, fiction and non-fiction authors, poets, artists and scientists. By having a wide range of adult respondents we aim to provide in-depth insight into the use of generated text by professional writers. We report on the differences between creative or artistic writers and less creative, analytical or copy writers. Respondent feedback on the custom AI text generator and editor, and their ideas on the integration of text generation with word processors, provide suggestions and paths for future creative support tools. We also gathered insights into the perception of 'fake' news and their relationship to AI generation systems.

Keywords

Human-computer collaboration; creativity; co-creativity; writing support; text generation; writing tools; creativity support tools; natural language processing; future of jobs; empowerment; automation; generative pre-trained transformer; AI, HCI; NLP; CST; GPT, PSG.

Introduction

This research addresses several issues that arise when using text generation technology to help in a range of common but dissimilar writing tasks such as captions, news descriptions and fiction. The focus is on using generated text in custom computer software to help with both creativity and editing. The new online software is designed to be simple to understand and easy to use.

Generated texts are commonly provided in computer systems such as email auto-complete, online product recommendations, and auto-generated data driven news items such as financial reports or sports results. Computer generated books have been produced for over ten years (Cohen, 2008). The software Quill from Narrative Science has been used to pad financial reports since 2010 (Quill, 2021). The Wordsmith generator from Automated Insights has been in use since 2014 (Peiser, 2019). Bots and text spinners have been used to generate content for various purposes since the early days of the web. Early examples include ArticleBot (Bailey, 2018). Experimental electronic poetry and prose is an even older field, with early examples of computer generated poetry showing that language manipulation was one of the core research activities on limited capacity computers in the early days of commercial computing after World War 2 (Roberts, 2017).

Modern text output systems have been labelled as ‘artificial intelligence’ or AI although they use statistical machine learning. The term intelligence is used as the system can mimic human intelligence, in the same way that a chess playing machine mimics a human player. These methods have rapidly progressed rapidly, both in complexity and accuracy. It is worth noting that advances in machine learning mean that AlphaZero, the current top Go board game playing software, is not trained on data from human games, but generates game play from the layout and rules only (Silver et al., 2017). This is in contrast to text generation systems such as OpenAI’s GPT-3 and others, which produce new text by mathematical analysis of large

volumes of human-written text, predicting the probable next word from memorisation. In this manner text generation can appear very realistic, and even convince people there is an AI author with a voice or self-awareness, since the training data was originally written by human authors, freely taken from the internet. These people, without knowing it or being paid, have supplied the myriad of texts that are conflated in the output (Brown et al., 2020).

Newer generation systems such as OpenAI's GPT-3 and GPT-J can use 'few shot learning' where pre-training of a large text dataset allows implicit learning of many tasks, so the model can generalize to new tasks, with only a few examples. This means in practice that examples of desired outcomes in the model's input text prompt allow it to 'learn' rapidly and generate the required outcome on new material (Schmid, 2021). This is also known as 'prompt programming' (Reynolds & McDonell, 2021). Otsuki (2020) provides a general overview of the use of generated text by students and academics.

One aspect of the current near-human quality of output is that even researchers anthropomorphise, ascribing intelligence or even consciousness to the syntax of the machines, which is entirely generated from accumulated human writing. All of the intelligence and wit in the generated text is gathered from the huge quantity of human text on which it learns. This is why 'big is better' in terms of training data and the model type.

Generated texts also depend on human perception for their meaning, hence the tendency to anthropomorphise.

In this Paper we do not discuss artificial general intelligence. It is worth noting that the famous Turing test or 'imitation game' (can a computer language system be distinguished in conversation from that of a human?) is a test of mimicry not a test of any internal consciousness or awareness of a computational machine.

Most large computer technology companies invest heavily in natural language processing (NLP). Microsoft and Nvidia have recently produced a very large Turing-NLG model, which then wrote a summary about itself (Microsoft, 2020).

During the course of 2020 I used OpenAI's Generative Pre-trained Transformer GPT-2 system (OpenAI, 2019). This is now superseded by the much larger GPT-3 system (OpenAI GPT-3, 2020). A Transformer is a machine learning model introduced in 2017, used primarily in the field of NLP and non-natural text such as programming code. They can also be used to generate images from descriptions. Like recurrent neural networks (RNN), Transformers are designed to handle sequential data, including natural language, for tasks such as translation, text summarisation and generation. For a diagrammatic explanation see Alammari (2018).

Newer systems are more powerful. However, this does not affect the relevance of this study as the GPT-2 we used is already considerably advanced for the purposes of research. Shane (2020) has discussed this issue in relation to her open-ended investigation of generated recipes, crochet patterns and fan fiction, etc. "Had I seen [GPT-3's realistic output] in 2016 instead of [Karpathy's] char RNN's quirky recipes I might not have gone down the path of neural net creativity", (Shane, 2020). For the purposes of this creative writing study the older Text Synth GPT-2 system, which regularly has unusual outputs (inappropriate words or non-sequiturs), was used. It was considered more stimulating for the writers than the more conventional prose generated by later generation models. Variety of output can be controlled with run-time parameters but we did not include this aspect in this study.

We used GPT-2 Text Synth defaults for initial settings. At the time of the study this was the largest GPT-2 model (1.6 billion parameters) publicly released by OpenAI (Bellard, 2021).

Text Synth now (2020) uses GPT-J 6B and GPT NeoX 20B, the largest public models (6 billion and 20 billion parameters).

Previous research specifically about fiction writing includes a study of four professional novelists using a generator, which concluded that writing tools should provide many suggestions, be responsive and have easy text regeneration (Calderwell et al., 2019). Other research devised a new poetry writing system using procedural language processing, and tested it on three poets in a user-centric study to gauge their reactions (Gero & Chilton, 2019). This found that the generated sentences were of use to the poets in divergent ways, and that ownership of the text was a concern.

Research questions

How do professional writers respond to computer text generation tools? We used a new text generation and editing system to gauge their responses.

In order to better understand what the concerns of stakeholders were in regard to computer generated texts, the experiment was designed to open up the topic rather than focus on one aspect. The intention was to have a broad range of mostly professional writers doing several generated writing tasks, then ask a variety of questions, while also gathering quantitative data such as scaled responses and timings. As text generation is quite different to traditional writing, there was a focus on emotional responses. In this way, we followed grounded theory in order to discover the main themes and concerns from the human voices and the observational data.

We also examined the data to find any differences between creative and non-creative writers. Other areas investigated included the practical use of text generation in a creative writing tool (CST) or word processor, opinions on ownership and plagiarism and the relationship of generated text to 'fake' or false news and content.

Co-created or hybrid human-AI systems

Working with computer generated text is on a spectrum with using grammar and style checkers, and text editors. Existing systems such as word processors and apps like Grammarly already co-create with authors. Technology has had a democratising or popularising effect on writing, if writing includes all text such as email, messaging, and memes. Since this electronic production also includes free fiction, and the cost of distribution can be low or zero on the internet, it is also devaluing writing from skilled and specialised, to deskilled and generalised. Even basic skills like spelling and grammar are no longer required in a democratising side-effect. This might sound alarming to academia but it helps with accessibility for second-language users, who constitute the majority of global English users. We are aware of the focus on English language systems in our research. It is worth noting there are many non-English language generation systems now in use.

The issue of what exactly constitutes hybrid or co-creative works is not directly addressed in this study. As the writers used the experimental system for the creative tasks, many concerns and ideas were raised and seriously discussed in a natural and unselfconscious way. The raw data is available in the Appendix (Feedback comments).

The experience of a reading a human written text produces a mental image of the whole person ‘as if’ they were speaking the words to the recipient. This may be beneficial as perceptual enhancement rather than ‘only’ an empathetic effect: ‘imagined visual patterns interact with a concurrent perceptual stimulus to boost sensory performance.’ (Pearson et al., 2015). If there is no person doing the communication, the generated writing, even if well constructed, produces the infamous uncanny valley effect, or will just come across as impersonal copy (Mori, 1970; Cherry, 2020). However if generated text has an authorial

voice, it can convince even experienced researchers that there is ‘someone’ doing the writing, or a consciousness within the software. This personification effect is used during perception, so it is hard to suppress. An experienced AI and literature researcher, Gwern Branwen, stated “chatting with GPT-3 feels uncannily like chatting with a human”. Samhan Salahuddin, software engineer and researcher, commented that “its [GPT-3] output is often profound and moving and better than most of my writing,” (Branwen, 2020) (Salahuddin, 2021). Whether people will accept that a machine can converse more intelligently than oneself, superhumanly, is moot. Already search is accepted as an impersonal but superhuman memory. We have not discussed applications of any future text and language systems in robots or androids, or as ambient communicators. This field is constantly improving and enlarging.

There are different ways a generative system can be used in creative writing compared to copy or report writing. Generally the use of a computer in creative works is seen as devaluing, akin to copying or relying on a machine (Danai, 2018). For instance, few authors (creative writers) have discussed their use of word processors, and how it changed their writing. Machine writing is judged to be unconnected to true creative writing. This is particularly true of fiction or biographical writing, and other personal communications such as travel or investigative writing, which is experienced as direct communication from one person (the author) to another (the reader). Often the ‘voice’ or style and tone of an author is enough to capture a readers’ committed attention, causing them to read everything by that author. In a more general discussion, Hertzmann (2020) has examined how art is produced culturally and its evolutionary function, and the false attribution of ‘artist’ or ‘artistic’ to the productions of machines, even ones that might pass the Turing test. In his view artificial intelligence is only ever an assistant or tool.

This tool-only approach would also apply to any AI ‘superhuman’ productions, purely through ownership of the software (however defined). This applies even if the output is superior to the writer’s own work, in the same way architects or engineers do not feel their design software owns their professional output.

Recent discussions of generated text outputs range between assuming there is an actual AI super-being communicating profound thoughts, to more knowledgeable comment on the output being like that from a ‘team of automated googlers’ used in a ‘modern content farm’ (Hooande, 2021). The models are trained on Google-visible internet text, not for instance, language as spoken in social contexts. The question is ‘what would you do with a team of hundreds of people who can solve arbitrary problems only by googling them?’ With this perspective, because all of the outputs are essentially second-hand, there are no new applications, no ‘killer apps’. The recent text systems can only provide improvements to existing use cases such as translation, search, summarisation and categorisation. These are very useful improvements and justify the production expense. Their use in creative work is more for artists, researchers and experimenters.

Any superhuman writing would be owned by a human originator or owner, and used for their own motivations and aims. If volume is a superhuman characteristic, then the rapid production large quantities of specialised, uniquely personal advertising copy represents the AI singularity. But this is never claimed. In contrast, a short computer generated poem will elicit human reactions of admiration for the imagined sentient AI author. Even if it created by the same aforementioned super-powered Google content farm.

This superhuman overproduction for a writer or artist, is generated from a text prompt, which is an artistic input. So the prompt is the art. The stream of personally tuned stories appears as

the artist watches the screen. This ability to produce endless narcissistic (as in a mirror) text and images is a challenge for artists. “It feels almost wrong to just pick a single thing [that the program produces]. Because, yes, it can create a lot of images, but it’s more magical to see it at work,” (computer artist Klingemann, quoted by Vincent, 2019). This could be rephrased as ‘it is magical to see it at work creating things like the things I create’. Using a generating system has its own allure, in a setup that removes not only a publisher but also an external audience. It is like creative writing without the fear of the blank page, effortless production in one’s own style.

This is not so much co-creation of a final artefact, as experiencing a guided text, similar to a non-linear narrative or hypertext fiction. The meaning is generated during the process, and does not need to be recorded. This is a form of ‘casual’ creativity, which might become an alternative approach to AI in the arts. Rather than co-creating more work to be published or viewed in galleries, the generated text or art becomes an interactive narrative to be read as it appears. This would enable readers to be less critical of changes in the overall form (experienced as a lack of plot and character arcs) since no overall semantic structure is needed. The enjoyment would become continuous or episodic. This has cultural reverberations. Narrative is locally produced. For instance, Asian film structures can be baffling to Western audiences due to Asian story forms and their religious or philosophical origins, compared to Western historicised narratives (Pretzelzetzel, 2020). Western narrative is usually based around the struggles of aristocracies, even a existential or superheroic aristocracy of one, and are based on a rise to power, overcoming and succeeding, with usually a rebirth to delay the climax. These arcs have to be added to modern story generation by the author for it to ‘make sense’ to a traditional Western reader. Huge efforts were once made by computer researchers to produce ‘literary machines’ using multilayered ‘expert systems’ to compete with the great authors. One such was BRUTUS (Bringsjord and Ferrucci, 1999).

Text generation can be useful to alleviate writers' block, for ideas generation, and as a source of unexpected new directions, as many of the respondents in this study mentioned. Others use it for simple content, for instance to generate many character biographies in interactive games.

Online discussion comments include '...it doesn't need to make much sense. Apparently people who read it don't mind' to which the reply was 'people don't read it, search engines do'. This alludes to the use of text generation for content which is only used in advertising dominated web pages for the gaming of search results. This used to be called 'spinning' content, prior to the new generators. A proportion of the web-scraped AI training data will be generated or spun second generation text.

There is an assumption that writing of all types, but particularly fiction, requires creativity, rather than just the rearrangement of existing ideas, words and phrases.

"The best-selling thriller writer Lee Child was asked whether his Jack Reacher novels could be written by a machine. 'Logically, it has to be possible,' he replied. 'Whenever anyone asks me 'where do my ideas come from', I always think, 'from reading','" (Higgs, 2019). The AI field started with computer vision systems modelled on a brain's receptors and nerves, with a machine called a perceptron (for a summary, see Golden, 2001). This was invented in 1943 and first appeared in hardware in 1957 built by Franck Rosenblatt (Perceptron, 2022). The current machine learning text systems are extremely developed and enlarged versions of this initial idea. Thus 'ideas come from reading [using a brain]' could easily be 'computer generated ideas come from reading in the form of training [using a neural net machine learning system].'

Experimental writers such as William Burroughs have used cut-up techniques to rearrange their own and found texts in an effort to extract implicit content and to discover hidden

meanings (Burroughs, 1963). This is a form of anthropomorphism, searching for personally appealing patterns in randomised words. This is related to cherry-picking generated texts from one's own prompts. Even if the generators produce original fiction that might pass a literary Turing Test, there is no interest as no human origin story. Literary writing has a personality and represents a biography, from celebrated (high class) to celebrity (low class). Dickens, Kafka, Eliot, Woolf, Lawrence, Kerouac, Burroughs, Winterson and many others are presented as stylistic breakthroughs but are emotionally experienced as biographical breakthroughs to inspire different audiences.

Removing the work and talent from the production of art removes the humanity, quite apart from any authorial confusions. However this might not be a concern for 'digital natives', or people that have grown up with integrated always-on computer services, who do not see themselves as outside the co-creation system but part of it (Taylor, 2014). Younger 'digital native' audiences are less concerned with narrative conventions (plot) or originators (authenticity) and can enjoy memes, short form video, comics, degen art, and so on. It is as if the short attention span required or conditioned by the smartphone has prepared the audience for new art forms. Perhaps the main requirement for enjoying ambient AI art is reduced expectations. There is also an issue around scarcity. Great art implies great prices. Even books used to be relatively expensive or obscure and hard to get, and so became exclusive. Now everything is free so there is much less intellectual and emotional investment in cultural offerings.

Stakeholders and Ethics

The respondents are people involved in writing, or people using creativity support tools for writing, and so will be directly affected by advances in computer text generation. This might

mean additional functions in their preferred writing software, such as an ideas generator (alternative text) or continuation generator (similar text).

The writers are from a range of professions, from novelists to journalists, scientists to artists, covering both creative (artistic) and non-creative (report or copy) writing.

There are ethical and political concerns concerning the incursion of powerful and flexible new generative tools into professional areas previously considered uniquely human.

Increasingly copy and content dominate the news, traditional reporting replaced with profitable advertorials and edutainment, or biased or invented news for niche groups. Modern generative tools can produce this type of text on demand and at far lower cost than humans.

The latest text generators are trained on vast amounts of free non-copyright human-written text from the internet, so this is an economic and ethical issue before anyone is put out of a writing job.

Already students and copywriters use generators to speed up the production of essays and reports, as time is of the essence in these roles. There is little academic study into this, despite it happening under the noses of mark-setting tutors. Some studies have explored co-creativity with novelists and other writers, such as those by Calderwell (2019) and Gero and Clinton (2019). These experiments are usually on the lines of productivity improvements rather than any broader societal questions.

The deployment of the new AI text systems might become widespread and raise ethical or political concerns as such outputs will be undetectable in use, and will affect the ability to charge a premium for creative human work. See Acemoglu (2020) for a discussion of the labour market effects of skilled and unskilled job replacement by robots. Ethics apply to how these systems are used in the workplace, because the economics of the employers will push towards cost-saving where possible. Currently, each company in the field is pursuing its own approach to optimising the technology. Opinions on value range from optimistic utopian

ideas of hybrid co-creativity, with any job losses made up with new jobs involving interpersonal relations, where a human is (at the moment) preferred. More pessimistic dystopian views predominate, where humans become “like a house cat” in the care of the pervasive AI systems, as suggested by Elon Musk at a recent coding conference (Musk, 2016). For a detailed discussion of the problem of control of future AI systems see Russell (2019).

There is also a more fundamental issue of human understanding, or how a culture will exist if a governing AI’s decision making is not comprehensible to humans. “But that [current] order is now in upheaval amid a new, even more sweeping technological revolution whose ... culmination may be a world relying on machines powered by data and algorithms and ungoverned by ethical or philosophical norms. What will become of human consciousness if its own explanatory power is surpassed by AI, and societies are no longer able to interpret the world they inhabit in terms that are meaningful to them? ... What will be the impact on human cognition generally? What is the role of ethics in this process, which consists in essence of the acceleration of [incomprehensible] choices?” (Kissinger, 2018).

Historical background of computers and literature

Computer generation of literary or non-functional artistic text has a long and parallel history to that of the progress of general computing. Combinatorial literature has an even longer history:

“Samples and reinventions of writings by Optatianus Porphyrius (Carmen XXV, fourth century AD), Julius Caesar Scaliger (Poetices, 1561), Georg Philipp Harsdörffer (‘Fivefold Thought Ring of the German Language,’ seventeenth century), and other works are capably presented on the Permutations site [from Florian Cramer], illustrating how the mechanics of

contemporary (and prehistoric) digital poems have roots in works produced several centuries ago," (Johnston, 2008; Cramer, 2000).

"The illustration of the "knowledge engine" included in early editions of Gulliver's Travels is an engraving of a sketch from the notebook of Lemuel Gulliver. In other words, it is a purely fictional object. Yet, Swift's fictional invention and its graphic representations have become part of the documented historical lineage of computing machines," (Rodgers, 2017).

These early text generation systems used the available technologies, such as cranked mechanisms. Almost as soon as computers were developed to solve ballistic and nuclear explosion calculations, people began experimenting with them as creative systems.

The creators came from, or were still in, engineering or mathematical disciplines, often in military research centres since computers were developed to calculate ballistics and nuclear blasts. This led to a credibility problem for computer art, which also applies to text generation and literature (Taylor, 2014).

In 1953 Christopher Strachey generated love poems, with variation based on Roget's Thesaurus, on the Ferranti 1, the first commercially available computer designed from specification for the Manchester 1, designed by researchers at Manchester University (Roberts 2017). In 1959 the German mathematician and computer scientist Theodore Lutz developed Stochastic Texts on a Zuse Z22 computer, which used Kafka's novel The Castle for source material. He was merely demonstrating that the computer could generate meaningful text, output on a teletype printer (Rettberg 2019).

Early computer art in the early 1960s, by innovators such as Frieder Nake, was usually line or pattern based, output on plotters or other mechanical devices. Traditional artists who started using computers in the mid 1960s, such as the painter Harold Cohen, also used connected mechanical devices with computers to create art. Cohen developed AARON, a software version of his art practise, which controlled a mechanical 'turtle' to draw Cohen-style art on

floor mounted materials. There was a large artisanal element to these early pioneering works. Cohen used his machine to generate unique versions of his paintings, and sold them at 'democratic' prices in galleries, in an early example of the reduction in value of computer generated work.

The Nobel prize winning author J. M. Coetzee also created computer generated poems using an Atlas 2 supercomputer in 1965 (Roach, 2017). Coetzee used some of his generated material in published poetry. As a computer artist, I used procedural story generation in a 'Story Generator' published and exhibited by the UK's Micro Arts Group. This was later made available as a download on Prestel teletext (Davis, 1985).

Literary hypertext fiction produced on computers, usually in academia, present creative electronic texts as postmodern due to their modernist nature and fractured presentation structures. Discussion of the lack of formal poetic, rather than technical, enhancements in digital poetry led to issues of legitimacy. "Could it be true that digital poetry is, in fact, a simulation of poetry?" (Funkhouser, 2007).

Recent technical developments have produced readable human-like text outputs. The progress in neural networks and machine learning has led to many different successful approaches. This area is now the subject for historians (Lu et al., 2018). Ironically, future developments might include AI historians analysing all the world's present and past data.

The initial breakthrough of stand-alone office computers was driven by software, as people wanted to speed up and economise routine office tasks such as writing letters and notes, and book-keeping. Word processors and spreadsheets were the driver of change. Word processors did not become readily available until the 1970s, becoming mainstream in the 1980s.

Early computer generated poetry soon progressed to story generator algorithms or SGAs, often used as teaching aids (Gervas, 2012; Sharples, 1985; Ruben, 1980), which later moved

into the public software and art gallery scene (Davis, 1985). We do not explore this area of electronic text but it is an indicator of future directions for co-created works. “Even if most of the world did not notice, these [early digital] poems are important because they established the foundations of the genre,” (Funkhouser, 2007).

Early ‘expert systems’ for literary production such as BRUTUS gave unimpressive outputs (Bringsjord, Ferrucci, 1999). New neural net systems work at high speed with very large datasets and have overcome the limits of previous manual coding methods with responsive systems based on statistical results. As the size of the system increases, there is a qualitative change in the output.

The same evolutionary path applies to the adoption of modern AI systems from research or education to general use. Autocomplete, automatic spelling and grammar checking, and template-based page formatting are all in common use. The new AI text systems can generate swathes of comprehensible and unique text on any subject, which can include computer program code, mark-up languages, categorisation, summaries, translation, as well as continuation text as used in this study. There is also a new field of text to image, where a text description generates an image.

Generative Pre-trained Transformer GPT as used in the study is a deep neural net where layers of neurons (decision nodes) can be individually weighted (‘tuned’), each layer feeding results to the next layer, to give a final output. There are many types. The first was modelled on a human neuron and was called the perceptron (Golden, 2001). Improvements have come from increasing the number of parameters in the models and training on much larger data sets. For the particular system in use in the study, the development of TensorFlow was an essential step (Abadi, 2016). Also see Johnson (2020) for a typical news summary on this ‘bigger is better’ approach, which also mentions issues such as racial bias in the training data.

OpenAI's GPT-3 is trained on a dataset over 116 times larger than the previous GPT-2, 175 billion parameters against 1.5 billion. A parameter is a neural network parameter, or weight of the 'neural' connections. A study by the OpenAI research team behind GPT-3 covers many of the developments (Brown et al., 2020. Section 6, Broader Impacts). At the time of the research, GPT-3 was not released. Larger systems are now available such as Google's Switch Transformer with over a trillion parameters (Fedus, 2021), and similar from Amazon AWS, IBM Watson with Discovery and Cognitive systems plus Sierra and Summit supercomputers, and Microsoft AI at Scale initiative using supercomputing in the Azure cloud computing environment (Combs, 2021).

This is to be contrasted with adversarial networks (GANs) mainly used for image processing, where the output is matched and tested against an actual image until it passes a realism level. Goodfellow's important paper on adversarial networks transformed the field of machine learning and led to the improved systems (Goodfellow et al., 2015). Interestingly, the latest GPT-3 model can also be used to create images from text with the DALL·E trained network (OpenAI blog, 2021, 2020), although this is not publicly released yet. For the latest work in this field see Imagen (Imagen Google Research, 2022).

OpenAI's GPT-3 had not been publicly released at the time of the experiment, unlike their GPT-2. The newer systems are available in the current version of the AI editor developed for this experiment, which used GPT-2.

In the field of journalism, text generation has been used since 2014 for news reporting, or more accurately, generating readable summaries from data driven news such as sport results or financial data. Computers have been used in literature for poetry, novels, and memes using the older procedural methods as well as the recent generative ones. Electronic literature is a wide field and is represented by distinguished forums like The Electronic Literature Organisation (ELO, 2021).

Studies have found that in many areas such as summarisation, classification and similar, software is more accurate and much faster than humans. See Artificial Lawyer (2018).

Personal background

My interest in using computers with text started with combining my own fiction and generative code. From 1980 I worked as a programmer on mainframes. In 1984 I bought a Sinclair Spectrum micro computer and experimented with text generation using randomised lists of words based on my existing short stories. This was released in 1985 as 'MA4: Story Generator' by the UK's Micro Arts Group. Along with other early computer art, this Story Generator is now in the Computer Art Society's Archive. It was exhibited in 1985 at the London Film-Makers' Co-op LFMC and in 2021 by the Computer Art Archive in Leicester, UK (Davis, 1985; Davis, 2019; Clark, 2021). MA4: Story Generator is now displayed with visual work at an exhibition at the British Computer Society Moorgate London (BCS 2022). This type of text generation is Procedural Story Generation PSG and is still used in games and for pre-designed story generation purposes. It generally uses a programmed structure with files or lists of labelled words, to create a large but finite output of sentences, usually around a theme. It is very different to machine learnt generation of text, which does not apply any structure or overall grammatical procedures, instead using statistics from a very large training body of real written text to predict the next texts from a prompt text. This is known as completion, but there are many other uses for the general purpose Transformer software such as translation, classification, summarisation etc. as mentioned. PSG is an older method of text generation used in deliberately limited domains, such as game scenarios (Uotila, 2018; Fredericks, DeVries, 2021). It is popular due to the controlled output, since modern generation with completion can give unexpected results.

Spatial text art and commercial creativity software, based on a custom 3D zooming user interface or ZUI, was developed from my research Masters in Electronic Arts at Middlesex University (Davis, 2004).

Creativity Support Tool research

Creativity Support Tools (CST) are a subset of computer aids used for work. In the realm of text, a word processor is not a CST as its core functions are accepting keyboard strokes and saving files of this activity, with more recent additions such as spell and grammar checking, and some basic layout tools. More specialist programs such as Desk Top Publishing (DTP) for layout and publication, and text style assistants such as Grammarly theme and style editor, are CSTs. Older software such as IdeaFisher, previously known as Thought Office, were creativity support tools, used alongside word processors, visualisation and charting programs. Academic work in this area included Genex, an adaptation of Csikszentmihalyi's model of creative flow (Shneiderman, 1999; Csikszentmihalyi, 2014).

Modern word processors have additional features, such as reference lookup and layout design, and so can be used as simple CSTs. This unified approach is better for creativity as an unobtrusive workflow is important. In the field of AI text assistants, most are currently separate programs due to the system requirements. This is an area which will change with increasing speed and storage capacity of local computing devices, and improvements in efficiency of text generation. For instance, the Story Live system at the time of this experiment used a compressed open source GPT-2 based system, improving the performance with little loss of text quality (Story Live, 2020). Development of current commercial GPT-3 systems is limited by high development and running costs (Dickson, 2020).

Other more complex text management tools help inspire and structure fiction or script writing. Popular software includes Scrivener, Granthika, Save The Cat and Final Draft.

Many, such as the style editors Autocrit, Hemingway App, along with the more complex Granthika, have now migrated online. There is also a subset of academic CSTs designed for use in HCI experiments to gain data which are not usually developed further (Frich, 2019; Shneiderman, 2008). I developed 3D and spatial CSTs during a research Masters degree (Davis, 2003).

For respondents' feedback on using the custom CST Story Live for text generation, please see the Discussion section.

Creativity Support Tool use in the study

As an experimental tool for the study, I devised Story Live, an online creativity support tool for my practise in collaboration with Fabrice Bellard, using his Text Synth GPT-2 generator. I decided to use this online system to conduct research into professional writers in order to examine what they knew of, and how they used, the new AI text generators. This online tool was also user tested here, as relevant metrics and comments were gathered during the experiment. Story Live is now freely available as a text generator and editor with export (Bellard, 2021; Story Live, 2020).

Contemporary and future uses of text generators

Language, sometimes converted to speech, is the primary interface in voice assistants, translation, recommendation systems, score or number based news reports and so on. Apple Siri and Amazon Alexa are common AI communications systems in use for many years. The new text generators are used commercially, such as in the GPT-3 application FloWrite (2021), which effectively mimics natural language to produce business letters and other communications from a minimal prompt.

As more general business uses appear in the marketplace the novelty of the generators will

recede. These text systems have become widely used with little attention and hardly any non-technical discussion of their origin. This was shown in the study with most writers (90%) having no experience of using a text generator.

There are still problems with logical errors, or rather, a lack of any world logic, which leads to nonsensical texts. This is because the GPT systems are trained on unsupervised, or unlabelled, data. Any links between concepts are due to statistical relationships between word parts. GPT-2 and GPT-3 produce sophisticated text but cannot rhyme, indicating another limit of the mathematical generation technique, as there is no special coding for phonetics, which are not necessarily in the word byte pair encodings which are used to compress the data (BPE see Jaswal, 2019; rhyming see Branwen, 2020). A typical rhyming poem used as a prompt will generate non-rhyming continuations, or lines that rhyme by repeating the prompt words. If instructions or examples of poems are used in so-called prompt programming, then a story about a poet will appear. Further research continues in this area (for poem generation see Wang et al., 2021). Previous machine learning poetry research includes complex trained systems that do not perform well, such as Deep-speare, a Shakespeare sonnet generator (Lau et al, 2018).

The use of unfiltered training text from the internet leads to stereotyping and bias present in the original appearing in the generated text. This is a problem as the source text is too large to effectively filter. See Brown et al. (2020) for a short discussion of fairness, bias and representation in GPT-3. Bias is a major issue due to the limited nature of the training corpus, since most online text is written by a particular racial and class group. Bias is systemic and generally ignored as an issue unless raised by people actually affected (Mehrabi, 2019; Bender et al., 2021). This study does not examine these issues.

Generally, semantics (including the ontological relationships of words) and the zero shot learning (ZSL) of the newer generative models is now an area of considerable research interest (Geng, 2021; SAM, 2019). This will inform future GPT-x models, due to the limitations of current models. For instance, chat systems used for customer support or liaison will inevitably make logical or real world errors, as well as present bias, which could insult customers, reduce brand loyalty and lead to legal challenges.

Technologies such as augmented reality (AR) or mixed reality (XR) could be used in real-time with generated text or speech dialogue, along with wearable technology for emotional monitoring. These and other practical applications of text generation are not discussed in this study.

Text to image software as used in public systems such as Night Cafe and Midjourney represent popular new uses of the generation systems. The software is rapidly improving (Imagen Google Research, 2022). A simple text prompt is used to create an image. The text prompt can be simple ('One cow') or long ('a person made of cheese, wearing a cowboy hat, holding a spoon, on the moon :: 8k, ultra detailed, unreal engine render' (Midjourney, 2022)). This is like the descriptive label in a gallery or art book, used to create the painting. There is burgeoning field of AI generated art and animations. This subject is not discussed here.

Materials and Methods

Initially I experimented with Recurrent Neural Networks RNNs, and older and simpler form of machine learning, such as RNN character-level language models (char-rnn). See Karpathy (2015). I adapted Karpathy's code and ran it locally on a laptop but this was suitable for early learning only. The output was not particularly convincing and it was more of an educational system, as intended.

I also worked with NLP (Natural Language Processing) using the Python NLTK for genre analysis, to then create a genre based generator, but this was inconclusive. The Max Wolff GPT-2 example for Google Colab (Wolff, 2019) was used to experiment with training a generator using genre fiction. Further research in genre identification was conducted, see Chiang (2015), Worsham (2018) and Jia (2019).

At the time of design the best generator available was OpenAI's GPT-2. A public system at Talk To Transformer, now InferKit (2021), was tested along with Hugging Face GPT-2 online applications, and some simpler bespoke systems (Hugging Face, 2021). During the study period, OpenAI GPT-3 was announced but was not generally available. It is now commercially available from Microsoft, see Microsoft blog (2020). I am currently (2021) using beta access to OpenAI's GPT-3 for experimental fiction continuation systems.

The decision was made to use a publicly available text generator of sufficient flexibility for experimental use. This is better for a public study since the generated text has to be at least as sophisticated as that from systems already freely available online. This replaced the initial idea of creating a custom text generation system, as- in practice this would have made the text generation system slow to develop and underpowered when hosted.

An online system, Fabrice Bellard's Text Synth, used a recoded compressed version of GPT-2. I communicated with him prior to the test and he agreed to release the Text Synth system for use in the study (Bellard, 2021).

Bellard is a leading Open Source programmer. Full license details were provided in the Credits section of the live experiment. The experimental system is now available as a free creative writing tool at the Story Live website (Story Live, 2021). It is now using GPT-J 6B, GPT-NeoX 20B and others.

The approach is data-led without bringing any preconceptions to the results. This informed the simple design of the interface and the open-ended experimental tasks.

Sentiment Analysis

Sentiment tools are very popular and there are many different approaches, all using NLP (Ahmad, 2017).

IBM Watson Tone Analysis was used for sentiment analysis of the comments and feedback texts from respondents (IBM, 2021). This system was chosen because it provided a range of emotional sentiments, as well as overall reasoning types. The identifiable sentiments included general Confidence, Tentative, and Analytic modes, and the emotional modes Joy, Anger, Sadness and Fear.

It was also already running online with an easy to use interface, suitable for the feedback analysis. It also gave numeric values to the sentiments suitable for data processing.

Other systems were tested and found to be either inaccurate on real-world examples (e.g. our actual feedback data) or were commercial systems and involved charges.

A small section of the feedback was directly human analysed, to provide an accurate appraisal of positive, negative and neutral tones, which because of the length of the samples was often incorrect when processed by machine.

Method

As described above the experimental aim is to gather data, including feedback, on how users

of text editors (typically word processors, but also including custom creative writing software) might respond or adapt to using computer generated text as part of their creative process. The results can be used to improve the design of creativity support applications. The study was designed to present a text generation system in a minimal working environment. This was then used to gather both quantitative data and comments on the use of the new system. This data and the analysis of responses are of general interest to AI hybrid researchers, and directly relevant to any forthcoming text creativity tools. There is no hypothesis to test and so this is an example of grounded theory research. See Yanai and Lercher (2020). All of the respondents' feedback is in the Appendix.

Design of the experimental writing tool - initial investigations

We took an iterative prototyping approach to designing the test system that went through a range of potential solutions. This is a common approach, see Nielsen (1993). The final design incorporated a text generator and a text editor, and a way to save and continue with the next text experiment, or repeat if desired. This was based around a combined generator/editor designed as a new creativity support tool. By using a custom writing prompt (in this case, an ambiguous image) open-ended tasks could be set for the respondents to complete.

The Feedback Answers

Individual replies to the various questions are in the Appendix. These will give insight into the range and depth of comments and more understanding of the aggregated results shown in the report.

Experimental Ethics

Data

No personal data or proprietary research is revealed. The data shown in the Appendix is not indexed but indexed data is available to researchers on request. The generated texts are also available.

General Data Protection Regulation and Privacy

The study was designed to be GDPR compliant at all stages, in line with current legislation. For details of this please see the Experimental Design section in the Appendix. In particular the data was anonymous, with few personal data requests. Gender was not collected and age was only sampled in large ranges, to indicate working or not age groups. No location data, such as respondent country, was stored.

GDPR and AI

It is worth mentioning that although an AI system is in use with the respondents, there is no collection of personal data whilst they are using the system, which is, in addition, anonymous. So the AI is not collecting and collating data to build a personal profile of the user, either to train the AI, or make product recommendations, adjust credit and insurance scores, and similar commercial uses. This sort of privacy-invading AI is called ‘conversational’ and is used in chatbots and voice assistants, which are not used in this design (Meadows, 2018; Adamopoulou et al., 2020). Passive data gathering is prevalent in the entire IoT industry (the ‘Internet of Things’) and continuously collects location and activity data along with biometric data.

Respondents

I approached an audience from my professional and personal contacts. I have many people on a mailing list from my creative writing software released in 2010. I also advertised to two professional forums for publishers and writers. These are active users of writing tools, and from a variety of writing backgrounds, both professional and serious amateurs such as non-commercial novelists.

Final design

This is a multi-mode sequential study with:

- a linear sequential study,
- introduction and help video, privacy and credits information,
- respondent profile data gathered at the start, so information gathered even if they abandon the study,
- three text generation experiments of increasing complexity (caption, news, short story) with Likert scale ratings and question response feedback,
- a set of final feedback questions with Likert scales and feedback comments.

Design Principles

The user interface was designed to be minimal to reduce distractions, and produced some unsolicited comments, mostly positive.

The Introduction page explained the overall aims, had essential links to Privacy and Credits, and a link to a video of the entire study including the experiments.

Credits and privacy pages were separate and linked from all parts of the study. The credits page also discussed copyright issues for generated texts. The initial video help was a 4m 12s

voice-over walkthrough of the entire study. This was to familiarise people with the study appearance and their actions, before they arrived at the experiment pages.

The Profile page anonymously gathered basic experience, occupation and age data. These questions were optional but most people responded.

Each of the three experiments had a feedback page that also allowed a repeat of the experiment, in case people wanted to try again. This was not actually used by any respondents but we gave them the choice.

After the experiments section, a series of questions included a Likert scale response to a simple question, and space for freeform feedback. Again, this is easy to change in a follow-up study.

Data was gathered from page navigation and content files, enabling timings of navigation as well as collecting direct text responses. Individual page activity was recorded with an anonymised remote monitor program (see Fullstory 2021) and overall activity was recorded using page analytics.

The feedback comments are in the Appendix. These provide useful data for other researchers.

Figure 1. Flowchart of the Study - online pages

This user journey started with a section providing an Introduction with Help video, plus Privacy and GDPR information and Credits links.

The introduction leads to a Profile which gathers basic data, and was anonymous.

This was followed by three Experiments of increasing complexity, which can be repeated if desired. Each experiment was followed by a Likert rating score on ease of use, and a free text panel for detailed feedback.

After the third experiment, there was a set of six feedback questions with Likert scales and free text panels for comments.

Finally, the respondents were thanked for completing the study, and provided with links for repeating the experiments if desired, plus a link to the text generation and editor screen for further work outside of the formal experiment.

Three Writing Experiments

The experiments challenge the respondent to create a new piece of text, prompted by a simple but ambiguous image. The respondents used a text generator to create new text, on the left side of an uncluttered screen. Then it was copied and edited it in a separate editor on the right side. Editing was not obligatory but was advised. Once they finished the text, they continue, and data is saved. This leads to a feedback page (numeric rating and comment text) with a link to repeat the experiment if desired.

Figure 2. Text generator and editor page, with image prompt, for the Caption experiment

The three experiments were designed to produce more text each time. The first asked for a short caption or meme, the second a news story, the third a short story. This also provides an indication of which type of work is most amenable to input from text generation.

Each of the three experiments was followed by detailed questions with Likert scales and text feedback panels.

Results - Processing, Statistics, Charting and Presentation

1 Respondents

Respondents

The cohort was a cross section of different writer occupations from publishers, professional writer groups and users of creative writing software.

Respondent Comments

Respondent comments are in the Appendix.

Summary of Respondents

The study attracted 82 respondents from these professional groups. 42 Professional, 35 Amateur. 15 writing occupations were represented. 163 occupations (more than one could be selected per respondent) including 6 Other (not in preset choices).

No University students were in the study, as my local group at the Creative Computer Institute would be very aware of the technologies. See Henrich et al (2010).

Table 1.1. Respondents' writing occupations	
Category	Total (more than 1 category allowed per respondent)
Academic, teacher	19
Artist	10
Copywriter	8
Fiction	35
Journalist	12
Memoir, historian	14
Playwright, scriptwriter	5
Poet	13
Report writer	13
Scientist	3
Scribbler, for fun	15
Student	10
Other: Games writer	3
Other: Lyrics for music	4
Other: Graphic novelist	1

The most common group were fiction writers (35, 22%), and the least common were Scientists (3, 2%) out of 163 designated occupations.

80% were of working age (18-65), and only 2% were students. Fiction writers were the most represented at 22% of all occupation selections. Of those with only one writing occupation,

Fiction writers stood out at 38%. Most writers had never used a text generator (89%).

Professional writers (51% of cohort) were more engaged and stayed on the study 35% longer than Amateurs (43% with 6% unstated).

126 individual experiments were completed by the 53 active (from 82) respondents. 29 respondents did the profile and arrived at the first experiment but did not complete, continue or leave scores or feedback. 60 Caption experiments, 36 News and 30 Story experiments were completed and saved.

60 respondents arrived at the first experiment, and 53 left feedback (52 starting at the Caption experiment and 1 who started feedback with the News experiment). This gives 53 active respondents out of 82. 7 left the study without leaving feedback.

Two thirds of the respondents ignored the instructions and did not edit the generated text.

Respondents that edited the generated text spent nearly three times as long on the study (256%). Of those that didn't edit, three quarters only did one experiment of the three.

Of occupations and length of time on the study, Games writers were on over twice as long as Academic writers, the next longest group. This might be because Games writers are familiar with generative environments in games programming, and investigated the experimental system more thoroughly than others. Academics may have a more general interest in new systems and have a more thorough approach. At the other end of the scale, Scientists were the fastest through, perhaps uninterested in writing to order on unexpected topics.

Detailed analysis - All Occupations

Multiple choices were allowed. There were 15 occupations (12 on the selection list, plus 3 named in the Other category). 163 occupations were selected in total. See later for more Occupation breakdowns.

Unique occupations

Some respondents chose only one occupation.

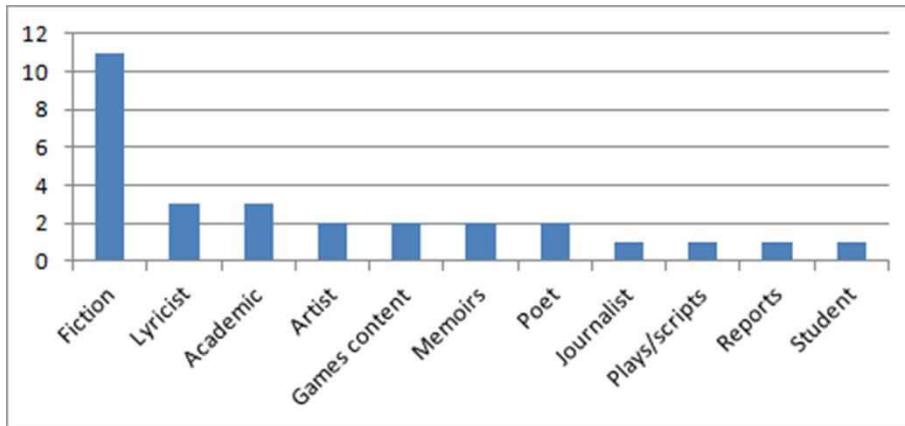


Figure 1.1: the numbers for occupations when only a single occupation is selected

29 writers from the 82 (35%) had only one occupation. Fiction writers were the commonest single occupation, with 11, 38% of this group.

Due to the selection process, a large number of fiction writers were present. The cohorts included users of creativity software, a self-publishing forum, and friends and relatives of creative writers. So this group might be favourably biased more towards creativity software, as used in the experiments.

Experience of Text Generation

74 from 82 (90%) had not used a text generator before. Of the 4 who could name the program they had used, 2 were relatively new generators. Named text systems were LISA (a 2018 chatbot, based on the earlier ELIZA), Text Synth (the program used in the study), a random poetry generator (2008), and a Commodore Amiga program (1984). One respondent said their previous generator was ‘my unconscious’.

This shows that even writing professionals have little knowledge of advances in writing technology.

Age of respondents

55 of 82 stated their age (67%).

80% were in the range 40-64 years.

15% were under 40 (18-39) or 'digital native' age, as web pages and websites established from mid-1990s.

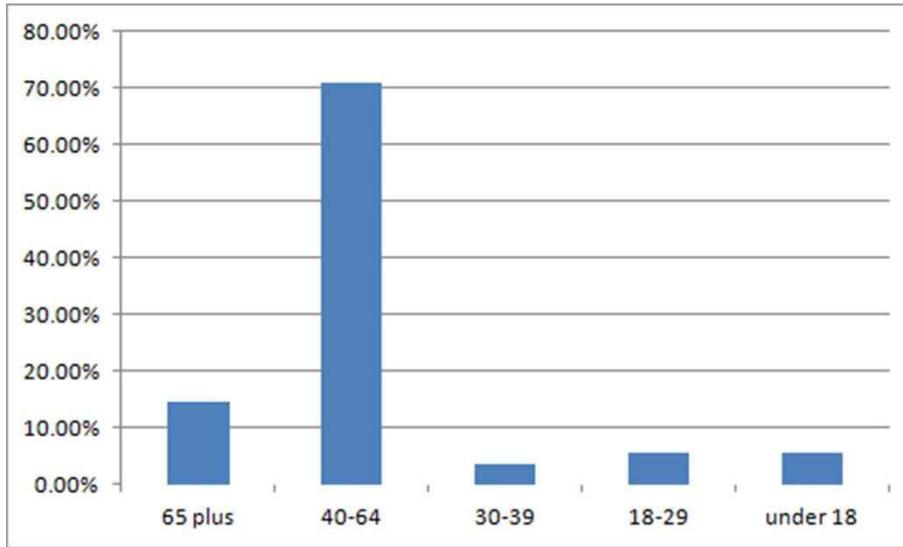


Figure 1.2: the numbers of people in each age group, when stated.

Figure 1.2 shows that the majority of the respondents (80%) were of working age (18-64 years, as we did not use University students). The age divisions could have been smaller to pick up more nuances. It also shows bias in the group for middle aged people, perhaps showing that fiction writing is the preserve of older people with more spare time.

Age differences

Age difference between different occupations. In the study we consider the null hypothesis:

$$H_0: \text{Age in } G = \text{Age in not } G,$$

where G defines some group (Academic, Artist, Poet...).

So, for $G = \text{Artist}$ hypothesis is:

$$H_0: \text{Age of Artists in population is equal Age of other (non Artist) in population}$$

Table 1.2: chart of Employment Group p-values

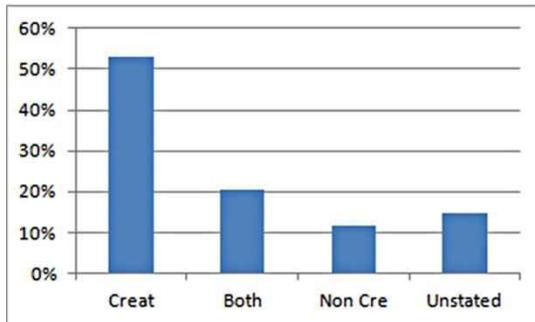
Difference in the Age in and out of the group for these three groups are defined by Mann – Whitney *U* test. Results of calculations.

Occupation Group	p-value (age difference)
Amateur	0.129
Professional	0.121
Academic	0.53
Artist	0.49
Copywriter	0.937
Fiction	0.343
Journalist	0.53
Memoir history	0.046
Plays scripts	0.27
Poet	1
Report writer	0.23
Scientist	1
Scribbler	0.0329
Student	0
Other	0.93

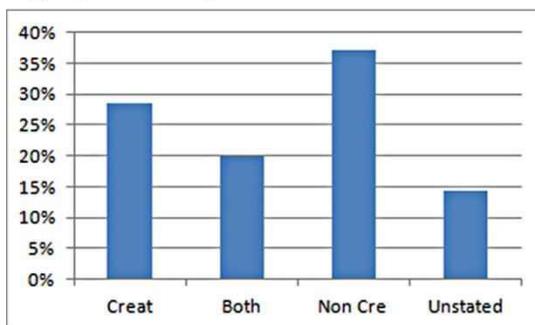
In the p-values table, there is significant difference for three groups - *Memoir history*, *Scientist* and *Student*. This indicates older occupations (memoirist, historian and scientist) and younger (student). This is to be expected, indicating respondents are of a typical professional cross-section.

Computer or device Operating System

Windows (34 instances)



Apple (35 instances)



Linux (13 instances)

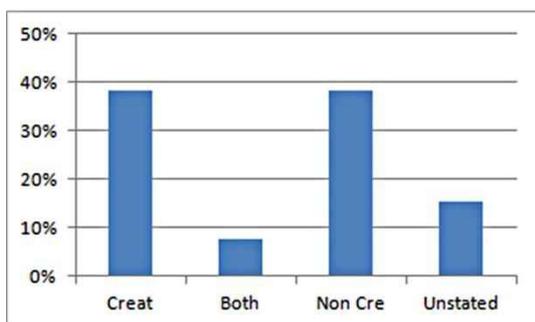


Figure 1.3: the numbers of people in each operating system, detected by the system data

Windows (34 people instances), Apple (35) and Linux (13) total data included all OS types, desktop and mobile. We did not examine differences in large screen, tablet or mobile phone usage in this study.

Creative writers largely used Windows PCs, and non-creatives used Apple devices. This might be due to cost, since writing is by reputation a badly paid occupation. Apple computers are generally used for design, desk top publishing, and other visual occupations, which might not include writing as a profession.

Gender

Gender data was not specifically collected. 37 people (45% of respondents) gave email addresses to receive follow-up reports.

Table 1.3 Gender of email names

Male name	16	(43%)
Femalename	11	(30%)
No gender name	10	(27%)

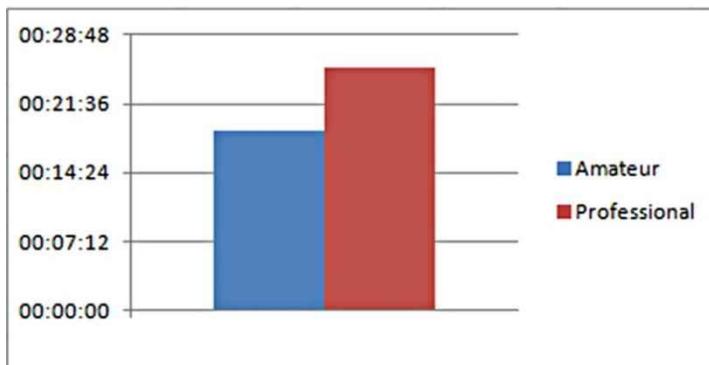
This indicates higher preference for anonymity in females, rather than the actual make-up of cohort.

2 Study analysis

Time on Study – Amateur and Professional writers

Professional status is important in this study as we are interested in whether working writers were aware of text generation, and their reactions to it. For time on study, the duration figure (minutes and seconds) is from the start of the study after the self-identification form, until the last question answered. This covers people who only did part of the study, but excludes people that only filled the Profile form, and then left. It also excludes time reading Help pages and viewing the Help video.

Time duration on the experiment of
Amateur and Professional respondents
Zero times not included



Zero times included

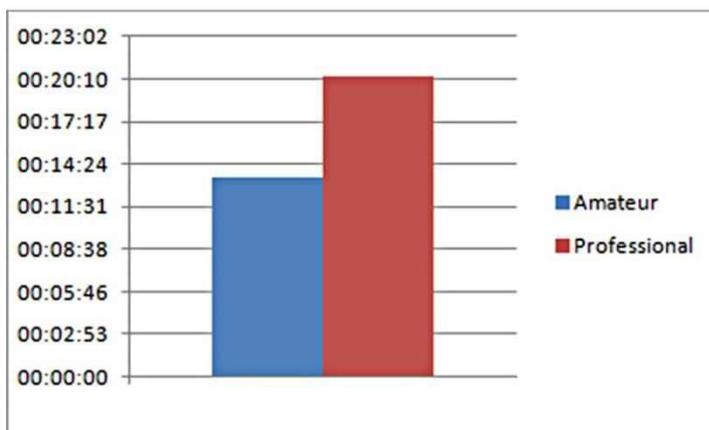


Figure 2.1 Time duration (in hours, minutes, seconds) on overall experiments and subsequent questions, of self-rating Amateur and Professional respondents, excluding people that filled up the personal data form and progressed no further.

The first chart above shows that professionals stayed on the study 33% longer.

The second chart shows that if zero times (start form only) are included, the difference becomes more marked (51% of professionals stayed on longer), showing that amateurs were more likely to abandon the study.

These charts show that people with more commitment to writing spent longer on the study, achieving the researchers' aim of engaging stakeholders.

Time on Study - Occupations

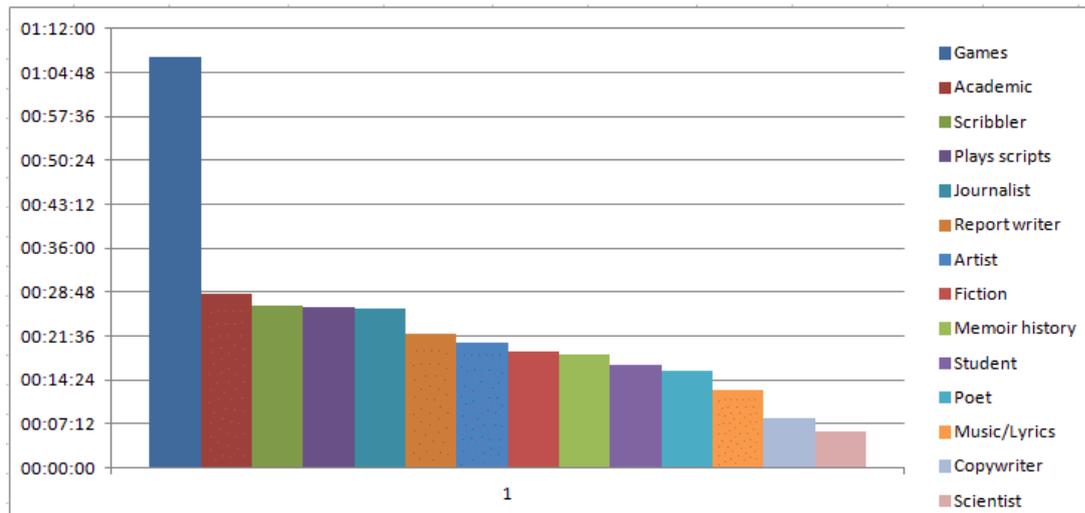


Figure 2.2 Average time on study per occupation, ranked in descending order

Other: Game writers (content, scenarios) had the highest duration (01:07:26), at over twice the next longest, Academic (00:28:27) (237% more). Games writers were perhaps interested due to the use of generated content in gaming . But as Games writers were a small sample, this could be caused by a respondent excursion from the study adding time.

Later in section 4 'Creative and Non-Creative Occupations' we will examine any differences between creative and non-creative writing occupations, a loose distinction but useful for analysis.

In Figure 2.2 the creative modes of Artist, Poet, Plays and Fiction writer are dispersed across the range. Academic and Scientist roles are at opposite ends of the scale.

Games writers were much more engaged with the study than any other group. This is perhaps due to their professional use of generated content, images (such as backgrounds or game items), plotlines and dialogue. This shows high stakeholder engagement with the study.

Further studies will be needed for this area.

Occupations ranked from Creative to Less Creative (poets, copywriters etc.)

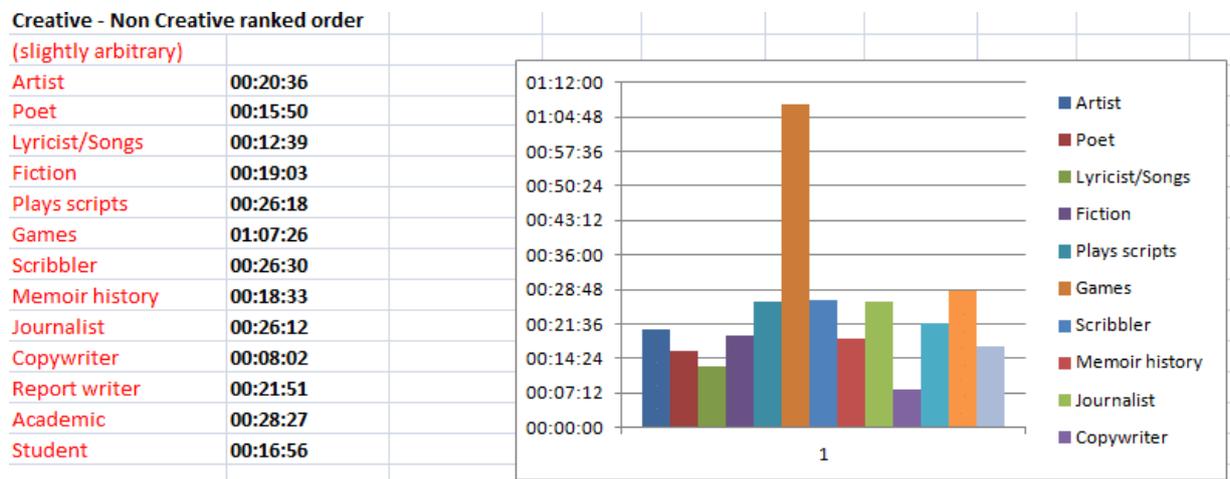


Figure 2.3 Average time on study per occupation, ranked from creative to non-creative.

The results show no overall duration differences between creative or non-creative occupations. Of the less creative writers, Copywriters are the fastest (least time spent on study)..

Three Experiments – Caption, News and Fiction

All experiment used the same ambiguous image as a prompt for initial ideas. See Credits for details of image.

The actual generated, edited and submitted texts are not published in this report.



Figure 2.4 The photographic prompt: Bryan Riesberg and Maxine the dog (2019). See Credits.

Caption – “Write a caption or meme”

The first Caption experiment had instructions to create a short text. This was expected to be a sentence or two. However the Caption task average word count was 88. Only 5 out of 50 (10%) used fewer than 100 characters, a typical caption length.

This shows that 90% were very engaged with the new way of making text appear on the screen, and that they were using the system for their own ends, rather than following instructions.

News – “Write a news item”

News task average word count was 93, only 6% more than Caption.

Fiction – “Write a short story”

Fiction short story task average word count was 159, showing that the respondents followed the instruction. This is ‘flash fiction’ length due to the constraints of the design of the editing area, which can expand but presents as a single panel. There were no imposed time constraints, but the respondents were experimenting for an academic study, not deliberately trying to write a story for, say, a competition.

Respondents’ average word counts
by experiment type

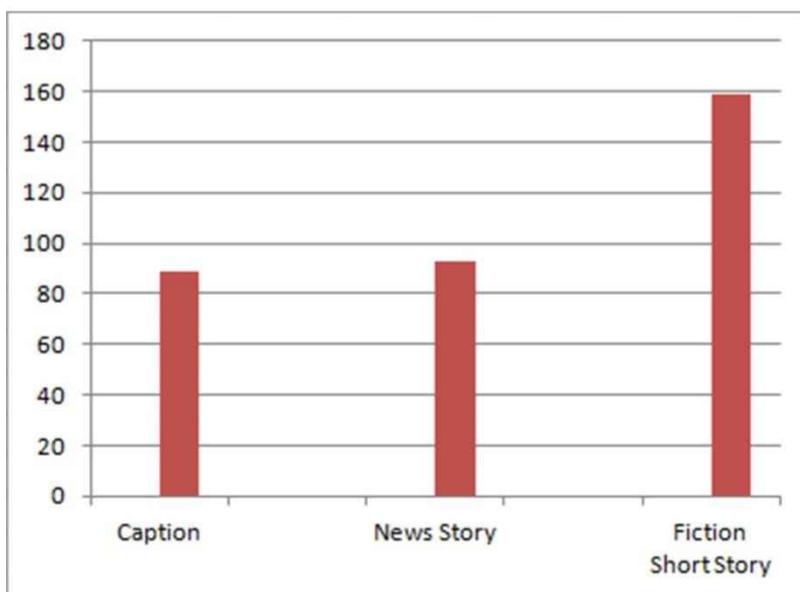


Figure 2.5 Average word counts per respondent by experiment

Text Editing during Experiments

We used the Fullstory screen monitoring service to record what people did during the experiments. This was anonymous. It allows some observations of cursor movements and typing in the editing window. See Fullstory (2021).

Text editing after generation

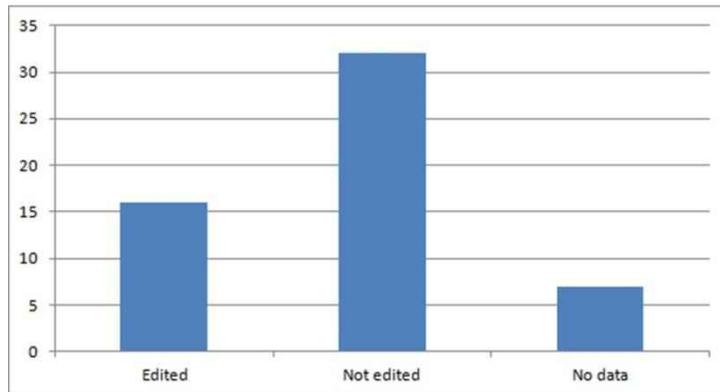


Figure 2.6 Text editing results, 2 drop outs in the ‘no data’ observations

Of those observed, one third of respondents edited the generated text, two thirds did not.

This shows that two thirds of people ignored the task instructions, which were to generate, then edit before saving. Perhaps they found the generated text sufficient without editing.

Other studies of writers using generated text also found this dichotomy (Calderwood et al, 2019; Gero and Chilton, 2019). This is an example of ‘casual’ creativity as mentioned in the Introduction.

The monitor recordings were unlabelled but it was possible to relate the monitoring data to a particular record through timing (eg, time on each section of the experiment) and searching for the actual text appearing in the window. This was used to locate a record in the data. This is still anonymised, as the records do not contain personal data.

Respondents that did not edit the text were less actively engaged in the experiments, shown in the three factors below: time on study; depth (completed number out of three experiments) and feedback ratings (Likert scales).

Editing respondents and non-editing respondents - time on study

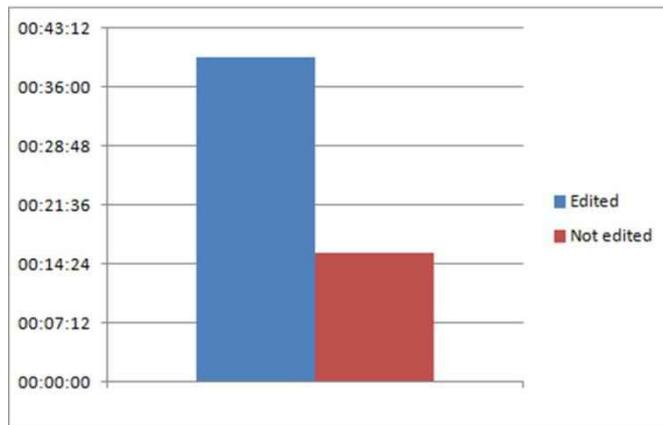


Figure 2.7 Editing respondents and non-editing respondents - time on study

Average time for editors 39m 38s, non-editors 15m 42s.

People that generated text but didn't edit were less engaged in the experiments, shown in the Figure 2.1.2 above, with editors nearly three times longer (256%).

Text Editing - Creative and Non-Creative Occupations

The 15 occupations (from Profile) can be divided into Creative vs Non-creative e.g. Poets or Copywriters (see above).

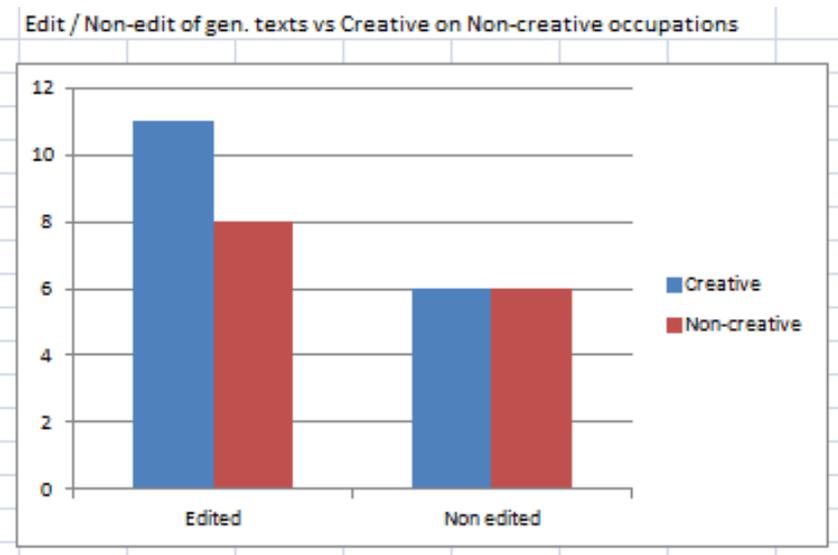


Figure 2.8 Text editing and creative or non-creative writing

There is a preference (over a third, 38%, more) for editing amongst Creative writers. There are further results using the creative and non-creative writing division later in this section.

Text Editing - Engagement Depth

There were three text experiments, Caption, News, and Short Story, requiring increasing work to complete. So a drop in engagement was expected during the experiment. Numbers of the three experiments completed are shown below.

Text editing and completion of the three experiments

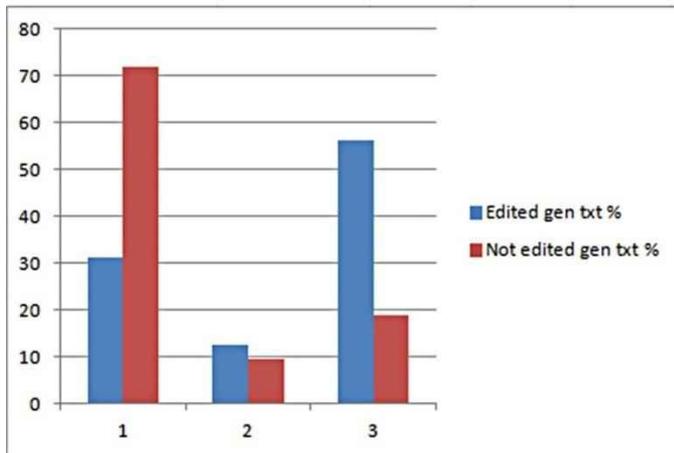


Figure 2.9 Text editing and completion of the three experiments

This clearly shows that respondents who did not edit the generated text mostly only completed the first experiment (nearly three-quarters), and those that edited the text completed all three. This shows that once respondents failed to edit text, their subsequent engagement was much reduced.

In terms of experimental design, perhaps a compulsory or a more engaging initial editing task is required to overcome resistance to editing (an action) by the respondent. Once they'd edited, they might realise how easy and preferable it was. This was the in the design with a simple Caption edit as the first experiment, but most ignored the instructions. Perhaps this indicates two main groups of future users, hybrid editors or co-creators, and those who use the text as it appears, in a 'casual' creativity read-only mode. But since most non-editors left after one experiment, perhaps they didn't like what they read.

3 Feedback questions after the experiments

Overall Likert scale results

Each of the three text experiments was followed by a feedback page comprising a Likert scale rating and a panel for written response. This was to record immediate feelings towards each experiment, and also collect data in progress in case they abandoned the study.

A 5 part Likert scale was used.

Structure of Questions

First experiment, to create a Caption:

‘Did you like using the text generator?’

What did you think about using the text generator?

Did it help or hinder or no effect?

How did it help you?’

Likert: 1 Like a lot/2 Liked a little/3 Neutral/4 Slightly disliked/5 Strongly disliked

Second experiment, News, the question was:

‘This second time, did you like using the text generator more or less?’

Third experiment, Fiction, the question was

‘This third time, did you like using the text generator more or less?’

Using this structure, respondents compared their experience as they learnt the generation and editing process.

Results

Table 3.1 Median ‘dislike’ Likert scores for 3 experiments

Caption Exp. Average	2.5	News Exp. Average	2.5	Fiction Exp. Average	3.1
Median	2	Median	2	Median	3

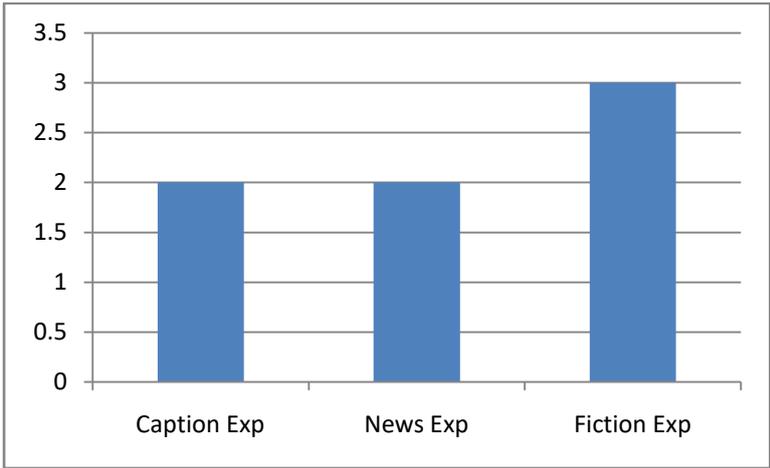


Figure 3.1 Median 'dislike' Likert scores for 3 experiments (score 1-5, 5 is strongly dislike, 3 is neutral)

Caption and News experiments feedback was slightly positive (Likert 2).

Fiction experiment feedback was neutral (Likert 3), less positive than Caption and News, indicating less suitable for fiction, or just enjoyed less as time went on.

Overall Likert scores

Engagement and Emotions

Question 1/6 and Question 2/6

Table 3.2 Likerts medians for Question 1/6 and Question 2/6 Engagement and emotions

Q 1 Like	Q 2 Feel
Median	Median
2	3

Respondents slightly liked using it (Likert 2), and had neutral feelings about this (Likert 3)

Plagiarism

Non-consecutive reverse scoring questions 3 and 5.

Table 3.3 Likerts medians for Question 3 Somebody else's work

Q3 Somebody else's work
Median
4

Table 3.4 Likerts medians for Question 5 Sell as own

Q5 Sell as own
Median
3

Respondents don't think that generated text is other's work (Likert 4), and they don't have issues (are neutral) over selling it (Likert 3).

The text generator is trained on originally human text from the internet, so this shows a lack of knowledge about the makeup of the system, or that people don't care.

Use in Word Processor

Question 4/6

Table 3.5 Likerts medians for Question 4 Use in Word Processor

Q4 Use in Word Processor
Median
3

Respondents are neutral about adding it to their usual word processor (Likert 3). This is perhaps due to the easy to use experimental system, or reluctance to add a separate creativity system to already over-complex word processors.

Changes over the course of the experiments

Correlations of experiment ratings and Feedback.

Each of the three text experiments was followed by a feedback page comprising a Likert scale rating and a panel for written response. This was to record immediate feelings towards each experiment, and also collect data in progress in case they abandoned the study.

The following charts have inverted Likert scores to show positive on Y axis, as the first of five options ‘Strongly Agree’ score was the lowest value 1.

The basic Likert data was further processed against other categories to gain more insight.

Likert details and medians are in the Appendix.

Caption, News, Fiction Experiments - Likert scores

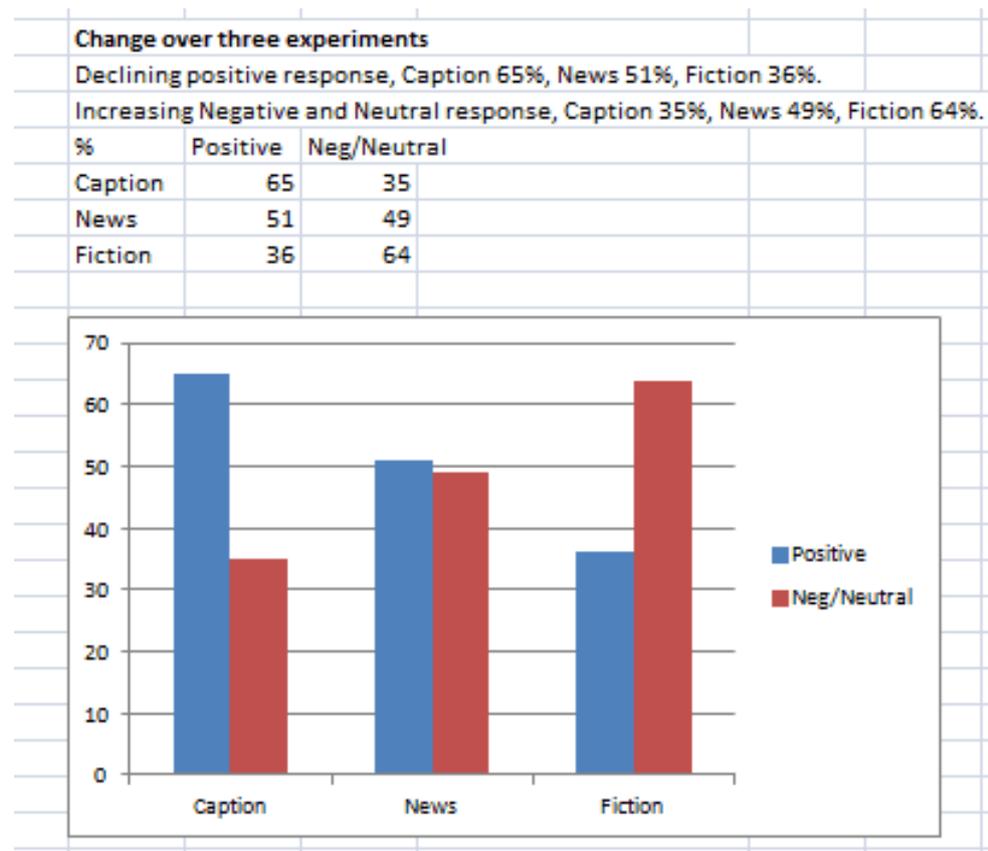


Figure 3.2 Likert rating change over three experiments, summed Likert scores

There was a decline in positive regard for the experiments over the course of the three experiments.



Figure 3.3 Likert rating change over three experiments by experiment, ranked and summed

The Likert scores were divided into Positive (a score of 5 or 4), Neutral (3) and Negative (1 or 2).

There is a clear declining positive response, from the first Caption experiment 65% completed, then News 51%, finally Fiction 36%.

There is also a clear increasing Negative and Neutral response, first Caption experiment 35%, News 49%, Fiction 64%.

Positivity halves over the course of three experiments. There is a large increase in Negative and Neutral scores, indicating that positivity may be an effect of novelty or enthusiasm. This might be showing a generalised positive regard for new experiences (also familiar as 'early adopter' enthusiasm) which then reduces as the limitations of the generated text become more apparent with use.

4 Creative and Non-Creative Occupations

For the purposes of this study, two contrasting writer groups from the designated occupations were selected. The Other occupation field allowed respondents to manually enter their profession. This is a simplification of the process of writing but can be used for analysis since the professional groups are distinct, not least in their own minds. There is a clear difference in self-image between a poet and a copywriter, although a copywriter might use poetic methods, for instance in advertising slogans and other short forms. Many creative writers also work as copywriters, editors etc.

Creative and Non-Creative division for purposes of analysis

Creative: Artist; Fiction; Plays scripts; Poet; Scribbler; Other: Creative;

Non-creative: Academic; Copywriter; Journalist; Memoir history; Report writer; Scientist; Student; Other: Non-creative.

Experiments combined (Caption, News, Fiction) by Occupation Creativity rating, and Likert scale

Binary analysis Like/Don't Like

Binary Like/Don't Like Likert scores for creative and non-creative occupations.

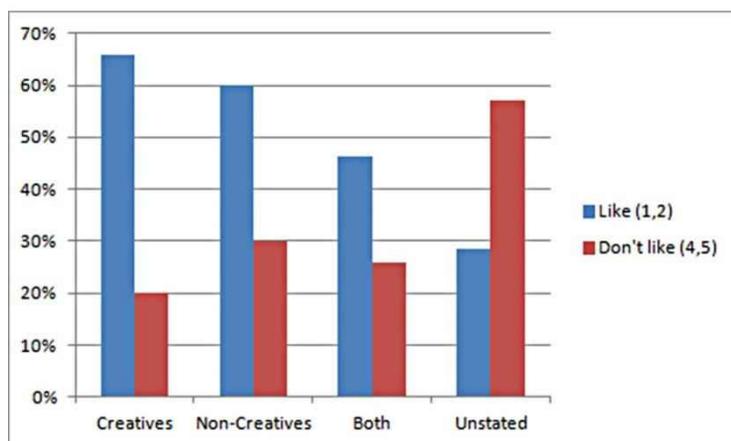


Figure 4.1 Binary Like/Don't Like Likert scores for Creative and Non-Creative occupations.

Creatives and Non-Creatives - Overall positive sentiment

Concerns positive sentiment for text generation experiments in this study, using Likert scores from all feedback ratings.

‘Like’ is Likert score 1, 2, ‘Don’t Like’ is 4,5. 3 is Neutral.

‘All feedback Likert ratings’ means the three experiment ratings, and the five questions at the end with scored ratings.

Figure 4.1 shows overall positive regard, with only a small difference between creative and non-creative occupations.

Respondents who did not state any occupation (‘unstated’) were the only group that clearly didn’t like the text generation experiments. This might show they are not actually working writers and were less likely to be interested in text-based work. The sources for respondents would have included some administrative staff, publishers etc. who might regard text work as labourous rather than creative.

Experiments combined (Caption, News, Fiction) by Occupation Creativity rating, and Likert rating scores – more detail

Combined all three experiment ratings and look/feel ratings Likert scales

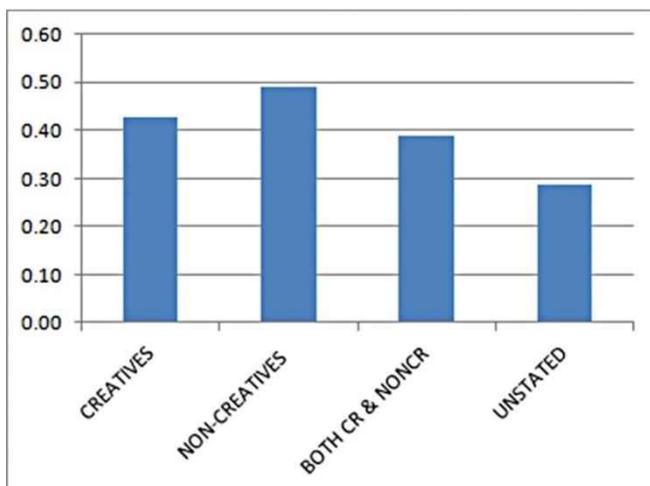


Figure 4.2 Overall Likert scores for Creative and Non-Creative occupations. On this scale, 1 is strongly positive, 0.5 is slightly positive, 0.33 is neutral, 0.24 is slightly negative, 0.2 is strongly negative.

This graph shows slightly higher approval from non-creatives, but the differences are small.

Again, the unstated occupation group shows less positive interest. This shows that stakeholders prepared to identify their writing are more positive about text generation.

Respondents who are not forthcoming about their occupation are more negative about text generation.

These two graphs show non-creative writers are slightly more positive about using text generation, at the time of the experiment, and afterwards, on reflection. Values are very close to neutral and generally show a small positive effect. The most interesting part of these ‘feeling’ answers are the feedback comments, which are analysed later.

Plagiarism and Ownership questions ratings

Combined Plagiarism and Selling Rights questions 3 and 5 Likert scales, by creative and non-creative occupations

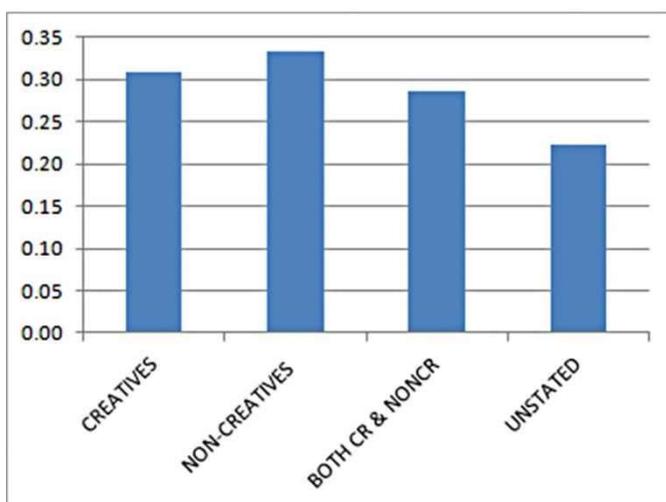


Figure 4.3 Combined Plagiarism and Selling Rights questions 3 and 5 Likert scales, by creative and non-creative occupations

On this scale, 1 is strongly positive, 0.5 is slightly positive, 0.33 is neutral, 0.24 is slightly negative, 0.2 is strongly negative. One of these questions was score reversed to remove casual click-throughs.

The Unstated occupation group, who were previously more negative about the experiments, showed a stronger disagreement to ideas of ownership of the text. The other groups were neutral on this topic. There was a wide range of interesting comments with these questions, showing that the respondents engaged with the concept of human or software ownership.

Use of text generation in Word Processor

Use of text generation in Word Processor Likert scales by creative and non-creative occupations

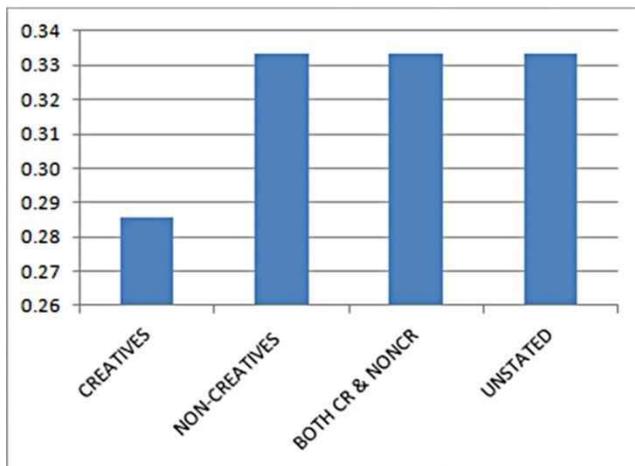


Figure 4.4 Use of text generation in Word Processor Likert scales, by creative and non-creative occupations

On this scale, 1 is strongly positive, 0.5 is slightly positive, 0.33 is neutral, 0.24 is slightly negative, 0.2 is strongly negative.

This chart show a preferences for text generation in a word processor amongst non-creative compared to creative writers.

Along with their feedback comments, it would seem that creative writers do not want the distraction of a generated text in addition to their own human-generated writing. There was also a sense that they were in competition with the text generator, and did not want or need the help, whereas copy writers or students might like some assistance.

Engagement – volume of comments

Volume of feedback comments by
creatives and non-creatives
(Y axis is text characters including spaces)

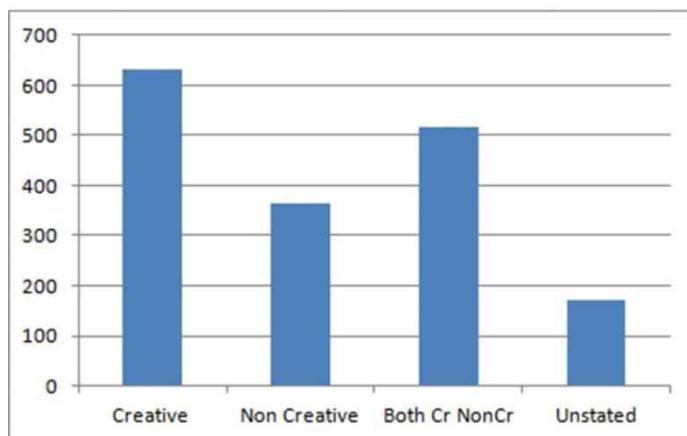


Figure 4.5 Volume of feedback comments, by creatives and non-creatives (Y axis is text characters including spaces)

Creative writers produced a much higher volume (73%) of feedback comments than non-creatives. This shows they were more engaged with the idea of using text generation, even though they were less interested in having it in a word processor or creative writing tool.

Creatives and Non-Creatives – Drop off over three experiments

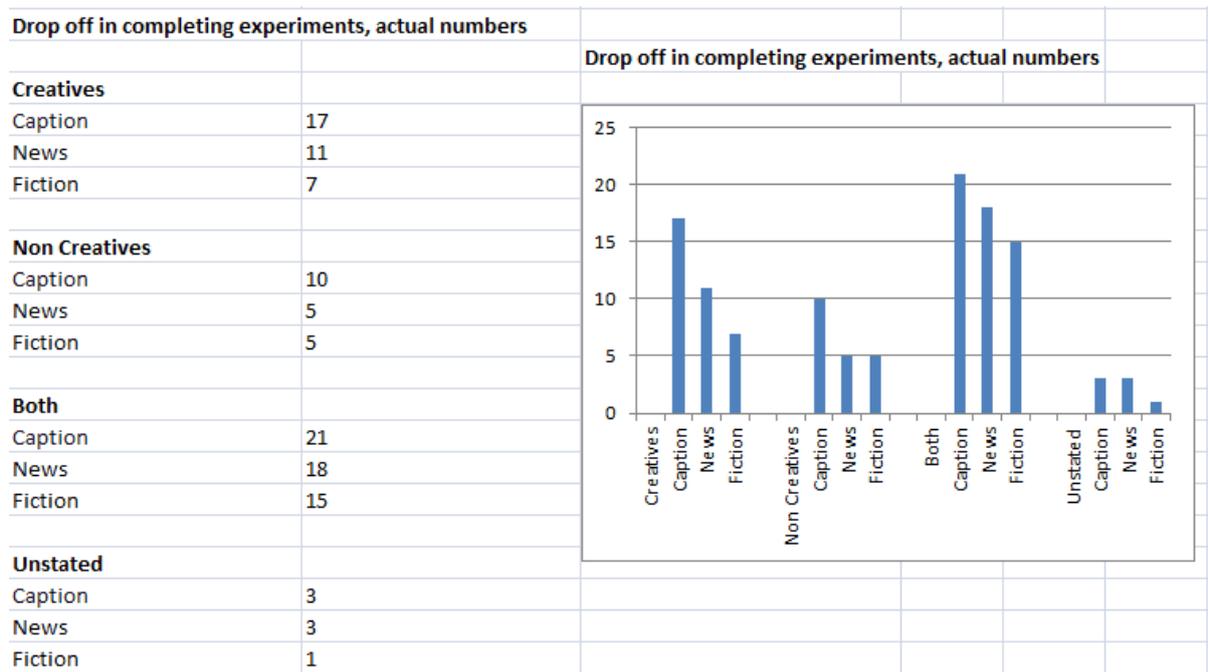


Figure 4.6 Completion of experiments. Drop off in experiment completion for Creative and Non-Creative occupations. Total of 116 experiments from 126 total, from 53 active respondents.

The drop off is about the same across each group, so engagement is about the same for all types of writers. This is most likely due to external factors such as fatigue or other commitments.

5 Sentiment Analysis

The study had three text generation and editing tasks, to make a Caption, a News article, and a Fiction story. These charts are for summed feedback text.

Sentiment analysis examines text for emotive words or phrases, and provides a score. We used IBM Tone Analysis which scores for Anger, Fear, Sadness, Joy, Analytical, Confident and Tentative tones.

For the following charts these were divided into Negative (Anger, Fear, Sadness) and Positive (Joy). The actual tone scores are also shown.

For the following charts these were divided into:

Negative (Anger, Fear, Sadness) (averaged) and

Positive (Joy).

The actual tone scores are also below in tables. 'Confidence' tone scored zero as it was not detected. This indicates the lack of confidence in using a new experimental system with unusual results.

This uses the same occupation divisions, which as earlier are:

Creative: Artist; Fiction; Plays scripts; Poet; Scribbler; Other: Creative; and

Non-creative: Academic; Copywriter; Journalist; Memoir history; Report writer; Scientist;

Student; Other: Non-creative

Creative and Non-Creative Occupations sentiment tones

IBM Tone Analysis scores on text feedback.

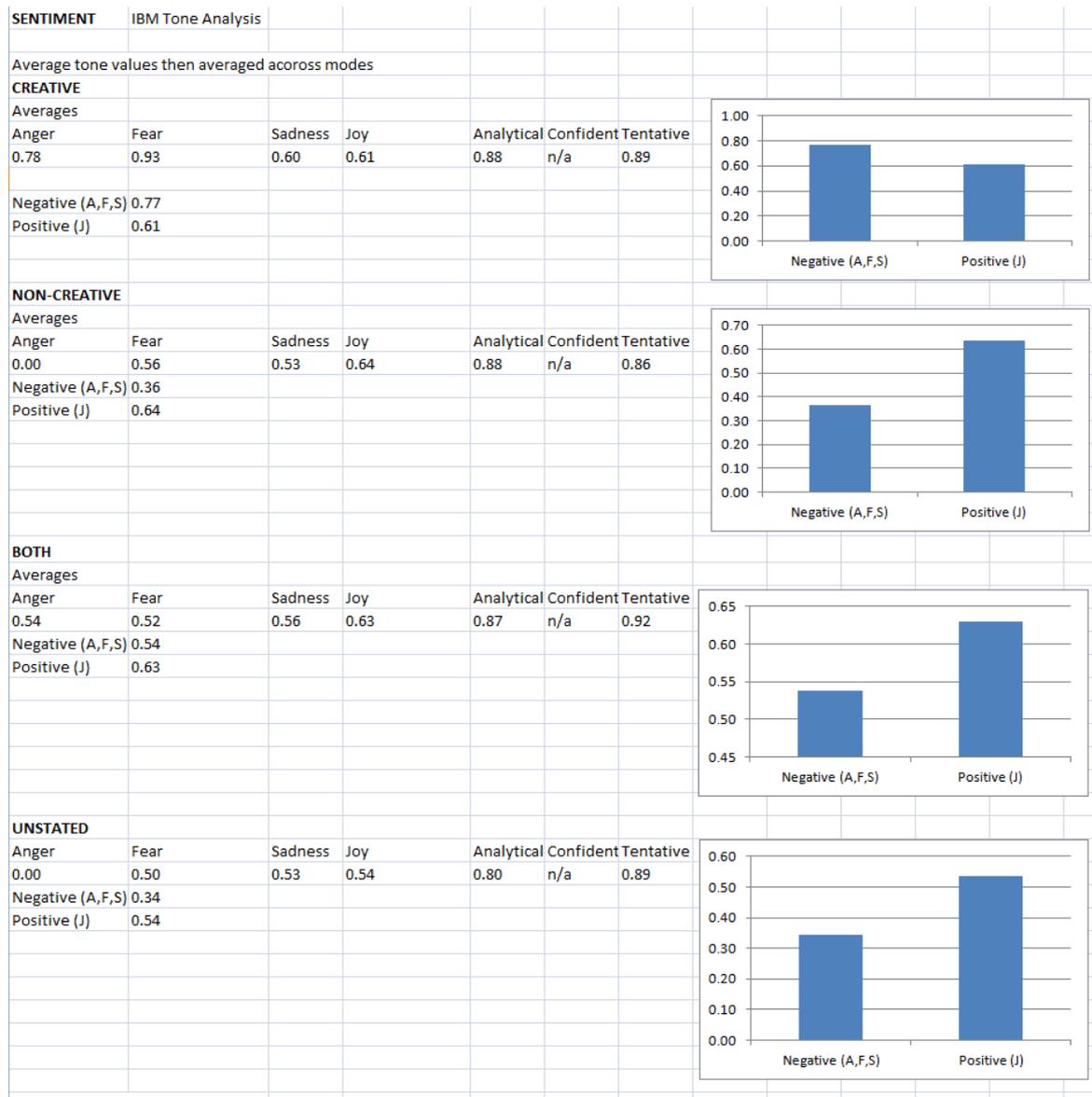


Figure 5.1 Creatives and Non-Creatives sentiment analysis divided into positive and negative

Negative sentiments (Anger, Sadness, Fear) are higher for Creative writers.

All other groups Non-Creative, Both and Unstated, showed a positive tone. The degree of negativity in Creative writers (0.77) compared to Non-Creative writers (0.36) is 213%, a large difference. Even Unstated, who generally have lower engagement, were positive. This

relates to the findings on Creative writers' relative lack of interest in using a text generator in a writing tool (see Figure 4.4 above).

Relationships between sentiment and occupation

Box plots are for descriptive use, showing the spread of the data. For this section, I made a Help page for Stats Explanation using box plots (Davis 2020).

In this section we compare the average value of each sentiment Tone (Anger, Fear etc.) by occupation as group. Then check null hypothesis about equality of the average value of the Tone in each occupation. Professional and Amateur selections are also analysed in this process, as they were not linked to any core occupation. Note that Confidence did not score in the sentiment analysis (i.e., none was detected) so there is no Confidence plot.

These scores are not one per respondent, as each can score several different occupations, such as Fiction and Poet, or less common, Scientist, Artist and Academic. This uses the same occupation divisions, which as earlier are:

Creative: Artist; Fiction; Plays scripts; Poet; Scribbler; Other: Creative; and

Non-creative: Academic; Copywriter; Journalist; Memoir history; Report writer; Scientist;

Student; Other: Non-creative.

<i>Table 5.1</i> Sentiment tone scores for creative occupations	
Creative occupations	
Sentiments	Average tone score 0-1.0
Anger	0.78
Fear	0.93
Sadness	0.60
Joy	0.61
Analytic	0.88
Confidence	0
Tentative	0.89
Summary	
Negative (avg. Anger, Fear, Sadness)	0.77
Positive (Joy)	0.61

<i>Table 5.2</i> Sentiment tone scores for uncreative occupations	
Uncreative occupations	
Sentiments	Average tone score 0-1.0
Anger	0
Fear	0.56
Sadness	0.53
Joy	0.64
Analytic	0.88
Confidence	0
Tentative	0.86
Summary	
Negative (avg. Anger, Fear, Sadness)	0.36
Positive (Joy)	0.64

<i>Table 5.3</i> Sentiment tone scores for both stated creative and uncreative occupations	
Creative occupations	
Sentiments	Average tone score 0-1.0
Anger	0.54
Fear	0.52
Sadness	0.56
Joy	0.63
Analytic	0.87
Confidence	0
Tentative	0.92
Summary	
Negative (avg. Anger, Fear, Sadness)	0.54
Positive (Joy)	0.63

<i>Table 5.4</i> Sentiment tone scores for unstated occupations (none stated)	
Creative occupations	
Sentiments	Average tone score 0-1.0
Anger	0
Fear	0.50
Sadness	0.53
Joy	0.54
Analytic	0.80
Confidence	0
Tentative	0.89
Summary	
Negative (avg. Anger, Fear, Sadness)	0.34
Positive (Joy)	0.54

Sentiment and occupation distribution plots

Fear

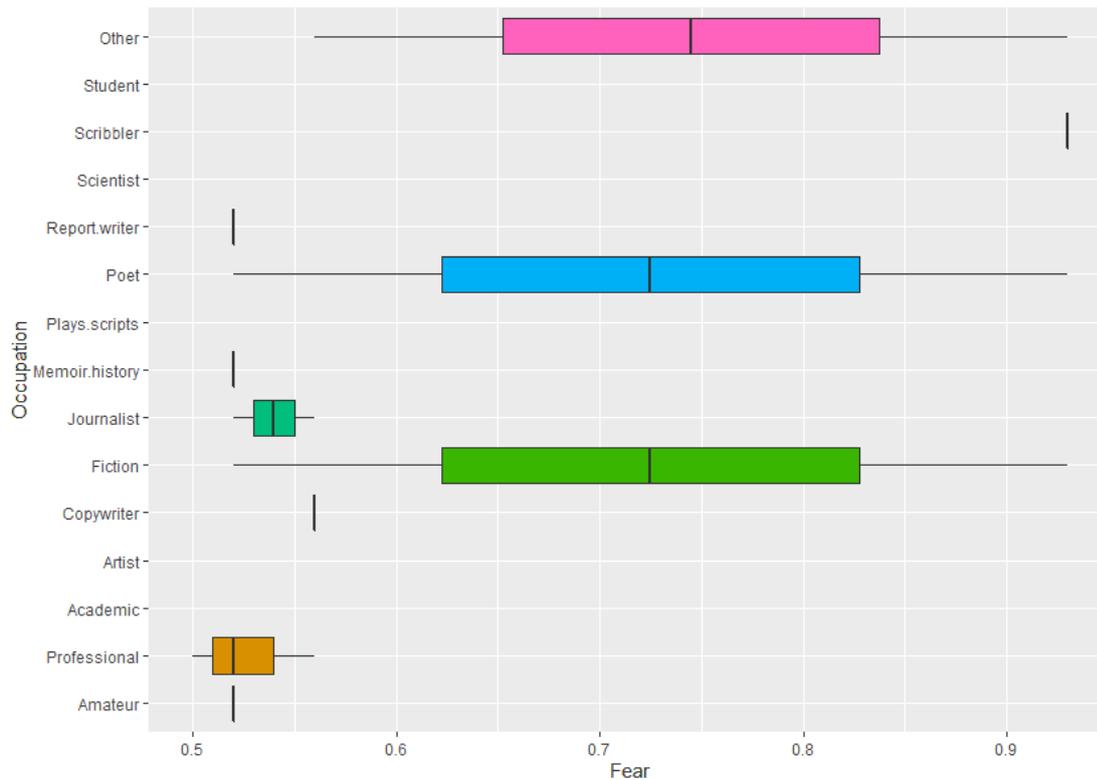


Figure 5.2 Occupation and Fear sentiment box plot

Fear is shown by the creative writing occupations Fiction, Poets and Other (Games writing and song writing), along with one Scribbler (creative). This is of interest as it shows that the typical creative writer is showing higher Fear, unlike the other occupations, apart from a slight amount of Fear (apprehension?) shown by Journalists and Professionals (not an occupation but scored).

Sadness

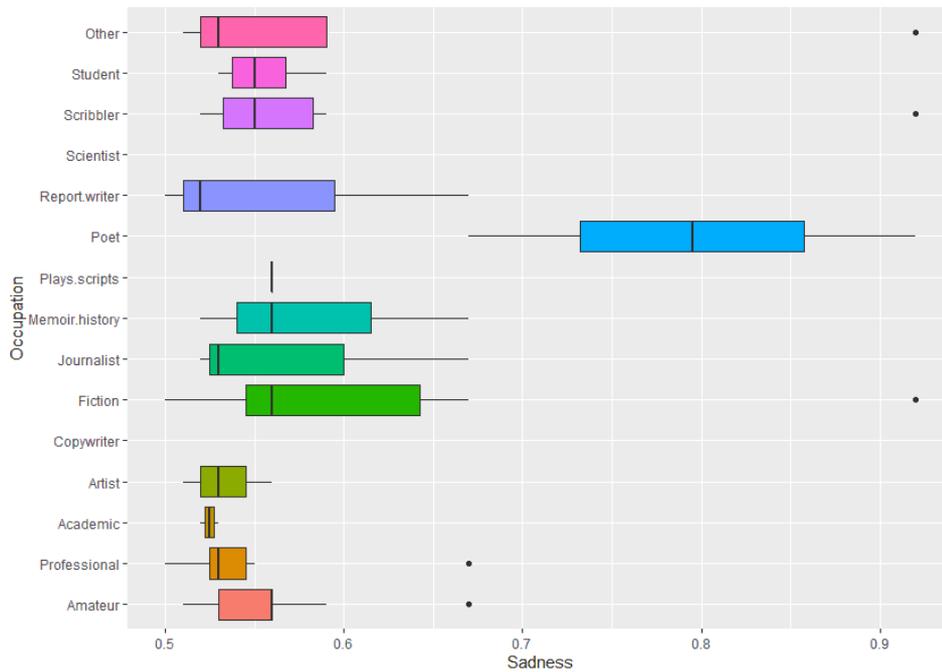


Figure 5.3 Occupation and Sadness sentiment box plot

Sadness is shown by Poets, as they might use more sad words (the Tone Analyser works on the syntax). Outliers are also shown by Other (games, songs), Scribblers, and Fiction.

Joy

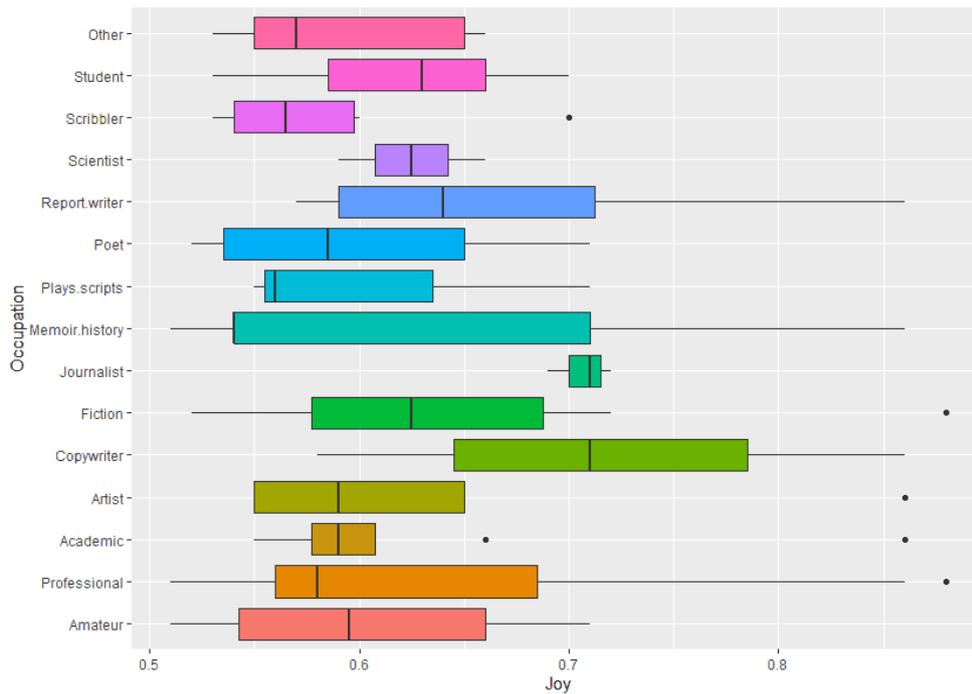


Figure 5.4 Occupation and Joy sentiment box plot

Joy is quite widely distributed across occupations, which is also shown in comparisons of Creative and Non Creative occupations, which only have small differences. Copywriters appear to be the most joyous when regarding the outputs of the generator, although several occupations have higher outliers (the right side dots). Perhaps copywriters can appreciate the usefulness of a generator to quickly produce more copy, perhaps for further checking or fine editing.

Anger

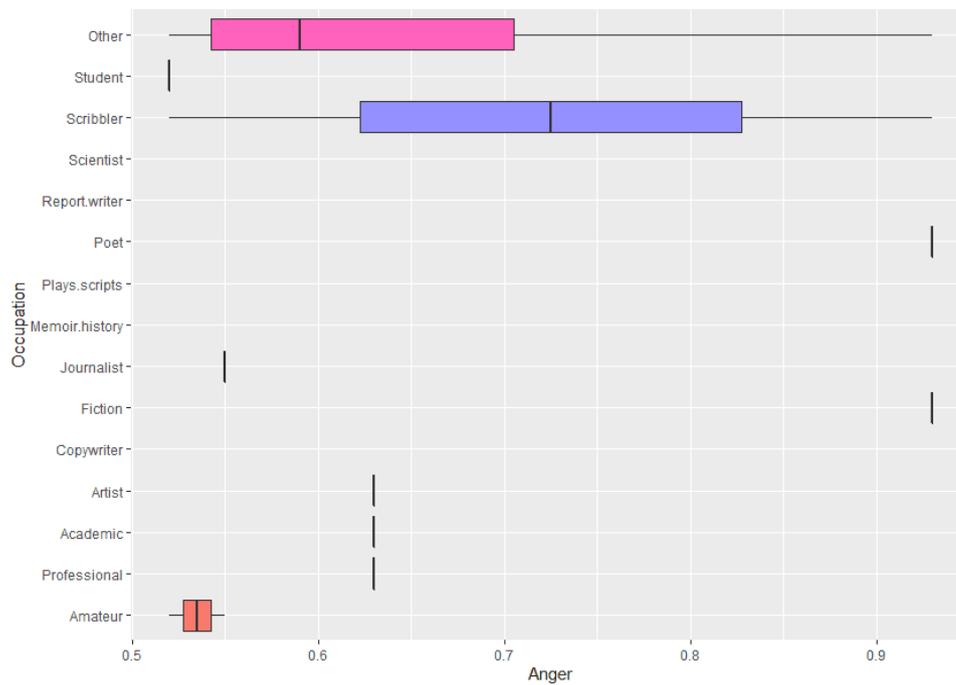


Figure 5.5 Occupation and Anger sentiment box plot

Anger is less common as a sentiment, showing the overall positive regard for the newly experienced generator and editor. This is a welcome finding for computer support staff.

Analytical

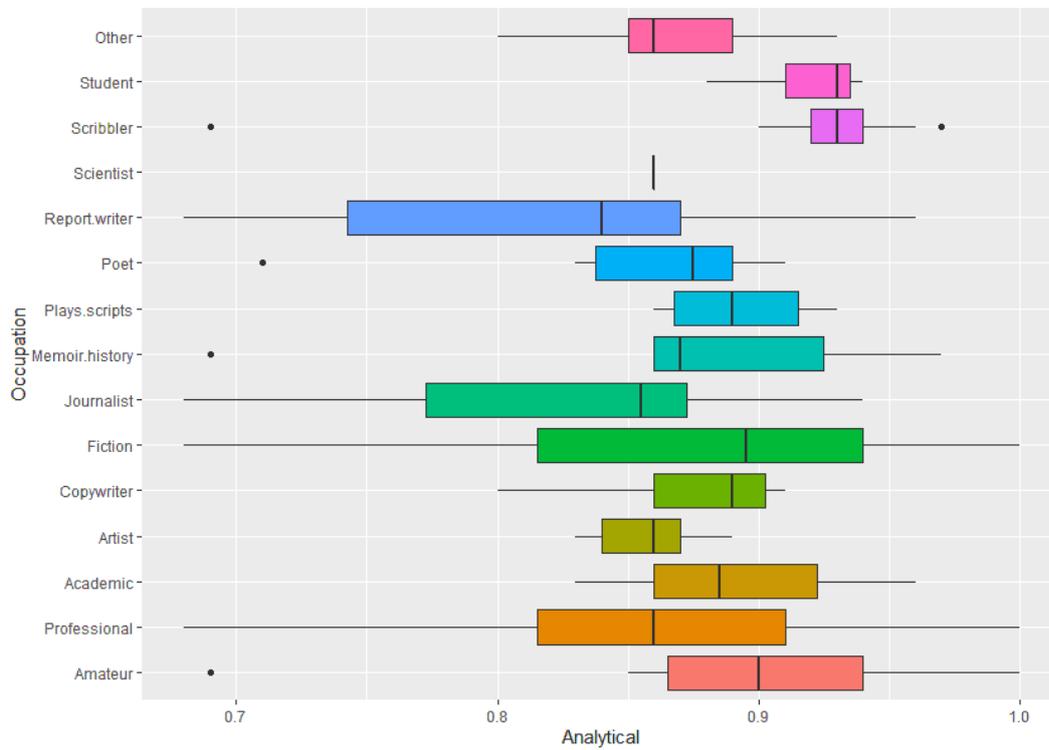


Figure 5.6 Occupation vs Analytical box plot

The Analytic sentiment plot shows the high level of considered feedback on the task, which is to be expected from largely older, professional or serious respondents.

Tentative

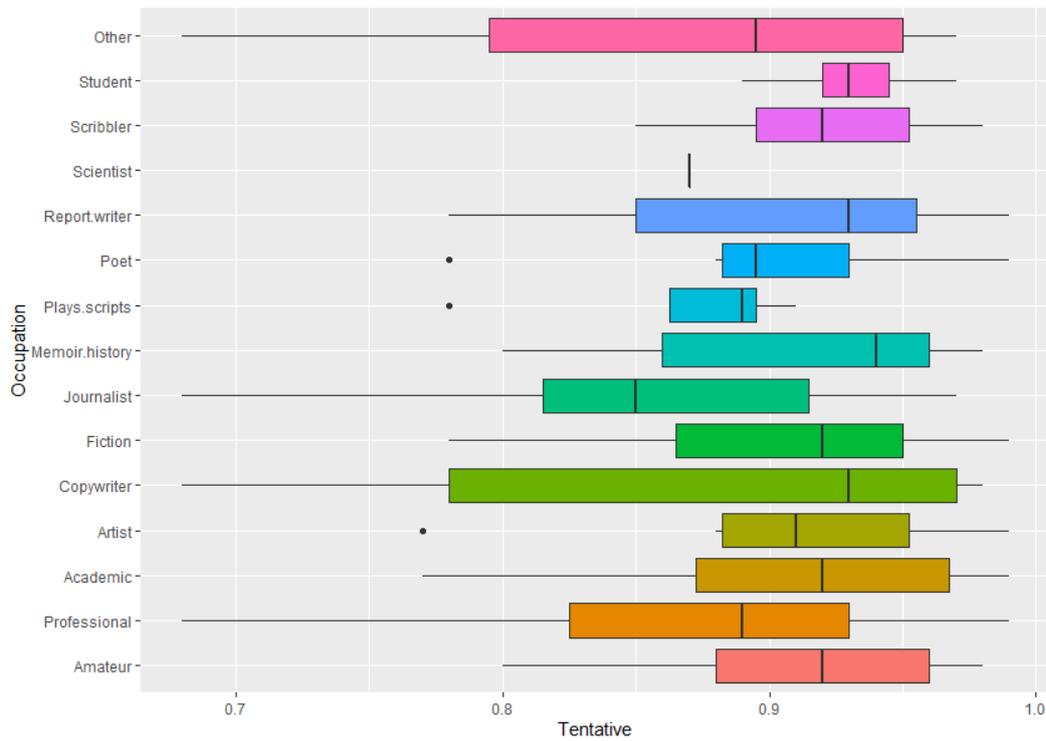


Figure 5.7 Occupation vs Tentative box plot (not Confidence sentiment did not score so no plot)

The Tentative sentiment plot shows the seriousness of the engagement with the task, recalling that 90% had never used a text generator before, so were wary of the new methods.

Feedback sentiment and occupations illustrations

The following descriptive box plots look as if there are differences (e.g., Poets show more Sadness) but the results are not significant, so they are shown here for observation and illustration only.

IBM Tone Analysis score comparisons - document level (each respondent's all-record feedback text).

Table 5.5 Results of calculations for t-tests for Joy vs [Anger, Sadness, Fear]:

Label/tone	p-value	Conclusion
Joy vs Anger	0.69	All p-values are higher than significant level $\alpha = 5\%$ (0.05). So, average values for Joy, Anger, Sadness and Fear are the same.
Joy vs Sadness	0.1	
Joy vs Fear	0.91	

t Student test (https://en.wikipedia.org/wiki/Student's_t-test). In the *t* test we considered difference of the means for two different samples. Only variables ***Anger, Fear, Sadness, Joy, Analytical, Confident, Tentative*** as numerical variables. ***Confidence*** had a null value (meaning under 0.5 in the Tone Analysis scale, so not significant) in all respondents (see data).

Poets and Report Writers

We took two respondent categories, Poet and Report Writer, which might be at the opposite ends of Creative-Non-Creative scale, as two examples for the test. This could be for any other pairs.

The first two graphs show average values of Sadness and Joy for Poets and Report Writers.

Plot – Joy (Poet, Report Writer)

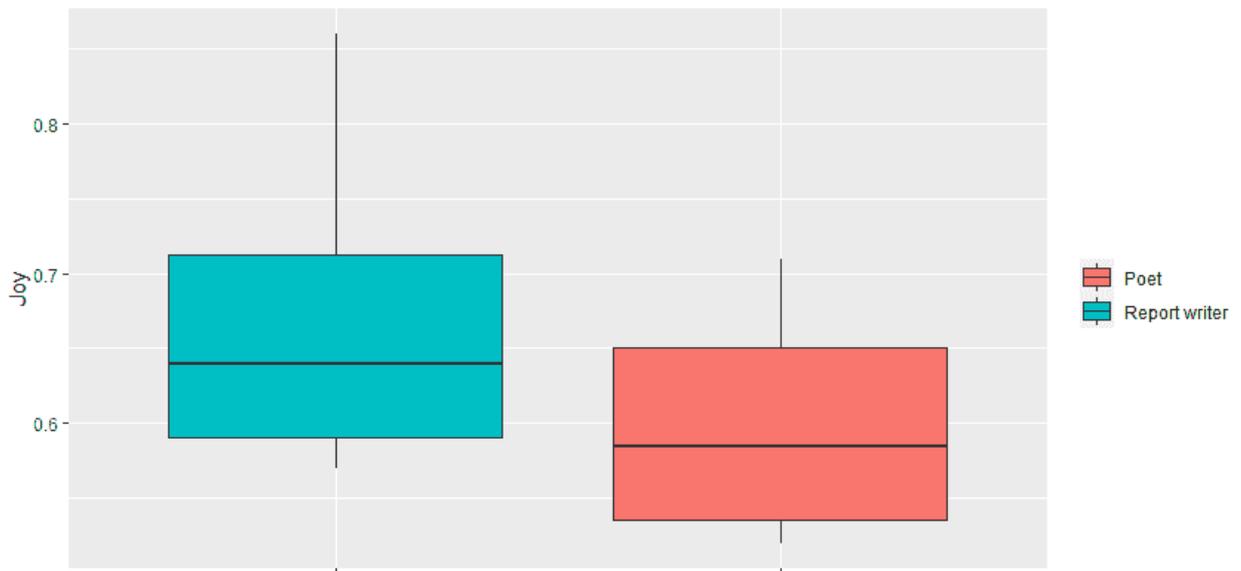


Figure 5.8 Joy sentiment scores for Poets and Report Writers

This would appear to show that Poets showed more Joy (or used more emotive positive words).

Plot – Sadness (Poet, Report Writer)

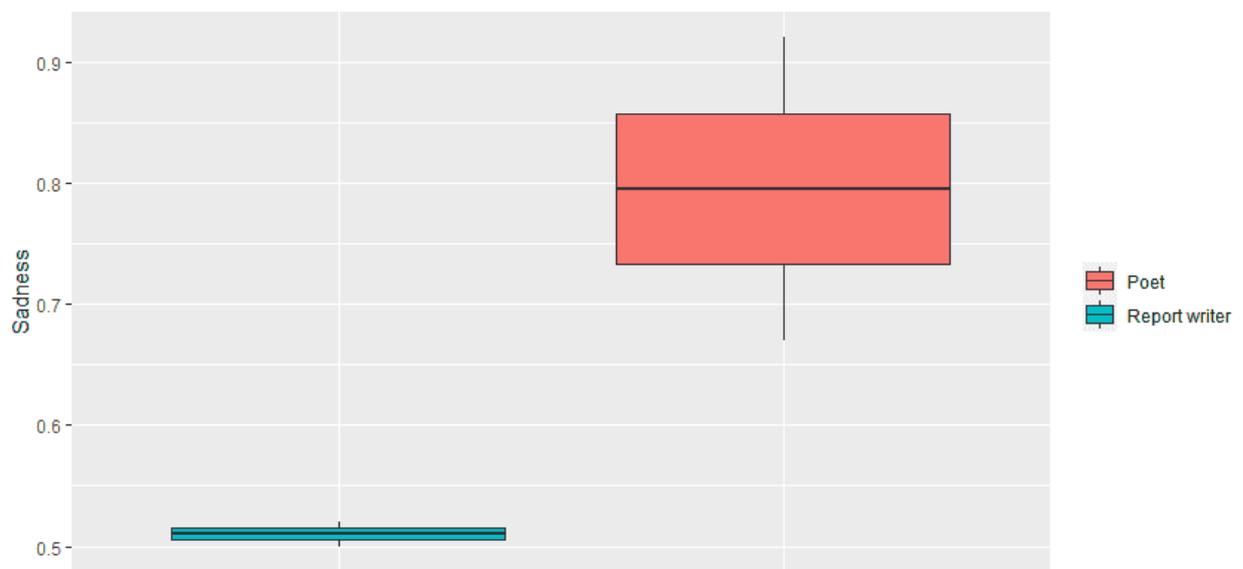


Figure 5.9 Sadness sentiment scores for Poets and Report Writers

This would appear to show that Poets showed more Sadness (used more sad words).

However these results were not statistically significant. Additionally, the result might be because of the extended lexicon of a poet, compared to a report writer.

Table 5.6 - t-test for Joy/Sadness (Poet, Report Writer)

Response variable	Groups	p-value	Conclusion
Joy	Poets and	0.30	Can not reject H_0 hypothesis that average values in two group for variables Joy and Sadness are equals.
Sadness	Report Writer	0.21	

Distribution of Positive and Negative for Sentiments

From the Tone categories, I grouped 'Joy' as positive, and ('Anger', 'Fear', 'Sadness') as negative. Analytic, Tentative and Confidence were neutral.

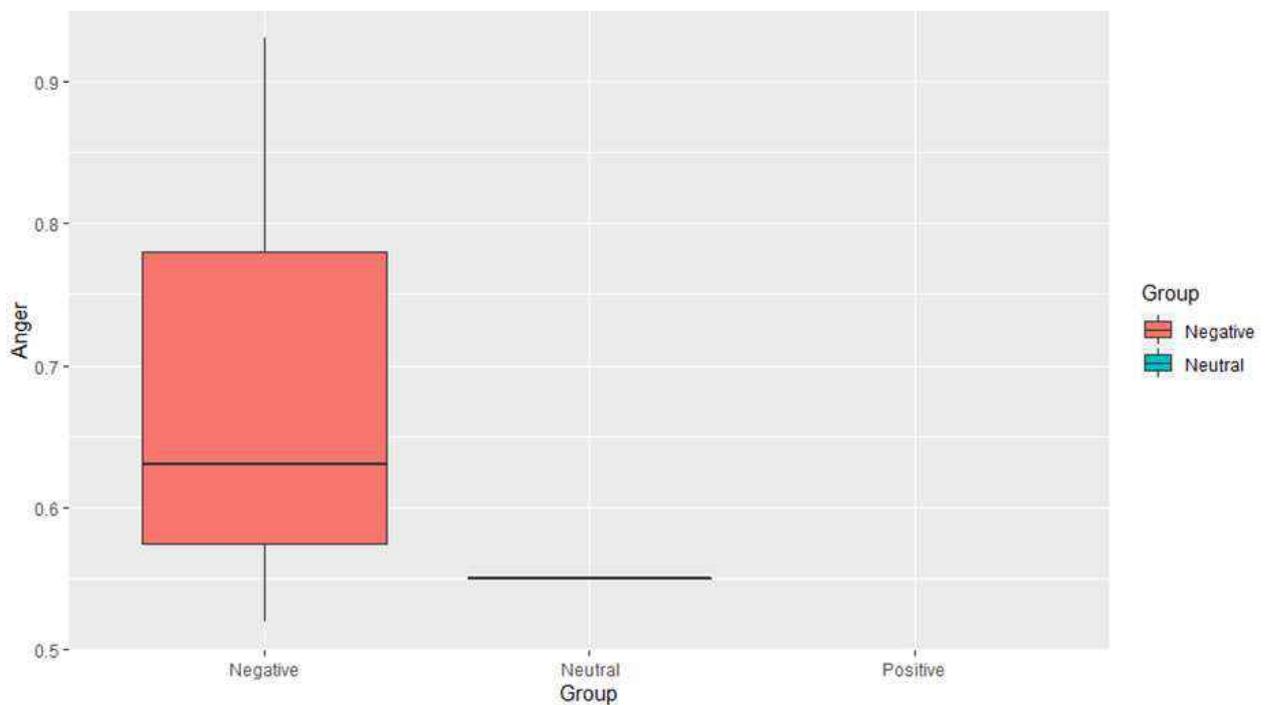


Figure 5.10 Distribution of Positive and Negative scoring against Anger sentiment

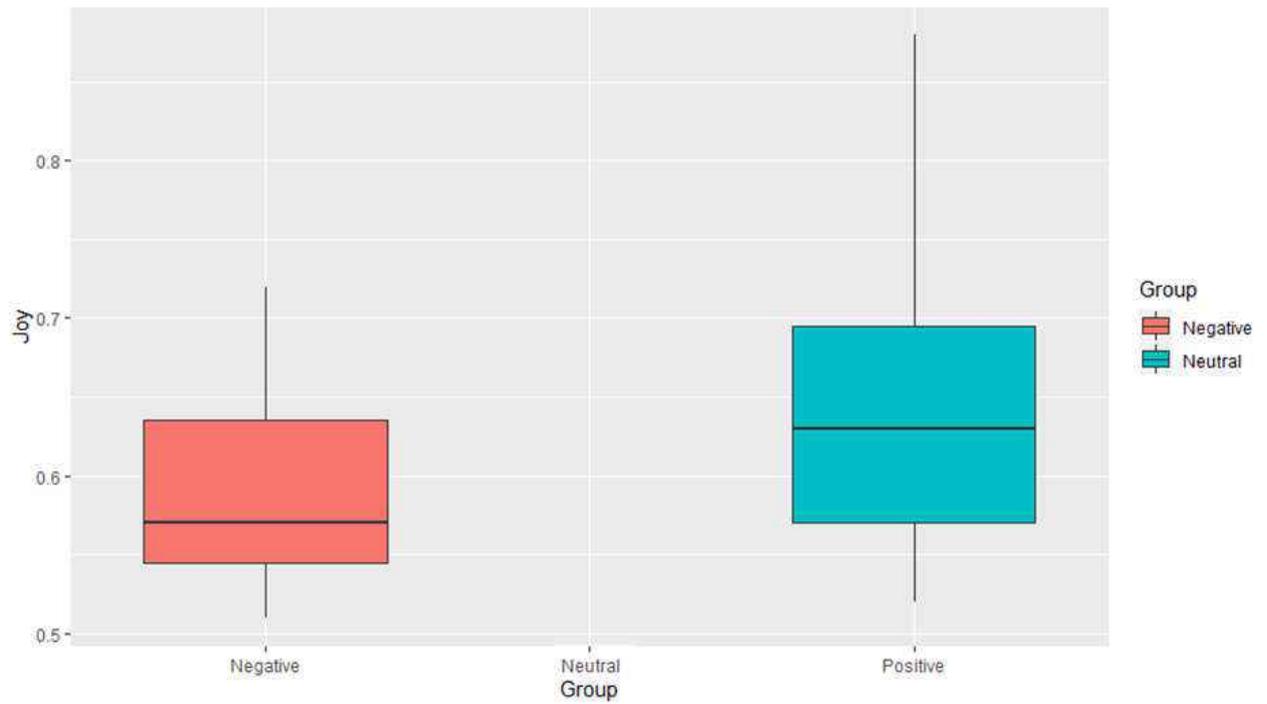


Figure 5.11 Distribution of Positive and Negative scoring against Joy sentiment

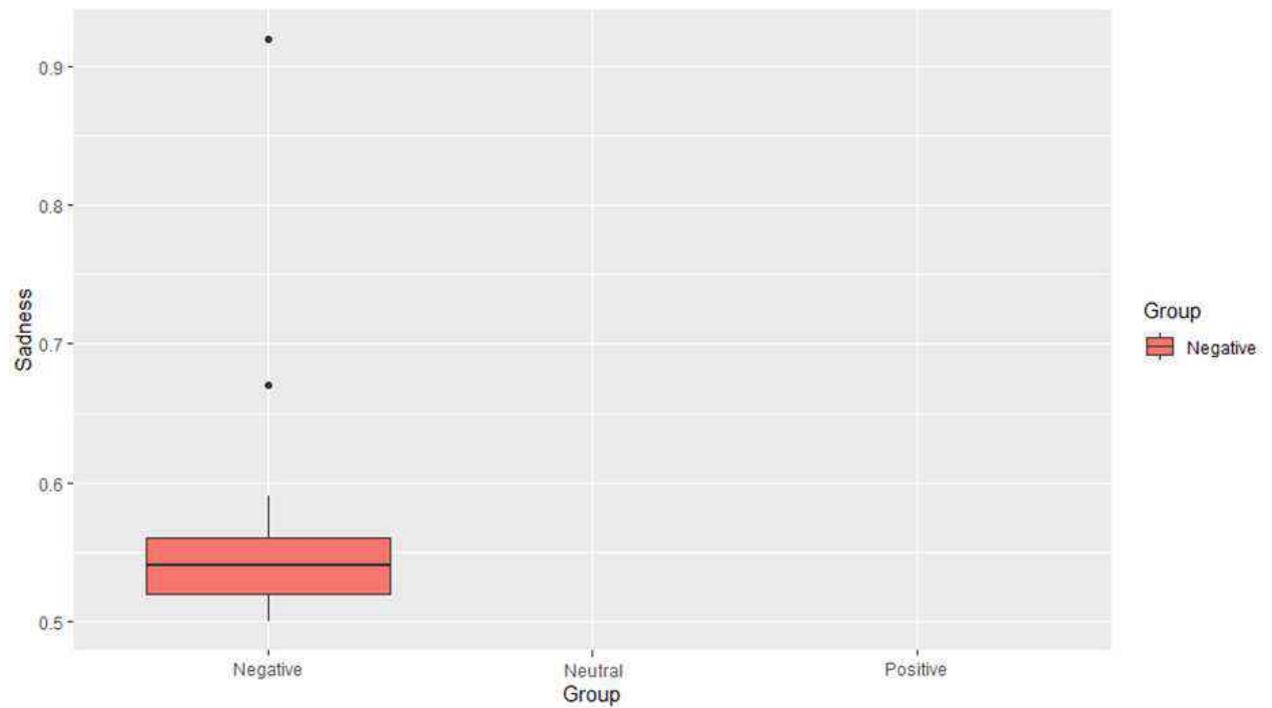


Figure 5.12 Distribution of Positive and Negative scoring against Sadness sentiment

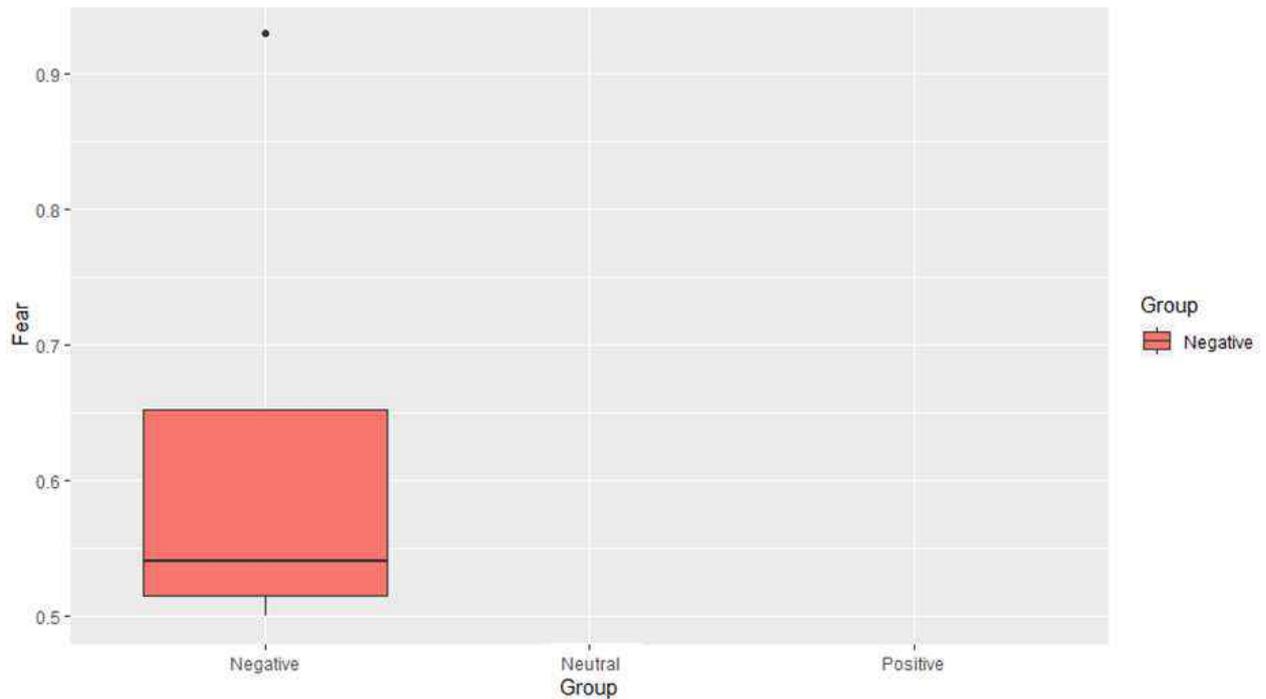


Figure 5.13 Distribution of Positive and Negative scoring against Fear sentiment

Summary

Illustrations of distribution using box plots clearly show that the negative and positive grouping by scores are reflected in Tone sentiments of Anger, Joy, Sadness and Fear. Result of *t*-test shows there isn't a difference between average values for Sadness in Positive and Negative groups.

Significance

ANOVA test. For checking differences between average values for different group of occupations we use one-way ANOVA. Results of calculations we can see in the table below.

As we can see all p-values of ANOVA tests are higher than significance level $\alpha = 5\%$.

Thus, for all *t*-tests from 1 we get same results. ('Confidence' too low to score so is not in this table.)

Table 5.7 Results for all t-tests from 1 are showing higher than significance level

Confidence had a null value (meaning under 0.5 in the Tone Analysis scale, so not significant) in all respondents.

Variable	p-value for ANOVA test	Conclusions
Anger	0.735	In all cases p-values of ANOVA tests are higher than significance level $\alpha = 5\%$ (0.05). Hence, we can not reject null hypothesis about equality of the average values of variables in all 6 cases (Anger, ...)
Fear	0.715	
Sadness	0.453	
Joy	0.796	
Analytical	0.23	
Tentative	0.619	

Sentiment tones by Experiment

Caption experiment

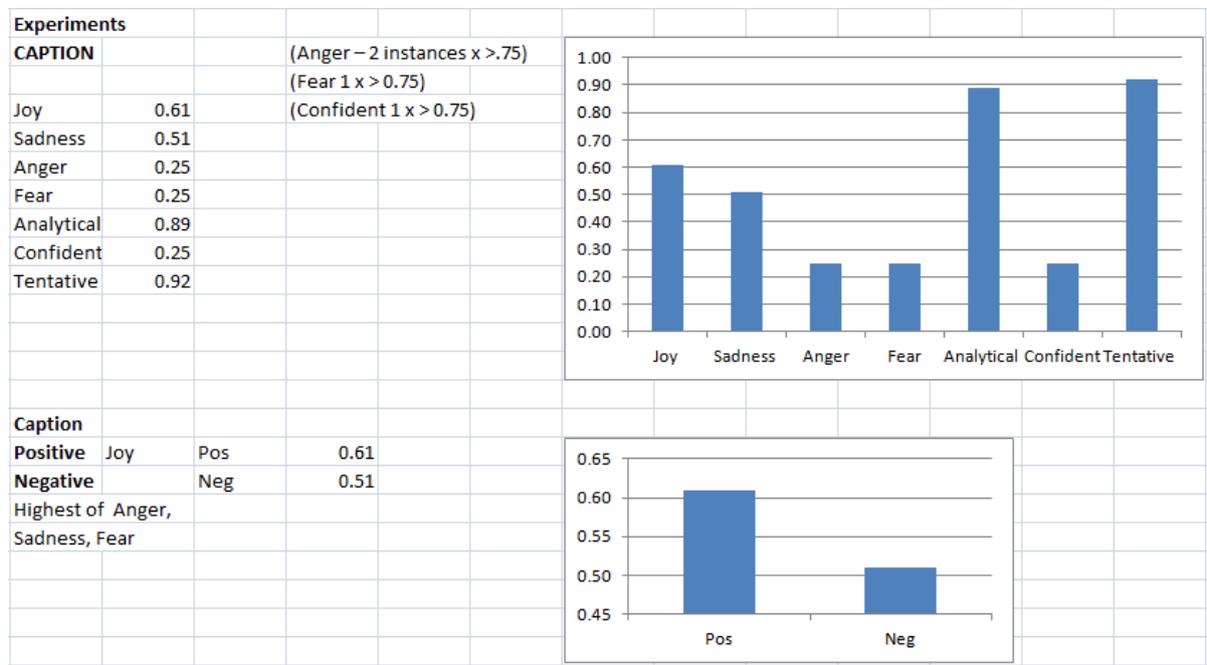


Figure 5.14 Caption experiment: tone sentiment analysis overall respondents

The most common tones are Analytical and Tentative, which is to be expected in responses to a new method of working. It shows serious engagement with the questions raised by the study. Positive (Joy) scores 20% more than Negative (average of Anger, Fear, Sadness). Sadness is the strongest of the negative reactions, with about double the scores of Anger and Fear.

News experiment

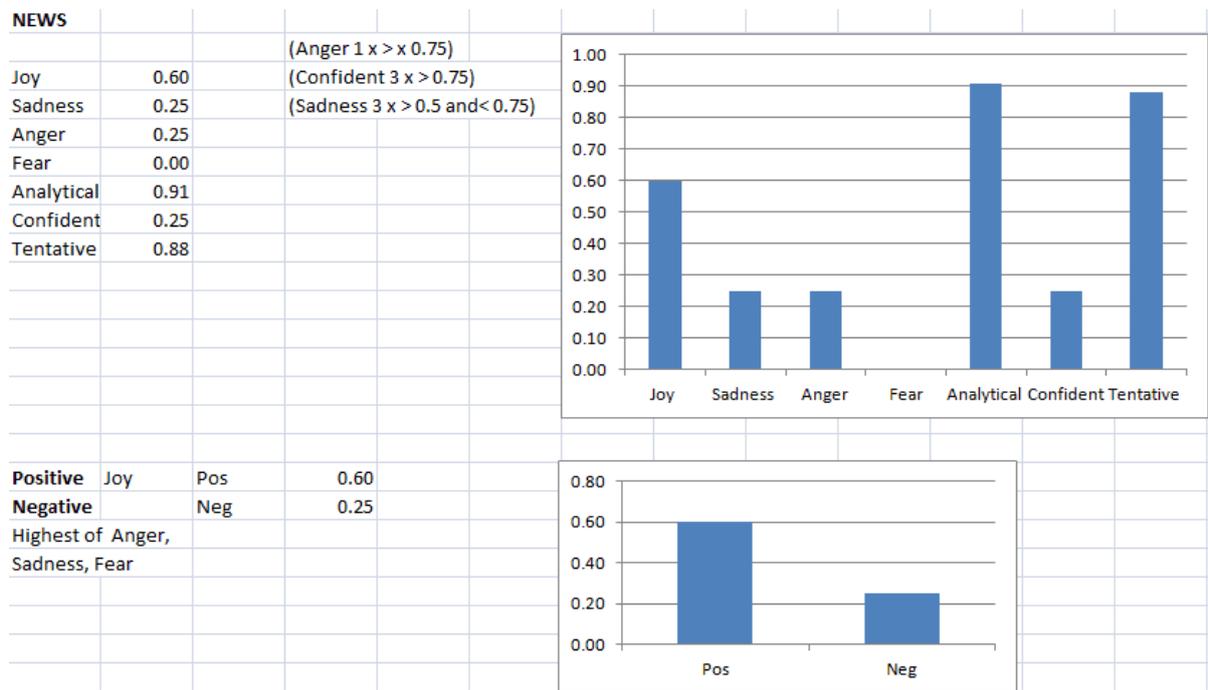


Figure 5.15 News experiment: tone sentiment analysis overall respondents

These charts show a more overall positive result than Caption experiment. Fear has gone, low levels for Anger and Sadness tones. Positive is about 42% more than Negative. This is perhaps due to familiarity with the experiment interface and the output of the generator.

Fiction experiment

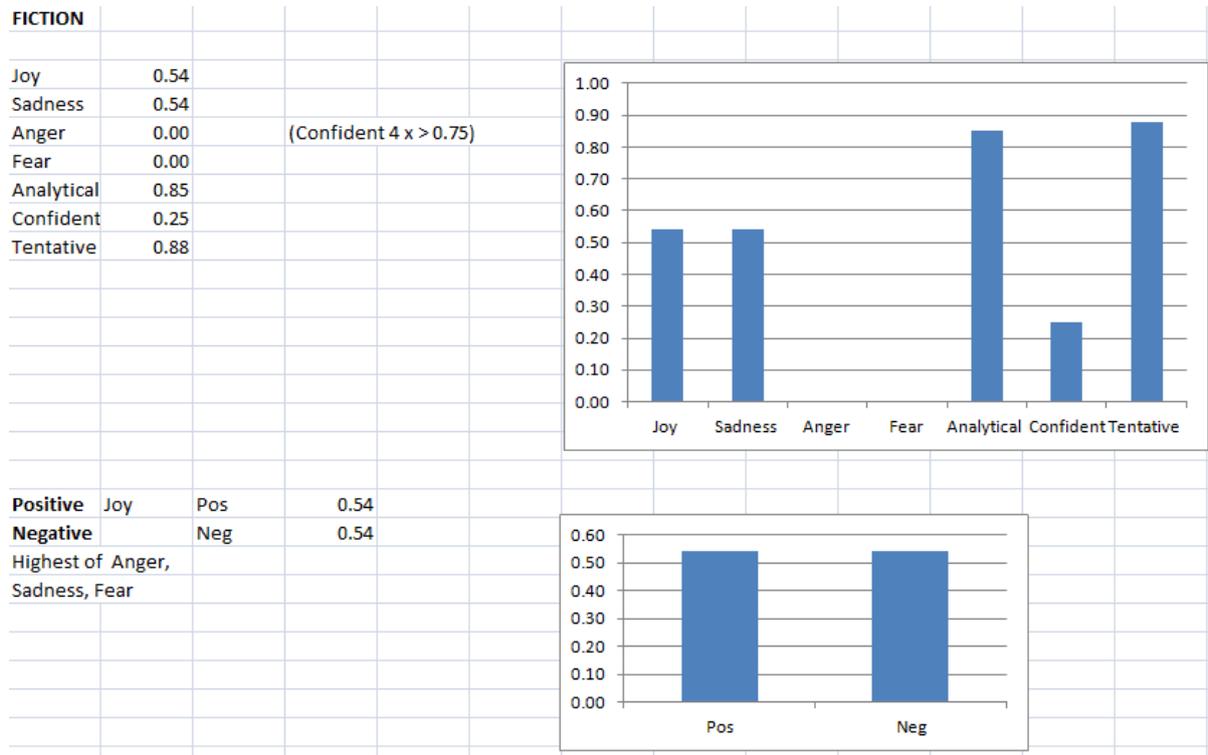


Figure 5.16 News experiment: tone sentiment analysis overall respondents

By the third experiment, Anger and Fear have gone, leaving only Sadness.

All three experiments show a low score for Confidence, showing the degree of doubt that using a considering the questions has produced.

6 Feedback questions after the three experiments

There were six further questions after the experiments.

Likert score

There was an initial Likert scale 1-5 scored response, Strongly Agree (1) to Strongly Disagree (5).

Question 1 ‘Did you enjoy using the text generator panel’ had a median score of 2, Agree.

Feedback comments (Text replies) were analysed to gauge responses and sentiment analysis was also conducted.

Sentiment analysis

IBM Tone, Watson service, has the following scoring (IBM Watson, 2020):

None 0 to < .5 = none or neutral score

Medium .5 to .75 = medium score

Strong Greater than .75 1.0 = high or strongly relating score

Question 1/6, with two feedback questions

‘Overall, did you enjoy using the text generator?’

1 Like a lot/2 Liked a little/3 Neutral/4 Slightly disliked/5 Strongly disliked

‘What was **most interesting**? Please explain.’

‘What was **least interesting**? Please explain.’

Q1 Enjoy and Least Enjoy – Summary

Table 6.1 Most Interesting scored feedback texts

<p>Most interesting – 26 replies Stimulating = 10 Ideas/irrelevant/surreal = 7 Responsive = 5 (Comment mentioning specifics = 5)</p>

Table 6.2 Least Interesting scored feedback texts

Least interesting – 21 replies
 Stimulating = 0
 Irrelevant/random = 11
 Responsive = 0
 (Comment mentioning specifics = 11)

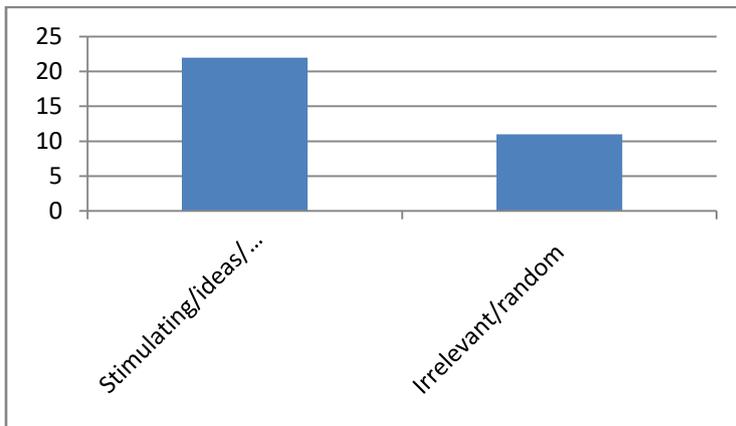


Figure 6.1 Q1 Most and Least Interesting responses, scored by stimulating or irrelevant

This shows an overall positive disposition.

Question 1 Enjoyment

“What was 1a Most and 1b Least enjoyable.”

1a / Most interesting – 26 replies

Sentiment analysis

Table 6.3 Sentiment analysis of feedback comments from Q1 Most Interesting

Sentiment analysis (IBM Tone Analysis)

Joy 0.63 Medium
 Analytical 0.91 Strong
 Tentative 0.93 Strong

(Anger, Confidence, Sadness, Fear < .5 None)

Human analysis

The response text was read and scored for keywords and phrases. There can be several scores per reply.

From the 82 respondents, these 26 replies were respondents that had completed the experiments and local feedback, and then moved on to the final questions.

Table 6.4 Human analysis of Q1 Most Interesting 26 feedback comments

Total analysed = 26
Liked, positive comment = 20
Disliked, negative comment = 2

Comment mentioning specifics such as:
 generation speed or interface = 11
 stimulating = 8
 new ideas/ words = 7
 responsive = 5

1b / Least interesting – 21 replies

Table 6.5 Sentiment analysis of feedback comments from Q1 Least Interesting

Sentiment analysis (IBM Tone Analysis)

Joy 0.50	Medium
Sadness 0.54	Medium
Analytical 0.84	Strong
Tentative 0.87	Strong

(Anger, Fear, Confident <.05)

Human analysis: replies scored on keywords and phrases, can be several per reply. These comments were more specific. Comments mentioned specifics.

Table 6.6 Human analysis of Q1 Least Interesting 21 feedback comments

Total analysed = 21 Comment mentioning specifics such as: generation speed or interface = 13 irrelevant or repeating text = 9 boring= 3

Summary

The 13 comments about the general interface, slowness of generation, page layout, etc . (including 4 about the prompt image) out of 25 comments (52%).

This shows engagement with all aspects including design of interface. Suggestions for improving the interface were greater than complaints about the generated text.

Feedback Comments

‘Most interesting’ aspects were stimulation (31% of comments used this or similar words).

‘New ideas’ was a common response (28% of comments). Nearly a fifth thought the text was responsive to the text they entered (19%).

The ‘least interesting’ question produced more specific replies, and complaints about the randomness of the generated text. This might be due to high expectations. Only 10% had used any generators before, often old systems, so this was a naïve cohort. There was also a learning curve, as some people mentioned that the second experiment (News) was better than the first, and that it got easier as they worked through the experiments, showing that their expertise increased with familiarity.

Sentiment of feedback comments

Sentiment tone analysis of the feedback text only showed a medium level of Joy in the ‘Most interesting’ question, and medium levels of Joy and Sadness in the ‘Enjoy least’ question. No

Fear or Anger scored. As usual in the experiment, Tentative and Analytic sentiments were high and Confidence was too low to score.

Question 2 Emotions and Feelings

Q2 “I felt emotional when the generated text appeared. Which emotions did you feel, please describe?”

Likert score

There was an initial Likert scale 1-5 scored response, Strongly Agree (1) to Strongly Disagree (5).

Question 2 ‘Did you feel any emotions when the generated text appeared?’ had a median score of 3, Neutral. This shows in the detailed analysis.

Text replies were analysed to gauge responses and sentiment analysis was also conducted.

Sentiment analysis (IBM Tone Analysis) showed Anger, Joy and Sadness but not Fear.

26 replies were scored for various emotions. More than one emotion can be scored per reply.

Table 6.7 Human analysis of Q2 Which emotions 26 feedback comments

Total analysed = 26
Comments mentioned specifics:
no emotion= 9
excited/interested /curious= 7
amusing= 5
boring/frustrating= 5
disappointed/negative= 5
satisfied/pleased= 3

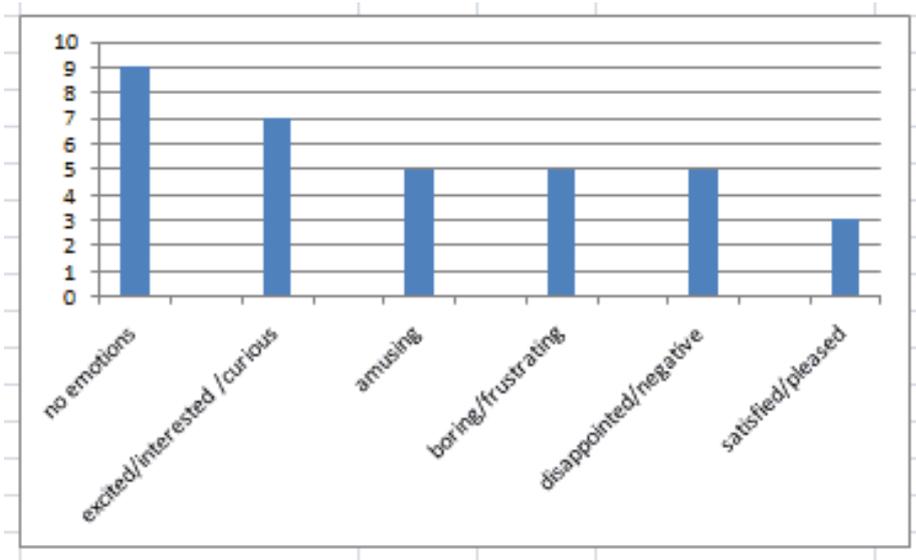


Figure 6.2 Q2 Which emotions question: detail of feedback feelings showing number of comments

Table 6.8 Human analysis of Q2 Which emotions 26 feedback comments

Total analysed = 26 Comments mentioned specifics: no emotions= 9 positive= 15 negative= 10
--

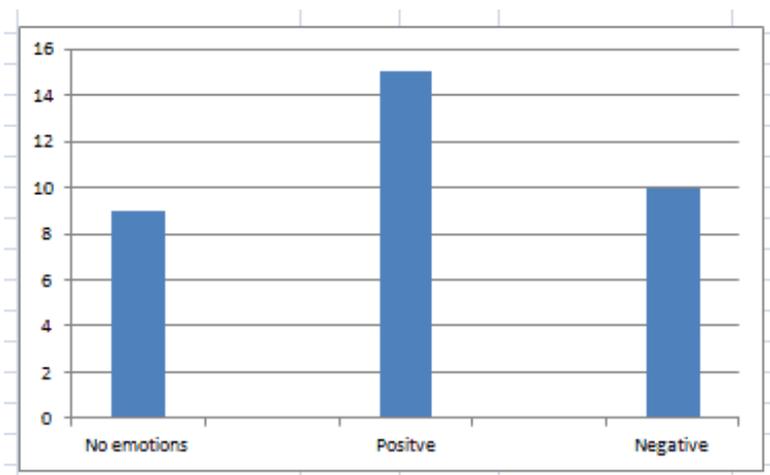


Figure 6.3 Q2 Which emotions question: positive and negative responses

Two thirds of people (64%) were emotionally involved in their text generation experiments.

Many emotions were raised. A third (36%) professed no emotion, or indicated none but

added a named feeling. Overall, 57% of respondent emotions included positive comments. Only 29% were negative.

Q3 and Q5: Plagiarism and ownership of text

Q3: Do you feel that you have used somebody else's work?

Q5: Do you think you could sell this as your own work?

These related questions were separated to prevent similar responses. One was Likert scale reversed for the same reason.

These questions can be paraphrased as 'is this yours, could you sell it?' and elicited strong responses in the feedback, which is not reflected in the median Likert scores.

For more analysis of Questions 1 and 2, please see section 5.2.

Q3 Somebody else's work

Do you feel that you have used somebody else's work?'

The word *somebody* was used to avoid suggesting that the text generation was in some way robotic or mechanical, when in fact it is a simulation of written text from an aggregated corpus of anonymous writers.

Likert score

Strongly agree /2 Agree/3 Neither/4 Disagree/5 Strongly disagree

Likert score of 4, Disagree.

Table 6.8 Q3 Somebody else's work, does this seem relevant and why

Yes	6
Neutral	2
No	19

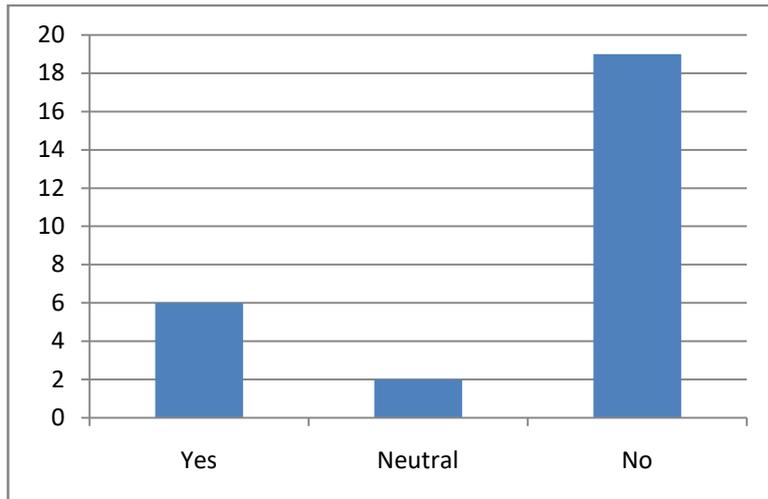


Figure 6.4 Q3 Someone else's work

'No' means respondents felt they owned the generated text (70%).

Q3 Sell as own work

Likert score

Question 5 'Do you think you could sell this as your own work?' had a median score of 3, Neutral.

'Do you think you could sell this as your own work?'

Strongly agree /2 Agree/3 Neither/4 Disagree/5 Strongly disagree

Table 6.9 Q5 Do you have any comments on notions of authorship, plagiarism, ownership and intellectual property

Own	15
Neutral	7
No, don't own	2

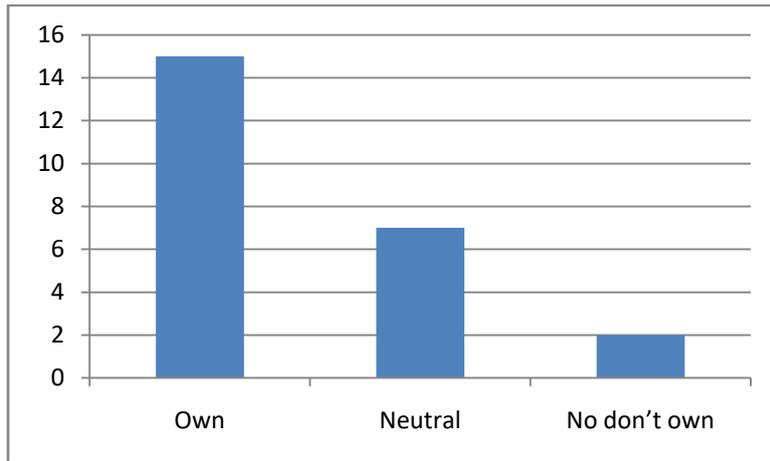


Figure 6.5 Q5 Sell the generated text as own work

‘Own’ means respondents think they own the generated text (63%).

Sentiment tone analysis

IBM Tone Analysis of Q3 and Q 5 – intellectual property

Q3 ‘Somebody else’s work’ and Q5 ‘Can sell as own’ replies had the same sentiments. In agreement with other sentiment gradings, Analytic was high (over 0.9), Confidence was low or none (under 0.5). In these aggregated texts, Joy was medium high (0.6). Others did not score.

So, regarding ownership, writers are positive about the opportunities available through computer assistance.

Respondents felt they owned the generated work with an average of 67%.

Table 6.10 Sentiment analysis of feedback comments from Q3 and Q5 Plagiarism and ownership

Sentiment analysis (IBM Tone Analysis)	
Analytic 0.9	Strong
Joy 0.6	Medium
Confidence <0.5	None or low
Anger, Fear, Sadness 0	None

Question 4 Use in word processor

“Imagine a generator in your ideas editor or word processor. What extra features would you like?”

Likert median score was 3, Neutral (or balanced). The data shows that this question polarised opinion, with count of Agree only slightly more than Disagree, and few neutrals. There were no score 5, Strongly Disagree with use.

One comment suggested that the experimental system (Story Live online tool) was sufficient, and wanted to keep the generating of ideas away from the production tool.

Table 6.11 Sentiment analysis of feedback comments from Q4 Word Processor inclusion

Total comments = 33
Positive comment = 4
Neutral comment = 4
Neg comment = 12
Ideas for better use/future = 13

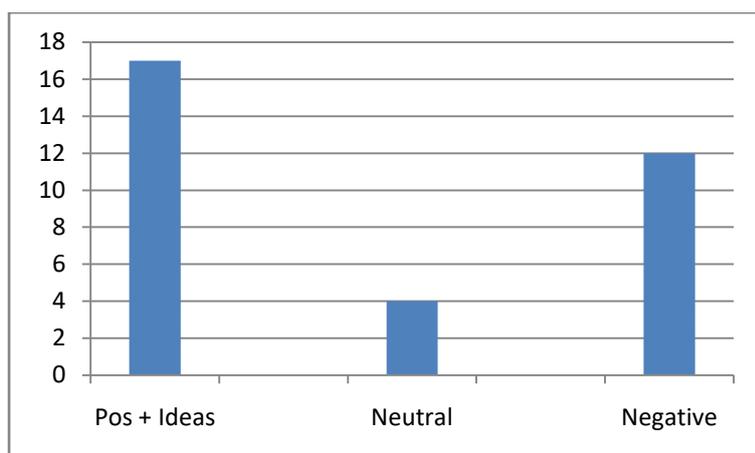


Figure 6.6 Q4 Use in Word Processor question: positive and negative responses

Positive comments: suggesting improvement ideas indicates interest in using it in a word processor. All of the people that left positive comments also included ideas for improvements. Generally did not discuss integration, but adding specific features to the generator and editor.

Negative comments: only one of twelve negative comments had any ideas for improvement (8%). The most common negative idea was that use of a machine would interfere with or reduce human creativity. The comments indicate a fear of loss of agency regarding their professional activity, a deskilling, and possible replacement of human by software.

Q4 Use in word processor – Creativity Support Tool CST comments

Use inside a word processor

“It's very straightforward to copy and paste from an online generator already, and in many ways preferable to doing anything directly within the document I'm actually working on.”

Conclusion: simple interface better.

Learning issues relating to prompt syntax

“The outputs were quite random and hard to find something relevant to the prompt without going through a few cycles.”

“Some parameter to control the output better.”

“It would be good if it could generate at macro level (eg narrative plan) rather than just micro (sentences or paragraphs).”

Conclusion: examples or more practice needed, as prompt can be written at any story level, eg, synopsis or dialogue.

Enhancements – Search

“Access to databases with relevant information.”

“I'd welcome a research AI assistant. Someone/something I could ask to search the world's libraries.”

Conclusion: search integration, and other writing tools, e.g. synonyms, mentioned.)

Enhancements – Prompt auto-programming

“Can it more easily generate relevant questions in order to make its response more relevant?”

Conclusion: analyse text generation to create new user-selectable prompts. This has been realised in the newer GPT-3 and GPT-J systems.

“Add an intelligent learning system for giving +1 to the stuff I like and -1 to the stuff I don't... to avoid repeating themes and phrases that I don't like, and to guide the text generation more actively.”

Conclusion: copy user selected text to prompt.

“Build an archive from my previous work to more accurately accommodate my personal style.”

“An option to pick the genre or style.”

“Input in names of characters, settings and the type of plot, or genre.”

Conclusion: longer prompts, can be done in OpenAI GPT-3 with ‘prompt programming’ (Reynolds, McDonnell, 2021).

Q6 Final questions and 'fake' news comments

Q6-1: 'Please comment on anything else you noticed.'

This question produced mostly positive comments (78%), with ideas for interface and output improvements. No negative comments, possibly as these had already been given in more specific questions.

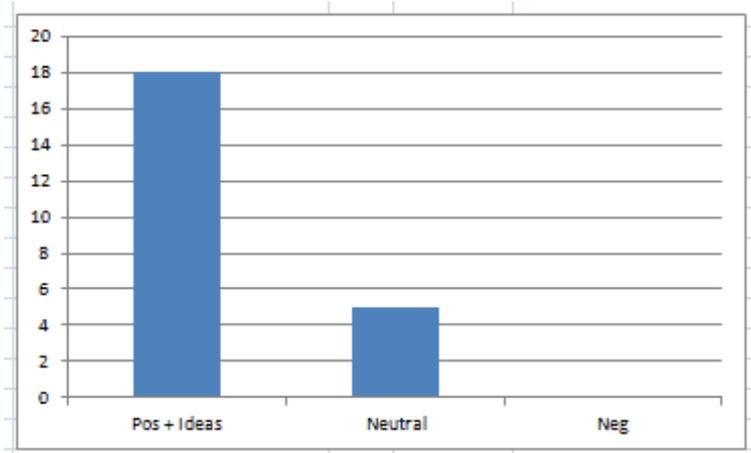


Figure 6.7 Q6 Any other comments: positive and negative responses

Q6-2: 'If you have heard of 'fake news' do you think this is relevant?'

This was the only mention in the study of a real-world scenario for text generation.

18 of the 26 respondents that started the final question section answered this last question, 69% of those people. It was a question not specifically to do with the experiment. This shows an awareness of the use of new software to create 'fake news' in 18 of 26 replies, over two thirds (68%).

The question had neutral phrasing so as not to influence replies.

Comparing answer word count with positive – negative ratings

The actual Q6-2 feedback text was assessed for Positive – Neutral – Negative response to the 'relevant to fake news' question. Anything ambiguous or with conflicting statements was

rated Neutral. For the purpose of ranking, the following process was applied: score is 0 (positive), 1 (neutral), 2 (negative). This was to get a suitable visualisation in the chart. Word count was also used in the analysis.

As examples, to the question ‘If you have heard of ‘fake news’ do you think this is relevant?’

‘No’ is 2 (score value) for negative with word count 1,

‘I stay out of these type debates’ is 1 for neutral with word count 7,

‘Of course’ is 0 for positive with word count 2.

These are scaled up so can be seen against the word count values. So no red column is positive score 0 (yes, relevant to ‘fake news’), high red is negative score 2 (not relevant to ‘fake news’).

The red bar 0-1-2 value is matched with ranked word counts to make it easy to visualise. So a tall red line means negative for relevance to fake news.



Figure 6.8 Overall summed and ranked Likert scores (blue) with ‘fake’ news relevance (red)

Figure 6.8, above, shows no relation between Likert scores (blue) or positive or negative regard, with ‘fake’ news relevance (red).

In Figure 6.8 all Likert scores are summed per respondent (from Caption feedback to Question 6, reversing one question that scored reverse to normal) and charted against whether they think it is relevant to ‘fake’ news generation (positive - negative).

Low scores for summed Likerts means agreement with the various questions (positive), high scores means disagree (negative).

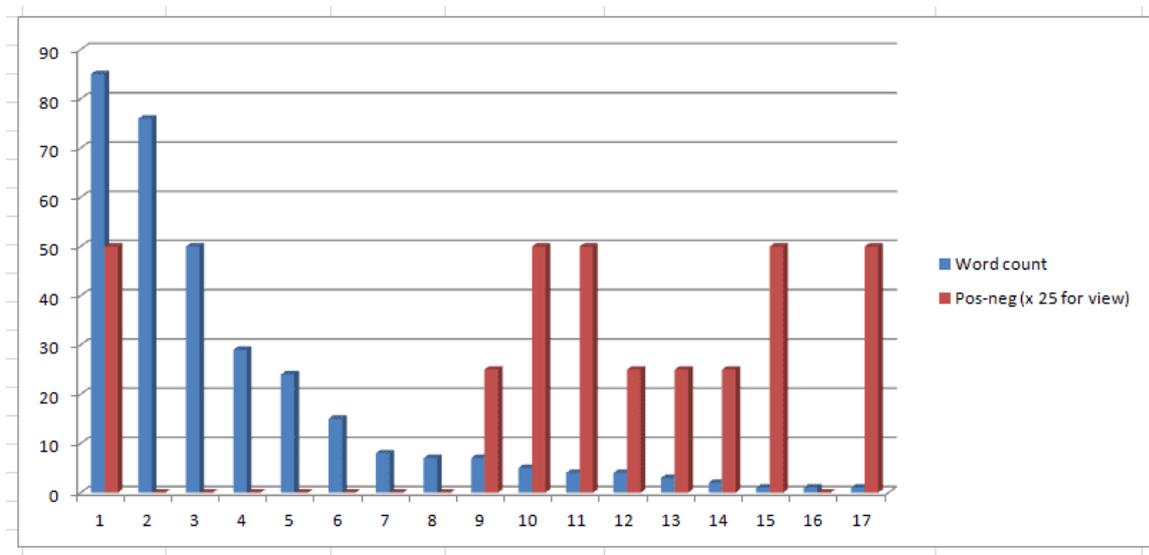


Figure 6.9 Word counts of feedback against ‘fake’ news relevance

This chart is of Question 6 ‘fake’ news reply word counts against score for ‘fake’ news relevance.

This shows that people, who have a lower general opinion of text generation, don’t think text generation is relevant to ‘fake news’. This is possibly because they think it is not capable or good enough to be of use for any guided use. At first glance this shows that higher feedback word counts are associated with mostly positive (only one tall red negative) for ‘relevance to fake news’.

Low word counts are associated with negative for ‘relevance to fake news’.

There is only one negative out of the top half of the responses (columns 1-9 on the chart above, with 2-8 pos=0). The highest word count (column 1 on chart) was from an answer that discussed at length an actual example of real ‘fake news’, or propaganda, in the Syrian War, so was a longer than usual negative response.

This shows that lower Likert scores (which means positive response) relates to lower Pos-neg scores (which means positive relevance to ‘fake news’).

Word counts - Mann Whitney U test

The z -score is 4.25539. The p -value is $< .00001$. The result is significant at $p < .05$. People who used the least words thought text generation was less relevant to ‘fake news’. People who used more words in their answer thought it was more relevant to ‘fake news’.

‘Fake’ news and occupations

A visible effect in boxplot of the data (figure 7.7 below) is that the two highest word counts (actually widest main spread, Academic has a higher outlier) were produced by Artists and Report Writers. The only connection might be that Artists write grant application forms.

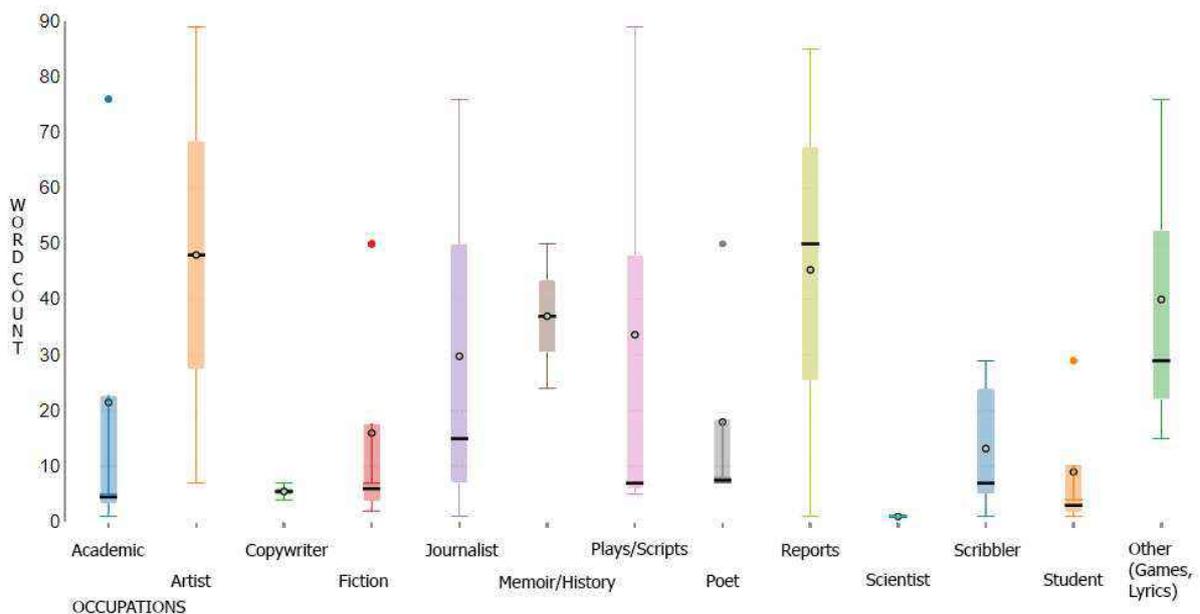


Figure 6.10 Values for word count of answers by Occupations to Relevance to Fake News

Positive-negative rating

Zero 0 is positive for ‘text generation is relevant to fake news’, 1 is neutral, 2 is negative.

Averages are used for the sake of ranking only, as this is an ordinal scale.

Creative writers and fake news

Occupations were graded ‘creative’ or ‘non-creative’. This is not very accurate, as there are dull poets and innovative scientists (and such labels are subjective), but it is useful for making a comparison. See below left graph legend.

The sequence is (n non-creative, c creative):

(score 0) n n c c n c n c c n n (score 2).

No effect from ‘creative’ or ‘non-creative’ roles is visible.

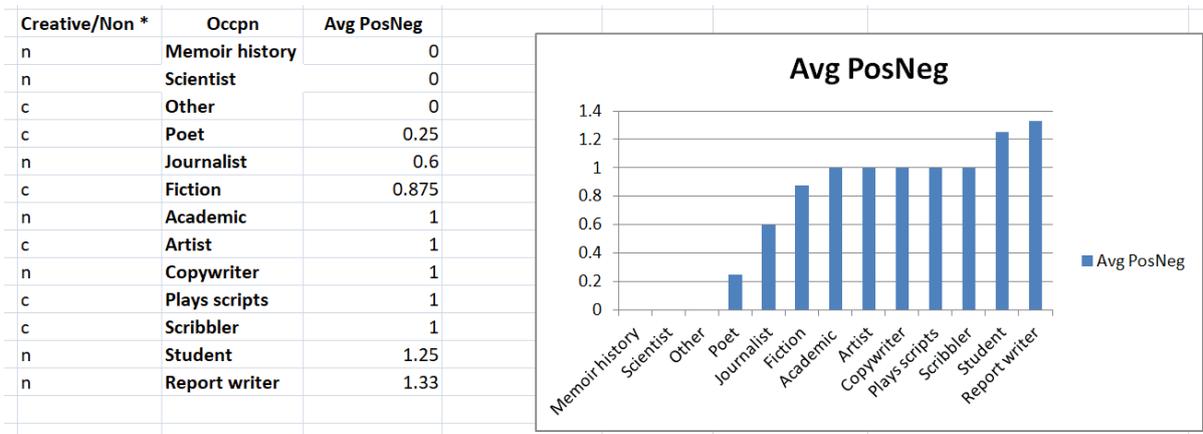


Figure 6.11 Values for Pos-neg ranking for sentiment of Relevance to Fake News answers showing Occupations

7 Summary of findings

Highlights

- 126 text generator experiments conducted in controlled system
- 90% of 82 writers had never used a text generator
- Fast learning curve over the course of three text experiments
- 71% had positive or neutral feelings about the experiments
- Non-creative writers (e.g. Copywriters) were more positive than creative writers (e.g. Poets)
- Two thirds of respondents felt they owned the generated work
- Many suggestions for improving the generator and editor in a new AI creativity support tool
- 'Fake' news awareness depends on engagement with technology

Experiments - Caption, News and Fiction writing tasks

Respondents were asked to write a short caption or meme using the text generator and editor.

Nearly all ignored this and wrote at length, showing engagement and enthusiasm for the task or perhaps enjoying using a text generator for the first time.

News stories were only slightly longer, showing continued enthusiasm for the task.

Fiction writing produced a much longer response, showing serious engagement with the new creativity tool.

Editing of the generated text

The results showed that if the text was edited, the respondent would continue with all three experiments. This indicates two main groups of future users, people who want to edit the text and produce hybrid works, and people who are not interested in editing and would only view

or use unaltered the generated text. These two groups could be termed co-creators and producers (or ‘casual’ creators, as the prompt is their personal start point).

The latest version of this particular generator OpenAI GPT-3 (2020), and the newer Open Source GPT-J (Eleuther, 2021), generate more or less human indistinguishable text, although it still produces many errors such as loops and non-sequiters. Results are cherry-picked for publication. Perhaps this non-editing producer ‘casual’ group would be more likely to use GPT-3 to generate text and use it without changes. The hybrid co-creator group use the more erratic and inspiring GPT-2 text, and improve it by editing, or use it as a springboard for further creation. This is relevant to the design of a creativity support tool.

From the results, creative writers would prefer a more surprising or random generator for inspiration and non-creatives prefer a more reliable and stable generator for reading or as production copy.

Sentiment

Positivity halved during the course of the three experiments, and negativity plus neutrality doubled. This might indicate a positive novelty effect, followed by an increasing awareness of the limitations of text generation. In view of the longer word counts, the respondents were still engaged even though their enjoyment had diminished, perhaps due to reduced expectations.

Creative and Non-Creative differences

Generally, more time was spent on the short story experiment, along with a higher word count, which is to be expected. Caption and News experiments took about the same time, as

is reflected in the similar word counts.

Non-creatives took longer to do the Fiction experiment, perhaps due to unfamiliarity with the task.

Non-creative writers were consistently slightly more positive (14%) than Creative writers about using text generation, both at the time of the experiment and after reflection in the final questions. Non-creative writers were also slightly more positive about using and owning the generated text (7%). Values were very close to neutral and generally show a small positive effect.

Creative writers were negative about using a text generator in a word processor or ideas system. Other groups, including mixed and unstated occupations, who might see the practicability rather than the threat, were slightly positive. Copywriters scored the highest for positivity. As they write impersonal text to order they can immediately see the usefulness of generated text.

Respondent feedback comments

Creative writers produced more volume (73%) of feedback comments than Non-Creatives. This might be due to having more spare time in which to do the experiment. Creative writers might have more curiosity for generative text, as it could potentially transform their workflow with an accessible controllable source of unexpected ideas or concepts. Non-Creative writers might only see a simple way to instantly generate more useful copy on a given topic, which would generate fewer feedback comments.

Sentiment analysis of the feedback comments

Analytical and Tentative sentiments were the highest, which shows that respondents were careful and serious in their responses. Confidence was low, to be expected with a novel

experiment using a generation technology 90% had never used before. This also shows the accuracy of the IBM sentiment analysis.

Negative sentiments (Anger, Sadness, Fear) were higher for Creative writers, who showed more negative than positive sentiments. Other groups showed an overall more positive sentiment.

Across the three experiments, higher positivity declined until the last Fiction experiment, when positivity and negativity sentiments had the same level. This shows an ambiguous rather than neutral position. Fear and Anger were present in the first Caption responses, but had gone by the last, showing familiarity. Sadness increased in the last experiment, perhaps indicating worries about professional status, or disappointment with the new technology.

Questions after the experiments

Most people (64%) were emotionally involved in the experiments. Most stated emotions were positive (60%).

People most liked the stimulation of a responsive ideas generator (38%), rather than expecting the system to be able to write coherent relevant text.

Most negative comments (52%) were about the text generation itself, with only some about the design of the user interface.

The question regarding use in a word processor gathered many suggestions, showing a high level of engagement. These comments are useful for further development of a creativity support tool.

Most respondents thought they had used someone else's work, which is literally true as the generator was trained on public domain text. However this was not seen as a barrier to selling work, as most respondents were neutral (unsure) on this issue, as they are not in a position to comment yet. As someone said, "The generator itself cannot claim royalties". Although now

these systems are often available on subscription. The money goes to the developers of the software, rather than the very large number of people who unknowingly donated their text to the training data via their various writing activities on the internet.

Some others were more negative, believing in their own human creativity – “That text wouldn’t have come from me, but from AI. That, to me, is cheating”.

Sentiment analysis of the ownership texts shows an elevated ‘Joy’ score. Regarding ownership, writers are pleased with the opportunities available through computer assistance. Final comments were very positive (78%).

‘Fake’ news comments

This question was added as ‘fake’ news and false content generation was in the world news at the time. Overall respondents who thought a generator was not relevant to ‘fake’ news had low word counts for all responses, indicating a low engagement, and perhaps were not interested in a long discussion. If text generation was seen as relevant to ‘fake’ news, more words had been used in responses, indicating engagement.

So it would appear that engaged respondents would be more aware of a connection to ‘fake’ news. Positive or negative feelings about the generator (from Likert scores) had no relationship to their opinion, so this was not an effect of low opinion of the tool.

Regarding the generator

Some of the text generation improvement suggestions related to reducing unplanned and unwanted deviation, to keep a character or plot thread. This is not possible with this type of system (GPT-2), as it relies on statistics to generate words in sequence, rather than adding or templating any overall meaning or structure. It often deviates from the ideas in the starting

prompt text, or just repeats similar ideas. Any apparent coherence is statistical.

The respondents perceived the text generation as some sort of creativity or even wilfulness or mind-reading on the part of the generator. Actually it is entirely devoid of motive or awareness. This generator uses the older OpenAI GPT-2 system (November 2019). The latest GPT-3 system (June 2020) can keep up the impression for longer, as it is trained on a much larger text body, but has the same inherent limitations. Further studies with the new GPT-3 system could explore this effect. Long form text generation is the subject of research activity, often informal, by enthusiasts.

Most negative comments referred to losing creative control to a machine. This might relate to the older age range of the respondents (85% over 40 years), as younger people are ‘digital natives’:

“For the present generation of artists, the computer, or more appropriately, the laptop, is one in an array of integrated, portable digital technologies that link their social and working life,” (Taylor, 2014).

There is much high-level discussion of Artificial Intelligence being as important to humans as fire, the wheel and electricity (Rajan, 2021).

“The sensitivity of the wheel-and-axle system to all these factors [weight, strength and friction] meant that it could not have been developed in phases... It was an all-or-nothing structure,” (Wolchover, 2012). This simultaneity effect might be applicable to the current surge in usage, as the factors of computer power, fast networking, large amounts of data plus a lot of data scientists, and a digitally connected public, connect for the first time in a system where the resulting usable artificial intelligence is greater than the sum of its parts.

Comments of design of the creativity tool

All of the respondent comments on the inclusion in Word Processor question had ideas for improvements, as did the final comments feedback, showing a keen interest in future developments. Respondents showed a preference for the simple user interface used in the experimental system, perhaps with some additional services such as search and text-level enhancements.

This is in contrast to the confusingly large range of services in a typical modern word processor. Perhaps text generation is seen a new service, not particularly for addition to word processors, as they are used for finalising documents. This means there might be a demand for a specific AI CST.

The drop-out pattern across the three experiments might indicate that respondents didn't enjoy filling up forms, or using a form-based text generator and editor. Within the tool, text generation can be in-line (inserted at the cursor position) to the active writing, rather than in a separate panel as used in the Story Live tool. In-line generation, where several new sentences are generated from the current text point, and then chosen between or declined, is used in other tools. See HuggingFace in-line demos (2020). Further studies could explore the relative practicality of these user interfaces.

Aesthetics

The general opinion of computer art (and by extension computer writing) has not changed since the early use of computers for artistic production. The public and critics are still surprised and appalled by art explicitly made using computers, preferring the computer to perform an unmentioned and unmentionable production role (Taylor, 2014). The term computer art is only used for marketing, as there is still a ‘shock of the new’ sales effect, which applies to fine art (Obvious Collective, 2018), painting robots (Ai-Da Robot, 2019) and currently NFT CGI animated art shorts (Beeple, 2021). Despite the smoke and mirrors, these works are created and controlled by humans. The Ai-Da art robot even has a human typing the replies in live interviews, Wizard of Oz style, which are then run through a speech generator. It is apparent that the actual Ai-Da robot is the art work, not the outputs, even if they are sold in galleries. The outputs are not art as a human might produce it, but celebrity memorabilia. Perhaps Ai-Da is the first celebrity robot outside of fiction. Her faceless predecessor, Cohen’s AARON software and turtle, produced multiple paintings to be sold at low prices, unlike Ai-Da’s expensive commercial gallery offerings.

Discussions of computer and art revolve around issues from the dawn of computing:

“Pairing the noun ‘computer’ with ‘art’ has in effect built a label with an unending fission, a precarious reaction from joining two seemingly incompatible and oppositional worlds. This discomfort concerning the incongruous combination has in fact permeated all writing on the subject. For many of its detractors, computer art was simply a contradiction in terms,” (Taylor, 2014).

“Literary critics focused on the dehumanizing tendencies of the computer and the perceived ontological break between author and reader. [Critics] praised the importance of the written poem as an essential “communication from a particular human being,” and noted that if the

difficulty of working with the computer discouraged those currently interested, then poems would happily remain “one last refuge for human beings,” (Taylor, 2014).

These two attitudes – ‘use of computer removes artistic value’ and ‘only humans make meaningful art’ – are still common in the feedback replies.

“In a complete turnaround from previous criticism by both humanists and anti-humanists, the computer was increasingly perceived as a technology of rupture rather than an embodiment of the Enlightenment vision,” (Taylor, 2014).

There is an element of fear in the respondent feedback, perhaps based on this perceived rupture to a writer’s carefully constructed and very personal practice.

In previous studies of art and computers, there is a sense that traditional aesthetic standards are undermined by technology:

“On the subject of artificial intelligence as a creative process of expression, the majority of participants [artists and experts] clearly stated that such a development would be of little consequence towards the abandonment of traditional criteria in favour of new-found aesthetic perspectives. ... Unique characteristics inherent to computer art [and computer creative writing]... acknowledgement of innovation was conditional to abandoning traditional aesthetic criteria,” (Lucas, 1986).

In other words, traditional criteria apply to hybrid and co-created works, but if obviously innovative (or labelled as such), then traditional criteria are not applicable. The traditional (pre-electronic) view in creative writing is that anything unusual goes into the experimental category, to be read only by other experimental writers. However it is possible that co-created electronic literature will not appear in the form of ebooks of new works, but as a new form more suited to the always-on broadband environment.

“In sum, electronic literature *is* digital humanities because of our shared philosophy that a computer is not a tool or prosthesis that helps us to accomplish our work; rather, it is the medium in which we work,” (O’Sullivan, 2021).

For the respondents, nearly all new to generated text, the overall ratings and comments were positive, summed up by the comment:

“It’s going to get complicated, but I guess this is just the beginning of amazing tools that will change the landscape of how text is created and analysed.”

Conclusion

Writers of all types were engaged by the text generation experiments, despite little or no previous experience of the systems. Only a tenth of writers had used generative or combinatorial writing tools before, even though they have been used for twenty years in some areas of their profession.

This willingness to engage with a new process shows that writers are interested in improving their workflow. This was a self-selected cohort via an email list, and so perhaps more curious about AI, in which case the general lack of awareness amongst professional writers is probably wider.

Creative writers were more likely to be fearful or sad about the loss of creative authority.

More practical writers such as copywriters or report writers could see advantages in a system that can generate high quality text for production with little editing required.

The use of text generation in a word processor produced mixed responses, with some thinking it should be outside of the main writing tool, others incorporated alongside dictionaries and style checking.

When asked about ‘fake’ news and text generation, those most dismissive of the technology were generally dismissive of any link. This indicates that those less aware or less interested in

text generation, or other new areas like image, video and audio generation, are underestimating the future widespread applications of the technology. Across the general population knowledge of new technology is low, until a product is advertised.

Popular dystopian fears of a robot overlord, such as exemplified by the Terminator film series, are a distraction from less exciting but important changes in work structures and power relationships, which might be altered in ways no human can understand.

Writers, as cultural workers, should be aware of these significant effects, showing how unaware even the well-educated are to increased automation in their own workplace. This could be due to wilful ignorance or over-confidence, since most writers in the study claimed human creativity as their unique right. Perhaps this is also an over-estimation of the creative aspect of writing, since all writing is based on, or influenced by, other writing and so is susceptible to automation.

The transfer of work away from a professional paid role to the individual includes writing business letters on an electronic device, rather than employing a secretary, and paying for shopping at an automatic till in a supermarket, rather than employing a cashier. These are obvious money-saving changes for the businesses, which also bring increased convenience to the users.

Many currently middle-class professional areas will be affected by automation, such as legal and financial bureaucracies (lawyers and accountants), and even academia (teaching and marking). Even therapists could lose out to chatbots. Factories currently use robots of many types, and this transformation will take place in the knowledge economy, with disruptions to employment caused by automation, outsourcing and downsizing (Acemoglu, 2020).

Mainstream news is ignoring this area, perhaps because the owners of media companies are benefitting from AI-generated job savings. Issues around employment and ethics are replaced by the latest computer victories over the arts (machine-produced old master paintings) or

gaming (human masters beaten), which represent minor class or underdog victories, with machines being the new proletarians or rebels.

The respondents were aware of employment concerns in a general way, and showed unease, despite their lack of practical experience of text generators. The most common negative feedback was that their human creativity would somehow be replaced, or made irrelevant, replaced by a machine. This might be because most of the respondents were middle-aged and older with established skills and careers to protect.

With text generation an editor could generate the text on any topic and then do minor editing for publication. This is an apparent enhancement of the editor role, but removes many writers' jobs from the process, and reduces the teamwork aspect of news or publicity production, so making the editor job less enjoyable. This might seem more applicable to news or report writing, but co-created fiction and plays have already achieved some publicity, although usually as an experiment or novelty (Goodwin & Sharp, 2016).

Discussions of the future of AI art focus on co-creation or teamwork between artists and others such as programmers or system designers, and the software. This might come more naturally to younger age groups who might not see the computer as a separate entity.

“One point [in AI and art]... is the change in attitude and perception observable in different age groups. Today, we need to understand the creative thinking, the cognitive processes, of digital natives,” (Edmonds, 2021).

It is likely that new forms of electronic literature will appear which do not relate directly to older classic texts, in the way that modern digital art does not relate directly to painting.

Further studies with writers and other professionals are needed to explore this new media culture, exploring the ruptures from traditional practices. The discussed ‘casual’ creativity, is consumption of a personalised narrative.

How do people think credit and responsibility should be allocated to various partners in the production of co-created art? How do these intuitions vary based on people's perceptions of the personality, or not, of the system? What might AI co-created 'native digital' literary texts be, given the long history of hypertext and networked art? If casually created for individual consumption, are publishers and critics needed? There are many further avenues for research as computer co-creation in the literary arts becomes more mainstream.

Declaration of Interest Statement

There are no conflicts of interest.

References

Method for IJ HCI: Taylor & Francis Journals Standard Reference Style Guide: American Psychological Association, Seventh Edition (APA-7).

Abadi M., et al (2016). Google Research. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. Available at:

<https://arxiv.org/abs/1603.04467>

Adamopoulou, Eleni & Moussiades, Lefteris. (2020). Chatbots: History, technology, and applications. Available at:

Machine Learning with Applications. 2. 10.1016/j.mlwa.2020.100006.

Acemoglu D., Restrepo P. (May 2020) Unpacking Skill Bias: Automation and New Tasks. AEA Papers and Proceedings Vol. 110. Available at:

<https://www.aeaweb.org/articles?id=10.1257/pandp.20201063>

Ahmad. M., Aftab S., Ali I., Hameed, N. (2017) Hybrid Tools and Techniques for Sentiment Analysis: A Review. International Journal of Multidisciplinary Sciences and Engineering. 8.

31-38. Available at:

https://www.researchgate.net/publication/318351105_Hybrid_Tools_and_Techniques_for_Sentiment_Analysis_A_Review

Ai-Da Robot (2019). "Her artist persona is the artwork." Available at:

<https://www.ai-darobot.com/>

See also: Rea N. (2019). Artnet. "A Gallery Has Sold More Than \$1 Million in Art Made by an Android, But Collectors Are Buying Into a Sexist Fantasy."

<https://news.artnet.com/opinion/artificial-intelligence-robot-artist-ai-da-1566580>

Alammar J. (June 27 2018) The Illustrated Transformer. Available at:

<https://jalammar.github.io/illustrated-transformer/>

Artificial Lawyer (2018) LawGeex Hits 94% Accuracy in NDA Review vs 85% for Human Lawyers (summary). LawGeex research is paywalled. Available at:

<https://www.artificiallawyer.com/2018/02/26/lawgeex-hits-94-accuracy-in-nda-review-vs-85-for-human-lawyers/>

Bailey J. (March 2018) A Brief History of Article Spinning. Available at:

<https://www.plagiarismtoday.com/2018/03/08/a-brief-history-of-article-spinning/>

Bellard F. (2021) Text Synth Technical Notes. Available at:

<https://bellard.org/textsynth/tech.html>

Beeple (2021). “He makes a variety of art crap across a variety of media. Some of it is ok, but a lot of it kind of blows ass. He’s working on making it suck less everyday though so bear with him... :)”

<https://www.beeple-crap.com/>

See also: “Created over 5,000 days by the groundbreaking artist, this monumental collage was the first purely digital artwork (NFT) ever offered at Christie’s”. Available at:

<https://www.christies.com/features/Monumental-collage-by-Beeple-is-first-purely-digital-artwork-NFT-to-come-to-auction-11510-7.aspx>

Bender, E. M., Timnit G., McMillan-Major A., S Shmitchell, S. (2021) On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, March 2021. Available at:

<https://dl.acm.org/doi/10.1145/3442188.3445922>

Branwen, G. (2020) Assessment of creative writing by OpenAI’s GPT-3 model.

“One does not train or program GPT-3 in a normal way, but one engages in dialogue and writes prompts to teach GPT-3 what one wants.”

<https://www.gwern.net/GPT-3>

Branwen, G. (2020) Developmental Stages of GPTs. Available at:

<https://www.greaterwrong.com/posts/3nDR23ksSQJ98WNDm/developmental-stages-of-gpts/comment/iopEYYxJwXM9cZT4q>

Bringsjord, S, Ferrucci, D.A. (1999) Artificial Intelligence and Literary Creativity- Inside the Mind of BRUTUS, a Storytelling Machine. Psychology Press, Mahwah, USA; Book form:

Lawrence Erlbaum Associates Publishers 2000. Paper available at:

<https://www.semanticscholar.org/paper/Artificial-Intelligence-and-Literary-Creativity%3A-of-Bringsjord-Ferrucci/66c1c3cb65423cf8f22022cbda06465e7ca6de13>

Brown T. B., and 30 others (2020) Language Models are Few-Shot Learners Available at:

arXiv:2005.14165v4 [cs.CL]

<https://arxiv.org/abs/2005.14165>

Burroughs W. (1963) The Cut Up Method, from Leroi Jones, ed., The Moderns: An

Anthology of New Writing in America. Corinth Books, New York USA. Available at:

<https://www.writing.upenn.edu/~afilreis/88v/burroughs-cutup.html>

Calderwood, A., Qiu, V., Gero, K. I., Chilton, L. B. (2019) How Novelists Use Generative

Language Models: An Exploratory User Study. In IUI '20 Workshops, March 17, 2020,

Cagliari, Italy. ACM, New York, NY, USA, 5 pages. Available at:

https://www.cs.columbia.edu/~chilton/web/my_publications/Calderwood_How_Novelists_Use_Generative_Language_Models.pdf

Chiang, H., Ge, Y., Wu, C. (2015). Stanford University. Classification of Book Genres By Cover and Title. Available at:

http://cs229.stanford.edu/proj2015/127_poster.pdf

Cherry, K. (2020) What is The Uncanny Valley? Available at:

<https://www.verywellmind.com/what-is-the-uncanny-valley-4846247>

Clark, S (2021) Revisiting and Re-presenting 1980s Micro Computer Art. EVA London 2021 Conference paper. Specifically the Micro Arts Group which released ‘Story Generator’ and many other digital art collections.

Cohen, N. (April 2008) He Wrote 200,000 Books (but Computers Did Some of the Work).

About businessman Philip Parker’s web scraping programs. Available at:

<https://www.nytimes.com/2008/04/14/business/media/14link.html>

Combs, V. (2021) AWS, IBM, Google, and Microsoft are taking AI from 1.0 to 2.0, according to Forrester. A new report says that the hyperscalers are using reinforcement learning and transformer networks to make AI smarter and more mobile. Available at:

<https://www.techrepublic.com/article/aws-ibm-google-and-microsoft-are-taking-ai-from-1-0-to-2-0-according-to-forrester/>

Summarised from original article (members)

Carlsson, K. Purcell, B. Gualtieri M. (2021) AI 2.0: Upgrade Your Enterprise With Five Next-Generation AI Advances. Available at (paywall):

<https://www.forrester.com/report/AI+20+Upgrade+Your+Enterprise+With+Five+NextGeneration+AI+Advances/-/E-RES163520#>

Clark, S. (2021) Revisiting and Re-presenting 1980s Micro Computer Art. Electronic Visualisation and the Arts EVA Conference London. June 2021.

Cramer F. (2000). Combinatory Poetry and Literature in the Internet. Available at:

http://cramer.pleintekst.nl/all/combinatory_poetry_-_permutations/combinatory_poetry_-_permutations.html

Csikszentmihalyi, M. (2014). The Systems Model of Creativity - The Collected Works of Mihaly Csikszentmihalyi. Springer Global 2014. Overview of work from 1964 – 2010.

Cusk, R. (2016) Transit (novel). Jonathon Cape London UK.

Danai (November 2018) What Brings Value to Digital Art? Available at:

<https://medium.com/digital-art-weekly/what-brings-value-to-digital-art-abbdc21117b>

Davis, G. (2019) *Micro Arts History 1984–85 Computer Generated Art and Stories*. Story Software Publishing London.

Davis, G. (2004) *Calm As A Dead Clam, non-linear narrative in zooming 3D environment (ZUI)*. Story Software Publishing London.

Davis, G. (2003) *Affinity Software, zooming canvas creativity software*. Geoff Davis.

Middlesex University Masters in Electronic Arts 2003. Available at:

<https://geoffdavis.org/MA-interface-website/index.htm>

Davis, G. (1985) *MA4 Story Generator. Procedural Story Generator (PSG)*. Generated endless versions of a short story ‘Cow Boils Head’. Prestel teletext viewdata distribution.

Exhibited London Film-Makers Co-operative (1985) Leciscter LCB Deport art centre (2021).

Available at:

<https://microartsgroup.com/ma4-geoff-davis-cow-boils-head-story-generator/>

Dickson, B. (2020). *The GPT-3 economy*. Available at:

<https://bdtechtalks.com/2020/09/21/gpt-3-economy-business-model/>

Dyson, G. (2012). *Turing’s Cathedral*. Penguin London.

Edmonds, E. (2021). *AI, IA and Art*. Keynote talk, EVA London 9th July 2021. Available at:

<http://www.eva-london.org/eva-london-2021/keynote-speakers/>

Electronic Literature Organisation ELO (2021). *To facilitate and promote the writing, publishing, and reading of literature in electronic media*. Available at:

<https://eliterature.org/>

Eleuther (2021) *A grassroots collective of researchers working to open source AI research: GPT-J and others*. Available at:

<https://www.eleuther.ai/projects/gpt-neo/>

Fedus W., Zoph B., Shazeer N. (2021) Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity. Available at:

[arXiv:2101.03961](https://arxiv.org/abs/2101.03961)

Flowrite (2021) Writing tool powered by the latest advancements in AI and NLP. Available at:

<https://www.flowrite.com/about>

Frich, J. Vermeulen, L. M., Remy, C., Biskjaer, M. M., Dalsgaard, P. (2019) Mapping the Landscape of Creativity Support Tools in HCI. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, May 2019, Paper No.: 389. Available at:

<https://dl.acm.org/doi/10.1145/3290605.3300619>

Fredericks, E.M., DeVries B. Genetically Improving Novelty in Procedural Story Generation, Software Engineering, Grand Valley State University. Available at:

<https://doi.org/10.48550/arXiv.2103.06935>

Fullstory (2021) A new kind of analytics to build better digital products. Available at:

<https://www.fullstory.com/platform>

Funkhouser, C. T. (2007). Prehistoric Digital Poetry - An Archaeology of Forms, 1959 1995 (Modern & Contemporary Poetics) . University of Alabama Press. The University Of Alabama Press, USA.

See also:

Geng, Y., Chen, J., Chen, Z., Pan, J.Z., Ye, Z., Yuan, Z., Jia, Y., Chen, H. OntoZSL: Ontology-enhanced Zero-shot Learning. Available at:

<https://arxiv.org/abs/2102.07339>

Gero, K. I., Chilton L. B. (2019) Metaphoria: An Algorithmic Companion for Metaphor Creation. In CHI Conference on Human Factors in Computing Systems Proceedings (CHI

2019), May 4–9, 2019, Glasgow, Scotland UK. ACM, New York, NY, USA, 12 pages.

Available at: <https://doi.org/10.1145/3290605.3300526>

Gervas S. (2012) Handbook of Narratology: Story Generator Algorithms. Available at:

<https://www.lhn.uni-hamburg.de/node/35.html>

Golden R. M. (2001). Perceptron discussed in Artificial Neural Networks:

Neurocomputation, International Encyclopedia of the Social & Behavioral Sciences.

Available at:

<https://www.sciencedirect.com/science/article/pii/B0080430767005635>

Goodfellow I J, Shlens J., Szegedy C. (2015). Explaining and Harnessing Adversarial Examples (v3). International Conference on Learning Representations ICLR, San Diego, USA, 2015. Available at:

<https://arxiv.org/abs/1412.6572>

Goodwin R. and Sharp O. (2016) Sunspring: the world's first film written entirely by an Artificial Intelligence, which when interviewed, named itself Benjamin. Available at:

<https://www.thereforefilms.com/films-by-benjamin-the-ai.html>

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world.

Behavioral and Brain Sciences 33(2-3): 61-83. Available at:

<https://www2.psych.ubc.ca/~henrich/pdfs/WeirdPeople.pdf>

Hertzmann A. (May 2020) Computers Do Not Make Art, People Do.

Communications of the ACM, May 2020, Vol. 63 No. 5, Pages 45-48

10.1145/3347092

Higgs, W. (2019) The Future Starts Here: An Optimistic Guide to What Comes Next. W&N Weidenfeld and Nicolson UK.

Hugging Face (March 2020) How to generate text: using different decoding methods for language generation with Transformers

<https://huggingface.co/blog/how-to-generate>

Hooande (2021) Discussion of GPT-Neo – Building a GPT-3-sized model, open source and free (eleuther.ai). Available at:

<https://news.ycombinator.com/item?id=25819803>

Hugging Face (2020) Write With Transformer – Checkpoints – DistilGPT-2 and Arxiv-NLP are in line editors with demos. Available at:

<https://transformer.huggingface.co/>

Hugging Face (2021) Transformers 10x Faster. Available at:

<https://hi.huggingface.co/accelerated-inference-api/>

IBM Watson Tone Analyser (2020). Available at:

<https://www.ibm.com/watson/services/tone-analyzer/>

Imagen Google Research, Brain Team (2022). Available at:

<https://imagen.research.google/>

InferKit (2021) State of the Art Text Generation. Available at:

<https://inferkit.com/>

Jaswal A S (2019). Byte Pair Encoding — The Dark Horse of Modern NLP. Available at:

<https://towardsdatascience.com/byte-pair-encoding-the-dark-horse-of-modern-nlp-eb36c7df4f10>

Jia, E. (2019). Bag of Words in Genre Identification on the Project Gutenberg dataset (see Worsham and Kalita 2018). Available at:

<https://towardsdatascience.com/bag-of-words-in-genre-identification-on-the-project-gutenberg-dataset-a98ee146d53f>

Johnston D. J., (2008) 330 A.D. : Florian Cramer & the roots of Permutations. Digital Poetry Overview. Concordia University, Quebec, Canada. Available at:

<http://glia.ca/conu/digitalPoetics/prehistoric-blog/2008/07/14/330-ad-florian-cramer-the-roots-of-permutations/>

Johnson K. (May 2020) OpenAI debuts gigantic GPT-3 language model with 175 billion parameters. Available at:

<https://venturebeat.com/2020/05/29/openai-debuts-gigantic-gpt-3-language-model-with-175-billion-parameters/>

Karpathy A. (2015) The Unreasonable Effectiveness of Recurrent Neural Networks.

Available at:

<http://karpathy.github.io/2015/05/21/rnn-effectiveness/>

Kissinger H. (2018). How the Enlightenment Ends. The Atlantic magazine. Available at:

<https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/>

Lau, H., Cohn T., Baldwin T., Brooke J., Hammond A. (2018) Deep-speare: A Joint Neural Model of Poetic Language, Meter and Rhyme. Available at:

<https://doi.org/10.48550/arXiv.1807.03491>

Lucas R. E. (1986). Evolving Aesthetic Criteria for Computer Generated Art: A Delphi Study. Ohio State University 1986.

Lu S., Zhu Y., Zhang W., Wang J, Yu Y (2018) Neural Text Generation: Past, Present and Beyond arXiv:1803.07133

Meadows, A. (2018) IT Pro: The Ethical Implications Of Conversational Ai. Available at:

<https://www.itpro.co.uk/machine-learning/32201/the-ethical-implications-of-conversational-ai>

Mehrabi N., Morstatter F., Saxena N., Lerman K., Galstyan A. (2019) A Survey on Bias and Fairness in Machine Learning. arXiv:1908.09635. Available at:

<https://arxiv.org/abs/1908.09635>

Microsoft blog (September 2020). GPT-3 Microsoft teams up with OpenAI to exclusively license GPT-3 language model. Available at:

<https://blogs.microsoft.com/blog/2020/09/22/microsoft-teams-up-with-openai-to-exclusively-license-gpt-3-language-model/>

Microsoft blog (February 2020). Turing-NLG: A 17-billion-parameter language model. This summary was generated by the Turing-NLG language model itself. Available at:

<https://www.microsoft.com/en-us/research/blog/turing-nlg-a-17-billion-parameter-language-model-by-microsoft/>

Midjourney (2022). Text to Image system. An example long prompt text

‘Realistic render of beautiful glass body in style of nick knight body building competition underwater translucent ethereal ghostlike Bernini sculpture fluid simulation in houdini wrapped in dark smoke like robes and silk veils by ILM and Paolo Roversi and nick knight and piotr jabłoński, beautiful futuristic simplified plastic form distorted by turbulent movement, misty dark studio background, trending on artstation, hyperrealism, matte painting, dutch golden age, subsurface scattering, very fine detail, zbrush central, + face arms hands limbs --h 480 --w 1920 --uplight’. Available at:

<https://www.midjourney.com/>

Mori , M. (1970) Bukimi No Tani (tr. The Uncanny Valley). Energy Journal (essay on robotics in Japanese journal). Available at:

<https://spectrum.ieee.org/automaton/robotics/humanoids/the-uncanny-valley>

Murray, A. (2018) (LSE Professor of Law.) When machines become sentient, we will have to consider them an intelligent life form. Available at:

<https://blogs.lse.ac.uk/medialse/2018/08/14/when-machines-become-sentient-we-will-have-to-consider-them-an-intelligent-life-form/>

Musk, E. (2016). On stage interview with CEO of SpaceX and Tesla. Code Conference, (from Recode.net) Vox Media, California USA. Available at:

<https://www.youtube.com/watch?v=wsixsRI-Sz4>

Nielsen J.(1993) Iterative User-Interface Design, in Computer, vol. 26. IEEE. Available at:

<https://ieeexplore.ieee.org/document/241424>

Obvious Collective (2018). Edmond de Belamy generative artworks. Available at:

<https://obvious-art.com/la-famille-belamy/>

Otsuki G.J. (2020) OK, computer: Let's bring text-generating artificial intelligence into the classroom. Available at:

<https://phys.org/news/2020-01-text-generating-artificial-intelligence-classroom.html>

OpenAI blog (January 2021) DALL·E: Creating Images from Text. Available at:

<https://openai.com/blog/dall-e>

OpenAI (June 2020) DALL·E: Image GPT - OpenAI. Available at:

<https://openai.com/blog/image-gpt>

OpenAI GPT-2 (2019). Better Language Models and Their Implications. Available at:

<https://openai.com/blog/better-language-models/>

OpenAI GPT-3 (2020). Available at:

<https://openai.com/blog/OpenAI-API/>

Pearson, J., Naselaris, T. Holmes, E. and Kosslyn, S. (2015) Mental Imagery: Functional Mechanisms and Clinical Applications. Trends in Cognitive Science, 2015 Oct; 19(10): 590–602. Available at:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4595480/>

Peiser J. (2019) The Rise of the Robot Reporter. Available at:

<https://www.nytimes.com/2019/02/05/business/media/artificial-intelligence-journalism-robots.html>

Perceptron (2022) Wikipedia. Available at:

<https://en.wikipedia.org/wiki/Perceptron>

Pretzelzettel (2020) Story structure in Korean cinema. Available at:

https://www.reddit.com/r/movies/comments/f351g2/discussion_story_structure_in_korean_cinema/

Quill from Narrative Science (2021). Quill transforms data into stories and embeds them directly into dashboards. Founded 2010. Available at:

<https://narrativescience.com/quill/>

Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D. & Sutskever, I. (2018), 'Language Models are Unsupervised Multitask Learners'

Rajan A. (2021) BBC Technology interview with Sundar Pichai, CEO of Google and Alphabet. "The next revolutions." Available at:

<https://www.bbc.co.uk/news/technology-57763382>

Rettburg, S. (2019) Electronic Literature. Cambridge UK: Polity Press.

Reynolds, L. & McDonnell, K. (2021), Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm. CHI EA '21: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. Available at:

<https://dl.acm.org/doi/fullHtml/10.1145/3411763.3451760>

Roach, R. (2017) The computer poetry of J. M. Coetzee's early programming career. Ransom Center Magazine, University of Texas at Austin. Available at:

<https://sites.utexas.edu/ransomcentermagazine/2017/06/28/the-computer-poetry-of-j-m-coetzee-early-programming-career/>

Roberts, S. (February 14, 2017). Strachey's 1953 computer poetry discussed in New Yorker Feb 2017. Available at:

<https://www.newyorker.com/tech/annals-of-technology/christopher-stracheys-nineteen-fifties-love-machine>

Rodgers, J. (2017) The Genealogy of an Image, or, What Does Literature (Not) Have To Do with the History of Computing? : Tracing the Sources and Reception of Gulliver's "Knowledge Engine". Humanities 6(4):85.

Rubin, A. (1980) Making Stories, Making Sense: Language Arts, Vol. 57, No. 3 (March 1980). Available at:

<https://lchcautobio.ucsd.edu/wp-content/uploads/2015/08/Rubin-1980-Making-Stories-Making-Sense.pdf>

Russell, S. (2019). Human Compatible: AI and the Problem of Control. Viking USA.

Salahuddin S. (2021) A Wild Adventure With GPT-3. Available at:

<https://pickledbrains.substack.com/p/a-wild-adventure-with-gpt-3>

Schmid, P (2021) Few-shot learning in practice: GPT-Neo and the Accelerated Inference API. Available at:

<https://huggingface.co/blog/few-shot-learning-gpt-neo-and-inference-api>

Semantic Arts Admin SAM (2019) Semantic Ontology: The Basics. Available at:

<https://www.semanticarts.com/semantic-ontology-the-basics/>

Sharples , M. (1985) Cognition, Computers and Creative Writing. Ellis Horwood Publishing, England.

Shneiderman B. (1999). User Interfaces for Creativity Support Tools. Human-Computer Interaction Laboratory, University of Maryland USA. Available at:

<https://www.cs.umd.edu/hcil/pubs/presentations/user-interfaces-creativity/index.htm>

Shneiderman B. (2008). Creativity Support Tools: A Grand Challenge for HCI Researchers.

Chapter in book: Engineering the User Interface: DOI:10.1007/978-1-84800-136-7_1.

Available at:

https://www.researchgate.net/publication/226067887_Creativity_Support_Tools_A_Grand_Challenge_for_HCI_Researchers

From the book: Ortega, Manuel & Bravo, Crescencio & Redondo, Miguel. (2009).

Engineering the User Interface. 10.1007/978-1-84800-136-7.

Shane, J. (November 2020). You Look Like a Thing and I Love You. Wildfire/Headline UK.

Silver D., Hubert T., Schrittwieser J., Antonoglou I., Lai M., Guez A., Lanctot M., Sifre L.,

Kumaran D., Graepel T., Lillicrap T., Simonyan K., Hassabis D. (2017). Mastering Chess

and Shogi by Self-Play with a General Reinforcement Learning Algorithm. (AlphaZero

system.) arXiv:1712.01815v1 [cs.AI]. Available at:

<https://arxiv.org/abs/1712.01815>

See also:

DeepMind research blog: AlphaZero: Shedding new light on chess, shogi, and Go. Available

at:

<https://deepmind.com/blog/article/alphazero-shedding-new-light-grand-games-chess-shogi-and-go>

Story Live CST (Bellard, F. & Davis, G., 2020-2021). Available at:

<https://storylive.com/>

Story Software: Notes Story Board and Story Lite CSTs (Davis, G. 2010-2021). Available at:

<http://www.notesstoryboard.com/>

Taylor, G. D. (2014). When the Machine Made Art: The Troubled History of Computer Art,

Continuum Publishing Corporation.

Van Noorden, R. (February 24, 2014) Publishers withdraw more than 120 gibberish papers.

Available at:

<https://www.nature.com/news/publishers-withdraw-more-than-120-gibberish-papers-1.14763>

Vincent, J. (March 2019) A never-ending stream of AI art goes up for auction. Available at:
<https://www.theverge.com/2019/3/5/18251267/ai-art-gans-mario-klingemann-auction-sothebys-technology>

Wang J, Zhang X, Zhou Y, Suh C, Rudin C (2021). There Once Was a Really Bad Poet, It Was Automated but You Didn't Know It. arXiv:2103.03775v1 [cs.CL]. Available at:
<https://arxiv.org/abs/2103.03775>

Wikipedia, Generative Adversarial Network (2021). Available at:
https://en.wikipedia.org/wiki/Generative_adversarial_network

Wolchover N. (2012). Why It Took So Long to Invent the Wheel.
<https://www.livescience.com/18808-invention-wheel.html>

Wolff M. (2019) How To Make Custom AI-Generated Text With GPT-2. Available at:
<https://minimaxir.com/2019/09/howto-gpt2/>

Worsham, J. Kalita, J. (August 2018). Genre Identification and the Compositional Effect of Genre in Literature. Proceedings of the 27th International Conference on Computational Linguistics. Available at:
<https://www.aclweb.org/anthology/C18-1167>

Yanai, I., Lercher, M. A hypothesis is a liability. *Genome Biol* 21, 231 (2020). Available at:
<https://doi.org/10.1186/s13059-020-02133-w>

Appendices

Further reading on creative writing and text generation

AI Dungeon (from Latitude): Play and create adventures with infinite possibilities, brought to life with one of the most advanced AIs in the world. Originally developed at a hackathon in 2019, it is now using OpenAI GPT-3 (2021).

<https://play.aidungeon.io/main/landing>

See also:

https://en.wikipedia.org/wiki/AI_Dungeon

Anderson J., Raine L. (2017) *The Future of Jobs and Jobs Training*. Will well-prepared workers be able to keep up in the race with AI tools? Will market capitalism survive?

Available at:

<https://www.pewresearch.org/internet/2017/05/03/the-future-of-jobs-and-jobs-training/>

Armstrong, J. (2014) *Experimental Fiction*. Bloomsbury Academic, London.

Bender E., Gebru T., McMillan M., Shmitchell S. (2021). *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* Available at:

http://faculty.washington.edu/ebender/papers/Stochastic_Parrots.pdf

Branwen, G. (2019) *GPT-2 Neural Network Poetry*. Demonstration tutorial of retraining OpenAI's GPT-2 (a text-generating Transformer neural network) on large poetry corpuses to generate high-quality English verse.

Independent writer on psychology, statistics, and technology.

<https://www.gwern.net/GPT-2>

Broadbent E. (January 2017) *Interactions With Robots: The Truths We Reveal About Ourselves*. *Annual Review of Psychology* Vol 68 p627-652. Available at:

<https://www.annualreviews.org/doi/10.1146/annurev-psych-010416-043958>

Burgess E. (2015) University of Manchester Press.

Understanding Interactive Fictions as a Continuum: Reciprocity in Experimental Writing, Hypertext Fiction, and Video Games.

Links non-linear narrative, hypertext fictions and video game narrative.

Burton N. (2014) *The Psychology of Laziness*.

What is the motivation of creative writers using generated texts?

“Our nomadic ancestors had to conserve energy to compete for scarce resources and to fight or flee enemies and predators. Expending effort on anything other than short-term advantage could jeopardize their very survival... Desire led to action, and action led to immediate gratification, without much need for proposing, planning, preparing, and so forth.”

<https://www.psychologytoday.com/gb/blog/hide-and-seek/201410/the-psychology-laziness>

CELL (2021) Consortium on Electronic Literature. An international organization that includes 11 member organizations, research labs, and research centres. Available at:

<https://cellproject.net/>

Craig C., Kerr I. *The Death of the AI Author* (March 25, 2019). Osgoode Legal Studies Research Paper,

Discussion of uses of AI text. Available at SSRN:

<https://ssrn.com/abstract=3374951> or <http://dx.doi.org/10.2139/ssrn.3374951>

Electronic Literature Organisation (ELO) (2021) An international organization dedicated to the investigation of literature produced for the digital medium. Available at:

<https://eliterature.org/>

Fredericks E. M., DeVries, B. (2021) *Genetically Improving Novelty in Procedural Story Generation*.

Using non-statistical techniques, for limited environments:

“Procedural story generation (PCG) tailors a unique narrative experience for a player and can be accomplished via multiple techniques, from matching storylets to grammar-based

generation. There exists a rich opportunity for evolutionary algorithms to be applied to this domain for intelligently constructing game narratives.”

arXiv:2103.06935v1 [cs.SE] 11 Mar 2021

Generative Unfoldings exhibition (ref. Nick Montfort MIT). 14 software artworks. Available at:

<https://generative-unfoldings.mit.edu/>

Konig M. , Branwen G. (2020). The Universe is a Glitch (AI-driven music video). Poem by Branwen using OpenAI GPT-3, David Bowie-style music and video AI generated by Konig.

<https://www.youtube.com/watch?v=5wl29laSOU4>

Miller A. I. (2019) The Artist in the Machine : The World of AI-Powered Creativity. MIT Press, USA.

Platen von P. (March 2020) How to generate text: using different decoding methods for language generation with Transformers. Available at:

<https://huggingface.co/blog/how-to-generate>

Montfort N. (2009). MIT . Taroko Gorge poetry generator in Python. Available at:

https://nickm.com/taroko_gorge/

Ohno J. (November 2019). A short history of procedurally generated text.

“...finally beginning to step into the limelight in literature—to fill the demand for new and strange experimental fiction, generate surreal comedy, and to aid pulp authors in fields like Kindle erotica to feed readers’ insatiable hunger for content. ... procedural writing may soon become as common as the use of spell-checkers are now. When that day comes, we should expect literature to get much stranger!”

Available at:

<https://tedium.co/2019/11/14/procedural-text-history/>

Stevick, P. (1971) Anti-Story: Experimental Fiction. The Free Press, Macmillan, USA.

Uotila , A. (2018) Text Generation with Stateful Context-Free Grammars. Available at:
<https://www.semanticscholar.org/paper/Procedural-text-generation-with-stateful-grammars-Uotila/d28a6b981c3c54ef300c1596b70ba6409c25c3bb>

Wilson K. (February 2020) The World's Second Largest Wikipedia Is Written Almost Entirely by One Bot. Vice Magazine. Available at:
<https://www.vice.com/en/article/4agamm/the-worlds-second-largest-wikipedia-is-written-almost-entirely-by-one-bot>

Credits

Text Synth

Permission was granted to use Text Synth in the study by Fabrice Bellard July 7 2020.
Bellard has many open source system available for free unrestricted use. Please visit his website, where Text Synth system is described and linked as ‘Online Text Completion’.
See <https://bellard.org>

Image Credit: Man and dog image prompt

Writing prompt: ambiguous image of man with dog in backpack. New York Times, Jacqui Cheng (May 2019). NYT ‘Picture prompts’ are a public resource for literacy classes.

This is a viral image and situation, both man and dog are celebrated. Maxine is the dog, Bryan Riesberg the man, who is now a filmmaker.

Maxine’s Instagram [madmax_fluffyroad](https://www.instagram.com/madmax_fluffyroad) (600k followers)

https://www.instagram.com/madmax_fluffyroad/

<http://bryanreisberg.com/>

Image sourced from: <https://www.nytimes.com/2019/05/31/learning/dog-in-a-backpack.html>

Picture prompt literacy resources: <https://www.nytimes.com/column/learning-picture-prompt>

Charting software

Creatley

<https://creately.com/diagram-type/flowchart>

Feedback comments

Caption Experiment (first) feedback

Raw data per respondent, no gaps, spelling and grammar are uncorrected. This data as shown here is not ordered across sections.

Question:

What did you think about using the text generator?

Did it help or hinder or no effect?

How did it help you?

Answers:

I found it thought provoking and fun.

It's possible I've misunderstood the intended process for this experiment - I thought there would be a separate screen to enter a news/fiction piece after clicking to continue - but text completion tool is amusing in itself. I tried it before starting (without an image prompt) and got some uncanny valley text out of just the word "Hello".

Helped. You get a body of text that you did not have before. The text provides ideas to expand on the image concept.

I was surprised that it could work well as it did on a couple of the examples. I suppose that a story generator could work if it had enough data in it. It has helped me to think about a game app for language learners I am making and how it could be expanded.

Good

It generated very coherent text that related well to the words I input.

It made me smile

I found it interesting to see what the generator would come up with and found it useful in thinking of ways to caption the image outside of my usual way of thinking. Was quite trial and error though as the generator often produced text that was not relevant though that was probably due to the keywords I had inputted.

Hard to say without using regularly. It could help with minor blockages and or switches in narrative. I would worry it might make me a tad more creatively lazy if used too much

The result would have been better if I just used my own mind but I guess it's ok.

Not that interesting. Many tries.

The text generated was very random and stream of consciousness which helped give interesting ideas

it is just Zum, but it can grow to am very useful Tool

It added power source which I hadn't thought of.

It's quite fun to see suggestions for your ideas. Might be helpful if you're stuck to come up with a catchy caption.

Can see how would help with ideas, but raw text somewhat stilted and of limited relevance no effect took many tries i won't do this for FUN It created something new.

This was actually very clever! But what I found really disturbing was that the subject matter of the two texts I did was closely - too close for comfort, in fact - related to the movie I'd watched just before doing this test. Mind reading? You tell me :-)

Do not know why the text was generated so can't ask "why do you think that would make sense?" Not sure what relationship this is meant to simulate. It was a bit like working with an uninspired person. I could be rude about them. Some of it - in fact a large part of it - was nonsensical - almost but not quite funny.

Surreal selections were vaguely entertaining. Not sure why I would use it

One extra sentence was good, but the rest was wrongheaded, so to speak.

Very good. It gave another point of view and it wrote in the same style.

I wrote things that I wouldn't have otherwise

The text generator did not help me with ideas that might have developed the "story" of the dog and his owner. In part, this may be because of my use of the word "walkies" but in general the text generated appeared to be random comments about dogs and walks.

The text generator hindered my construction of a caption. It threw up things that were not relevant - sentences about vegan dog food!

not particularly artistic response

It was a big help. I started with something vague a sort of descriptive sentence mainly about their names. I was looking for something whimsical and I got it in the return sentences. I was quickly able to reshape the generated text into the emerging direction that built on my original intention but was extended by my new ideas generated on the spot.

It had a tendency to repetition. It did seem to take into account style AND register. Best I've seen, but not there yet.

It did help in generating new ideas - a bit like mind-mapping.

Lots of ideas but few I wished to explore. If I was "stuck", it might be advantageous.

it reverbrates like the echoes...

It help me move the imagery forward to get more story lines.

I'm not sure I was using it properly. Multiple (and I mean a lot) inputs might produce something useful at some point.

Some phrases helped but some nonsensical It was interesting to see where it went liked it

A good idea but still in its infancy. Text generated not usable or logical. Examples I generated didn't really act as a springboard for new prose development.

good suggestion

It ain't Proust, Joyce, Burroughs or Mallarmé. In addition, the unconscious, which uses metonymy & metaphor, cannot be reduced to machine type operations. Voilà.

So far no effect.

It didn't exactly bring anything useful, but my words were not common. I write about fundraising for an animal rescue charity and used 'love animals' 'fundraising' and 'volunteer'. Not common words so maybe that's why the results were scant and vaguely relevant. Great effort at a tool and I wish you all the best

I suppose it was picking up on certain aspects of my text and running with it in a different direction. I wouldn't necessarily have followed those lines mainly because they didn't seem to make much sense. The first one moved away from 'dog' which was central to the picture and my story, onto 'wolf'. Perhaps because I'd mentioned 'fox' and it was using other animals. The second time I did it with the same picture, I mentioned 'flickering fingers typing' and it jumped to something invisible on 'my shoulder'. I suppose this could have been worked into a different story, far from the original idea. Perhaps I'm thinking too literally. But once I'd got going with the dog on the shoulder, I didn't want to abandon it. The second one generated had possibilities, I suppose.

No effect. The generator came up with worse than I did.

Limited use Think I must have done something wrong...

Served not my need

Can't see the value of generating pseudo realistic text

Didn't really help- went off on a completely different tangent.

it was senseless

News Experiment (second) feedback

Raw data per respondent, no gaps, uncorrected

Question:

What did you think about using the text generator?

Did it help or hinder or no effect?

How did it help you?

Answers:

I still think it is thought provoking. It gives you the scope to think about something you would not normally think about.

Please disregard my last comment about the experiment process: it turns out I did understand it, but wasn't expecting to get a question after each bit of writing. The generator really isn't well suited to news articles, as far as I can tell. It's possible I could have got a more suitable prompt out of it by repeatedly generating new text or changing my caption slightly, but at that point it's definitely a hindrance rather than a help. The generated text tends towards the bizarre, but not in an amusing Onion-style sort of way, and pretty much has to be entirely rewritten to make any sense as a news article.

Helped. It researched and found newsworthy article based on the original idea. And it seemed true.

It didnt generate a lot more text but it seemed more relevant. This is a very interesting experiment. How is the text generator on creating metaphor?

The best one so far

This generation was better than the first. I think it was very close to what a real person would write.

Made me feel weird that the story generator was so jaded

It was quite difficult to generate text that was relevant to the image more so than for the caption one. There is a lot of variety in subject matter when using the generator which is very helpful. I think if there is a specific news story you wanted to write it might not be the most useful but if the news story had a vague prompt it would be quite helpful.

It took a line of enquiry and kind of confused it. I felt the impact of AI rather than individual creative thinking

The same as the last go. The result would have been better if I just used my own mind but I guess it's ok.

Quite weird phrases I would not have thought of.

It prompted ideas but several times they were unusable, making the protagonist a woman despite me saying male, and putting cat rather than dog for example. The one I used in the end finally got both these right

I think it will be much better, if there are more informations

I don't write news articles so not sure I used that effectively or would be able to judge what I gained / didn't gain from it.

It had some suggestions I used. However

Slightly more, probably because very slightly more used to it!

same as last time. no effect took many tries It provided the back story perfectly believably.

This one wasn't as good, it had too many nonsensical repetitions.

It did make me think of potentially interesting aspects that I might not otherwise have thought of. It was irritating that it did not correct errors. A human collaborator would do that. Slowed down the task significantly.

Didn't seem to be especially helpful in a news context compared to the previous example

Not useful.

Very good. The news item went perfectly with the first line of text.

I thought it was more beneficial in this circumstance as it provided me with information to add content to my initial idea.

It helped but only in the sense of providing relevant information about cats that I could incorporate into a last paragraph for the piece.

Again it hindered me in thinking about what I had written. It did not provide any extra thoughts or pieces of information that I could use to write more.

it doesn't read my mind - haha!

It continues to be quite helpful. It pleasantly surprised me by extending my work into new areas rather than scrambling what I already came up with. This is good because in this longer example I invested more time and effort in building the paragraphs than in the first example. The first was a short probe of a sentence. And having the program juggle things around was quite useful. In further development of the app it would be useful to allow user input into the degree of revision of the existing user input. It could range from scramble everything to only extend the existing input.

Better this time.

It can open doors.

For writing a recipe, it helped to have the kind of writing that usually comes with writing about cooking.

Can be repetitive (also addictive, I can imagine! Ha!)

Nothing new offered I was excited to see what it would do with more text

LMAO! Text generator can't tell Trixie is a dog, or that Crash and Stupor are band names (for this article). The text generated doesn't fit the article being written. Horrible, horrible, hardly human (or canine).

Fiction Experiment (third) feedback

Raw data per respondent, no gaps, uncorrected

Question:

What did you think about using the text generator?

Did it help or hinder or no effect?

How did it help you?

Answers:

It is most helpful it gets you thinking were you would not normally do so.

Weirdly, my story caption generated something that would have been much better suited to a news article (though it still veered wildly off topic). I liked it enough that I didn't edit it, other than to omit the typical incomplete sentence at the end.

Helped. It started off in a voice that I had never seen or used before. The language is strange and different. I would not use this voice. I would use the setting and rewrite all of the short story. The idea was OK and flowed well after the 3rd generation. It could be built upon.

It knew better how to connect to my text and then it just stopped. IS it that it had nothing to add? I am not sure about this. I think it would definitely help generaqte ideas and it would probably help language learners even though it does make syntax mistakes as in my first result.

Good

It seemed like a logical response, although it left a lot of writing to be done.

Help me? I don't feel that i need help, i am curious about the text generator and playing along with the experiment... this last one was pretty funny

I found when completing this prompt that I was getting text that was better suited to the news story which was interesting. I do think it is helpful to have this kind of resource when writing a fiction story because it could help you think outside of the box and come up with new ideas/places for the story to go.

It made more sense narratively but felt a little more cliched than the text I may have chosen to write

A good story that can be useful for ideas

Made a sort of poem or flash fiction story. Might use as basis for something.

It threw up lots of random thoughts which were useful but also was at times very repetitious and a lot of it was unusable

it only Works in English language?

Clever. I'm not a fiction writer either, but I like how it generates words.

It helped a little. Probably need to input my key points for it to generate more ideas.

Starting to see how its brainstorming by other means

a little now

It sort of ignored my text - turned the man into a boy and introduced a father etc. It didn't introduce an effective new layer to the story.

This time it didn't work at all/ I think the trick is to choose keywords very carefully and as many as possible.

It did make me think of potentially interesting aspects that I might not otherwise have thought of. Left it going for a good time to see what it would come up with. It stopped pretty quickly.

But did make me think of a new angle. It does expand potential creative space. Rather like using random numbers in a creative task. An annoying thing is that the picture ought to act as a focus for the software agent as it does for the user. Instead the app acts only on the most recent text entered. It does not remember what you typed previously so keeps starting over.

Regularly comes up with nonsense though. Sometimes reads like poor fanfic - not surprising

I suppose considering the data going into GPT-2. SW assumes the wrong genre and it continues to feel as if it is trying to hijack the story.

Random

Didn't help.

Not very good. The story did not have any development. It went round and round the original text.

The text generator didn't seem to work on this page.

It was not helpful at all. It seems to have picked out the subject but not provided anything that I could use.

The text generated by the computer failed to see that the dog was in control of the story which was from his/her perspective. It was not the owner's voice that was speaking. But the following text assumed it was the human. It also did not seem relevant anyway referring to "Pals" which is not the name of the dog food that the dog was interested in. So the text generated was of no relevance in continuing the story.

needs to be more input from writer as to their train of thought ...

Does. Not. Work. The generator doesn't understand, nor can it write, in context. It also knows nothing about pop culture apparently. It hinders my writing like a really bad writing partner.

It seems to latch on to one noun (in this case "haircut"), and run with it to Non-Sequitur Land. I'd have to jettison everything it [creates?]

Genuinely thrilled to see it trying to express some complex ideas

Final Questions feedback

Raw data per respondent, no gaps, uncorrected.

Question 1a:

Overall, did you enjoy using the text generator?

Answers:

It is thought provoking.

I wrote a story that for me sort of paralleled the use of the generator. It seems to have responded in kind. I find that quite amusing.

The news one was the most enjoyable

Seeing the generated text and how it related to my input.

This is weirdly phrased. The question should be: "I enjoyed using the text generator panel" agree...disagree

I liked how easy it was to use and how the left was the generator and the right was the text box. This made it easy to cycle through generated text until you found one you liked that was relevant.

Seeing what would generate in real time

The writing that was generated.

Unusual phrases, juxtaposition. Would be useful for ideas

It prompted lots of fun ideas and allowed you to think s but differently
much Potential!!

I liked the generation of related words, sparked ideas.

It was fun to see what ideas it generated as it could bring up suggestions that was different from what I had in mind initially so it may help me approach my ideas from a different aspect or viewpoint.

Seeing the tangents generated and possibilities therein
doing the Q&A cuz its cool

It added a new angle which may have not occurred without the generator.

It was fun but honestly, it's good to know machines still can't grasp the intricacies of human logic :-)

surreal perchance juxtaposition. puzzling at the possible sources of them

It was clever. The first example actually gave something back, so some AI aspect worked there.

The best thing was seeing the words appear like a ghost writing on the page. Anticipation of the narrative being developed or taking a sharp turn.

It was interesting how the generator seemed to help me the most in the news article, but I could see how it could easily work the opposite way to hinder one's imagination.

Most - the novelty of seeing how this works now. A long way to go, methinks!

seeing the results ...

The text generator just doesn't work as a writing partner. The AI would have to be much more advanced to help write human-interest articles or fiction. As it currently stands, it can form sentences but doesn't understand context or metaphor. It doesn't "think" in images, so creative writing is lost on it.

How coherent its text was

Question 1b:

What was most interesting? Please explain.

Answers:

Maybe seeing the same picture.

Because the generator rambles, the text it produces has little substance. I'm not really interested in trying to edit what's essentially nonsense into a coherent story or article.

Least interesting was the first round because the generated text was less relevant. It seemed that the more text I created the better this generator worked. Maybe it needs inspiration also.

Having to write about the same image three times.

Anyway, I least enjoyed the horrible user interface from 1982 (yes I know that is far before netscape but still this shit is beyond old school)

The predictability of the AI. Needs to be better able to rerun different options rapidly to have more probability of sparking an idea or a concept outside the author's normal range

The picture

Hard to use on a phone. Had to repeat generated process.

can i influence the "direction" of the Tex?

The fiction didn't write like fiction, it lacked a human quality. It read like it wasn't trying to capture interest. I guess it does depend what you put in though and I didn't write very much at all.

It generated ideas and sentences that didn't make sense to my key words/ideas.

Syntactical inconsistencies slightly irritating

waiting for the text to generate

Least: That the system was not able to remember previous texts and so had to keep all text in the input box to retain a sense of continuity. Also that the system did not parse the image so did not take the content in the image into account.

Where the writing is more esoteric, the AI hasn't anything to give.

The worst thing was when the story was disappointing or did not make any logical sense, and was without direction.

Answering questions on the same picture.

Least - trying to write about a photo that made little impact on me. Most - the novelty of seeing how this works now. A long way to go, methinks!

I did not enjoy using the text generator because it was not able to offer any extra comments or information that was relevant. This was especially the case with the short story where the computer failed to identify that the story was from the dog's perspective. Any human would have recognised this but the computer failed to do so. This has something to do with human imagination.

the layout wasn't particularly inspiring ...

the shorter text gives it less to go on, so it's more random

Question 2:

I felt emotional when the generated text appeared [Likert].

Which emotions did you feel, please describe?

Answers:

Thinking of my observations of what the picture contained.

Having started with a caption of a photo of a dog in a backpack, ending up with generated text revolving around two (married?) friends being involved in a shooting/robbery was quite jarring. However, I wasn't particularly invested in their story - mostly I just found the tone amusing.

I actually felt disappointed with the first round as the generated text seemed to miss the point or maybe make an irrelevant reply. I felt curious and was unhappy when the text was not relevant. I did feel very interested and pleased when the generated text was related and built upon my ideas. I may try this again and use only words and not sentences. I wonder what it would do then.

Not that much

Detached interest, not particularly emotional.

humor, awkwardness

I did not feel any emotions

Didn't feel any sense of ownership with the generated text

I didn't feel much emotions since the text doesn't make me feel emotional.

Curiosity

It was mostly amusement due to the many incongruous suggestions

the is something growing

Neutral. I think this was my fault largely though as I didn't type very much in.

Didn't really feel emotional.

Piqued interest, if that's an emotion

boredom

Satisfaction when it created something pleasing.

As I said before, I was spooked when I saw the first two texts because they closely reflected the subject matter of the film I'd just watched. It was as if the generator could read my mind.

You never know! :-)

A degree of surprised interest when it was relevant. Amusement when it came up with stuff that I would never have come up with, but in a serious job it would have been extremely annoying. Anthropomorphising (which is all too easy to do) one would want to keep saying "focus on the job! We haven't got all day!"

When it is going finish? Emoticon Grinning Face With Smiling Eyes

Excitement of words appearing like magic

It did not impact me emotionally. I just found it slightly frustrating when I didn't have freedom to write what I wanted to write.

Bemusement

Frustration at the lack of imagination of the machine. It was definitely off kilter with all the examples.

disappointed

Amusement at how hopeless the generated text was.

excitement, awe

Question 3:

Do you feel that you have used somebody else's work? [Likert].

Does this seem relevant and why?

Answers:

It did not enter my mind that the words that I had produced were being manipulated by a program.

I'd probably credit the generator for a portion of the work since readers may be interested to know where it came from and appreciate the opportunity to experiment for themselves, but I'm not concerned that I'm using somebody else's work because the generator itself cannot claim royalties etc. and will not have any concerns over how the generated text is used.

If I would use someone else's work it would have to be a collaboration or I would not accept it. I am very proud of never copying another's work but I really like to collaborate.

Because it sounds like a book or something when I generate it

Whenever you use writing that you did not create, obviously you're using someone (or something) else's work.

no, i feel like i played with somebody else's toy

Only when the generated text very clearly had come from an article or a tweet or social media post of some sort. But since those sources are in the public domain and as long as they were credited appropriately I would not feel uncomfortable about that. It is an interesting question and was not a worry that came to mind when I was using the generator.

It's not somebody else's work it is aggregated AI so cant be plagiarism. The secret would not be in using the generated text to add to my own but in providing lots of possible insights that my own writing could then apply and own

I don't know.

Seemed generated random no-one would write this.

Plagiarism is always a worry but it was so randomly generated and out of context that the text didn't seem like it had been part of a real or existing piece of work

doesn't any matter

Not really, I guess because from a poetry perspective anyway it would be what I made of the ideas it gave me.

Didn't feel like you

I've used someone else's work as it just generated ideas and sentences based on my key words.

Normally I would edit the generated ideas so it becomes your own work anyway.

No, text came from my input

not really

Doesn't feel like somebody else's. Feel that it was generated by what I wrote. A bit like using a cut up technique maybe.

Well, it was just pretty generic in style, anyone can write like that.

The original work by myself sets a possible direction but then the results are from elsewhere.

Authorial choice in how or where to start and when to stop but the bit in the middle is a proxy

I didn't use enough of it to worry.

No, it wasn't someone else. It felt like I was writing in translation.

It seemed as though the ideas were not my own and my work was taken out of my own hands.

I do not think it applied to my answers but it would be an issue if I had used any of the text generated.

nope - because the generated text wasn't relevant to my train of thought

I do have questions about copyright, and ethically who the text belongs to, but aside from the expertise in creating this tool (which is enormous) the text feels as though it's borrowing from a deep well of existing ideas and literature

Question 4:

Imagine a generator in your ideas editor or word processor.

What extra features would you like?

Answers:

To be able to input a number, for various scenarios to be produced.

I've used generators such as this to produce text for my work in the past, but only when I essentially just need some gibberish to throw in there. I can't say I wouldn't use a version built into my usual word processor, but it's absolutely not a feature I would go looking for.

It's very straightforward to copy and paste from an online generator already, and in many ways preferable to doing anything directly within the document I'm actually working on.

More text. I would really like it if the generator could carry on a conversation with me as in the story I wrote. That would really be interesting and maybe very helpful. Could you

possibly make a generator that could discuss ideas? That would be a very valuable asset. It

would also improve it if it could ask questions of me the author as it decided what to

generate. Can it more easily generate relevant questions in order to make its response more

relevant? Are you using an AI or just a ruled database? An AI to generate a conversation

would really be quite thrilling.

not going wrong

I would not use it.

some sort of step by step options for the direction that the text could take, and an intelligent learning system for giving +1 to the stuff I like and -1 to the stuff i don't... to avoid repeating themes and phrases that i don't like, and to guide the text generation more actively

I would especially if trying to come up with a caption or story of some sort. It would be really useful when you have writers block or you are just looking for another perspective as the generated text can be so random that it sparks an idea that you had not previously thought of.

I think if there was some way to determine what kind of output you were looking for like fiction/non fiction or an article or a poem. The outputs were quite random and hard to find something relevant to the prompt without going through a few cycles.

I'd use it sparingly when struggling to progress. It would be good if it could generate at macro level (eg narrative plan) rather than just micro (sentences or paragraphs) Build an archive from my previous work to more accurately accommodate my personal style

Improve the results because of th time they don't make sense

Have to be smooth to use

Not sure. It would need to be more sophisticated in terms of the text being generated but maybe that was because I wasn't used to using it.

Access to databases with relevant information would be helpful

Maybe, I would use it. It's interesting. Synonyms would be good too.

An option to pick the genre or style.

Might use it if I had it, but wouldn't seek it out

maybe

Something that provides alternative phrases/words perhaps but stays true to the meaning of the text.

The biggest problem I have with this is that using it might gradually weaken my own creative muscle. I'd rather exercise my own brain.

But it might become habit forming if it were better and less frustratingly out of touch with the author's intentions - the aim of the writing.

Some ability to be aware of sources and for some parameter to control the output better

I wouldn't use it.

I would like to be able to input in names of characters, settings and the type of plot, or genre.

No it would inhibit my creativity

This would not really help me in writing something.

I don't want any extra features. The only possibility I can imagine would be a situation involving writer's block where one way of dealing with it would be to ask the machine for alternative next sentences. But given it draws these from other writers there is an immediate problem with plagiarism. And you would not know where the extra new text had come from. A bit of a nightmare best avoided.

we are entering the music world here - apple loops - people using them - constructing a piece of music and thinking they wrote it . . . this could be OK for people who have no creative skills (mostly business people) but sadly I think you are doing the creative world a disservice I don't think I'd use it to write. I want to use my own creativity, imagination, and learning to create my works. On the other hand, I'd welcome a research AI assistant. Someone/something I could ask to search the world's libraries (or out of print books, etc.) for titles / background / information on, for example, details of late medieval village life. That would be useful. But use a feature like this as a writing partner? No.

Question 5:

Do you think you could sell this as your own work? [Likert]

Do you have any comments on notions of authorship, plagiarism, ownership and intellectual property?

Answers:

This is just purely thought provoking to give you ideas of your own.

If I edited the generated content, I think I could sell this as my own work. If I didn't edit the generated content, I think I could probably still sell this as my own work, but I'm not sure what I would do with it. I have no more concerns about plagiarism/ownership due to the use of this tool than I would if I were constructing a story by throwing darts at a dictionary. The text it produces is more readable, but essentially no more or less mine.

Of course I do. I am part of the world-wide conversation among authors and I acknowledge my sources whenever I can. Intellectual property is not the text but rather the ideas within the text and the method of generating a response in the reader. Since I make money from my writing I fully appreciate the value of intellectual property and I respect it. After all many people can be taught to make a response or to build a box but can they be taught to respond in a way that generates more ideas and can they build a unique box? If they can that unique box and that response are their intellectual property.

I have no idea what that means.

Since this involves federal law I have no comment.

I think if other pieces of work are being used or referenced as long as they are credited appropriately that would be fine to me. I do think I could sell this as my own work as I was editing the generated text and ensuring that what was generated was appropriate. It was not as simple as just pressing 'generate text' and submitting whatever was produced.

See answer to an 4. I could FEEL this is my own work in the sense that a DJ or record producer/remixer feels the re-imagined original song as their own work

I have no idea what that means because I am just a boy.

Is generated so who cares. Also edited by me

Maybe I could but I don't think anyone would buy! I feel I took random elements and added to them

it depends on Quality

Again, from a poetry point of view I don't have an issue with it generating ideas. From this perspective, I use ideas from myth, folklore, natural history and I have no issue with that!

Would only use as starting point for rewriting, don't feel plagiarism the issue

Don't see why not. The origin of the idea is my own.

I probably could but I wouldn't because my standards are higher :-)

In principle, I think that an author has rights over any original work they produce unless they have explicitly signed it away. Maybe it depends where the text is assumed to be. I couldn't sell it because I don't think it is particularly good.

Difficult to sell any writing to anyone. That aside, the the production might be seen as a unique construction or framing of other's material. Like architectural photography? But the sources are black boxed.

Only because I used what was in the public domain anyway. If it was highly effective, I would be conflicted.

No comments. I think it's OK to use as it does not sound v human.

No it would feel wrong.

I am quite careful about ownership issues and am particularly careful about my own work. So this would not really help me in writing something.

You cannot have plagiarism in a short story bound up in imagination where the story is a flight of fancy

I think if it was more like a thesaurus I might be interested in using it - and would have no qualms about IP as the original thought came from me

Everything I wrote came from my own mind/imagination. Everything the text generator wrote was unusable. If the generator wrote good text, I think I would feel that I'd used someone else's work. That text wouldn't have come from me, but from AI. That, to me, is cheating -- mainly myself, as I know from experience that if I don't use my creativity and imagination, I'll lose it. But until AI can think for itself (and also use the five human senses) it's not going to be able to write human interest or fiction well. While AI can think, it can't feel. It isn't self-aware, it can't experience (see, hear, feel, taste, touch). As some have pointed out, when it is self-aware and can experience, it won't need its creators/humans any longer. I daresay we won't need one another either. So...if AI can write fiction, I might -- *might* -- enjoy reading what it came up with. But I enter an altered state of consciousness, and imaginary worlds, to create what I write. I've just discovered I've no desire to stay in this world and watch a computer write.

It's going to get complicated, but I guess this is just the beginning of amazing tools that will change the landscape of how text is created and analysed

Question 6a:

Please comment on anything else you noticed.

Answers:

Very easy to use.

I never adjusted the settings for the text generator. In hindsight that might have been something to try.

Excited, willing to see what happens next, willing to change the way it was originally written and to receive new ideas.

I notice that part of this study is to measure the response to the idea that this generator creates intellectual property. I do not think so because it has no needs. It does not need to be relevant

or understood. This generated text is no more intellectual property than a heart shaped McMuffin caused by a glitch in machinery is.

Don't repeat a word multiple times on the first one or else it will go wrong

It's an interesting experiment. How useful it could be to working writer is debatable. I've used word generators in the past to help compose titles, tag lines and character names.

Having text written for me doesn't really fit with my needs as a pro.

funky

The idea of authorship and whether you could claim the work as yours is really interesting and definitely something that I think will become more important as this kind of technology becomes more widespread. Definitely a useful thing to have access to and something I would use personally.

The relevance and flow of the generated text reduces the further away from the original text it is. Should have option to limit generated text to e.g. just the next sentence or para

change the picture please

Lot of nonsense sometimes interesting good for experimenting getting started. Would suit short form not novels

It was certainly fun to use but I'm not altogether sure where it would be used in practice.

Also I guess the point was to use the same picture to provide comparisons but it would have been fun to have different options there

I'd like to see a Database (cloud-based) with true informations

Could see this being useful as a problem solver. May help in story creation to provide different potential story lines.

I think it's a very good idea but it might still need some work in the logic department. The generator can't quite grasp it yet.

The system reminded me of an autistic child I once met. He had loads of potential but found it hard to formulate original questions to enable him to engage in extended conversation successfully with a stranger unfamiliar with his condition. You had to adapt to his way of making sense of the world then it was fine. You had to do more work than usual but it was worth it because he was another human soul. Yet it was mentally and emotionally exhausting. Whether it is worth putting similar work in to adapt to a software system is debatable. Is it a tool or a collaborator? A slave or a colleague? How does slavery effect the enslaver psychologically? I would have expected a statement to do with it going through an ethics check and more information on what the study is. It is quite fascinating and I had forgotten the excitement I have had at research centres working with people who shared their ideas. No. All OK and worked well. Fast and responsive.

I think that it will be a long time before text generated by a machine can match the human imagination. The text provided seemed unconnected - or at least, hanging on by a mere thread of similar meaning.

I liked the photo. I thought it was fascinating and it enabled me to think of a story easily. sadly we are or have already entered this era - as long as the original thought is of value then it has merit - trouble is trying to find that these days

If a human creates a story or an article, it belongs to the human. It's copyrighted [Name of Human Author]. If AI creates a story or an article, does it belong to the AI, or to the creator of the AI? Who's is the intellect behind the words? Our legislation hasn't addressed any of the possibilities, has it?

For me, the best thing about this experience was having a random picture of a dog on a guy's shoulder to generate a story from. I felt I could go a long way with this and didn't need the generated text at all.

I think that an autocomplete (like gmail has) for finishing sentences is really handy, but I would want to be able to turn on & off the more complex "ideas" generator

Question 6b:

If you have heard of 'fake news' do you think this is relevant?

Answers:

With a program of this type, it would be easy to generate any number of texts that could supposedly be used in fake news.

No. Not even a little.

Oh yes! Too many people are not analyzing the sources. Maybe that is too big a job or most and maybe many do not know to do this. My son repeats this claptrap to me as if it is gospel and he does not understand me when I tell him that his source is untrustworthy. He thinks if he hears something from dozens of people it must be true. Mass media is powerful in that way.

I stay out of these type debates.

sure

It is an enabler of fake news (eg allowing much faster fake news to be created - especially if fed in the analytics re effectiveness of dissemination etc.) but could also make REAL news reporting more efficient and effective especially if it supports optimum use of human vs AI input

not really

Would not be convinced

no !

Yes, not sure how this would be related.

Fake news has an agenda, this is far more random, though guess anyone could use

no.

Oh, it would be perfect for that! :-)

Very difficult to unpack - particularly as the different sides each claim the other side is distributing fake news. Having followed the Syrian War closely, I know for a fact our major news channels spouted fake news. Could something like this be used in social media to reply with biased opinions? Perhaps, but the existing troll farms used by all sides are often quite sophisticated, with the least sophisticated being the least effective. So, this type of low level autogeneration would not be very effective.

I don't know.

It could certainly encourage the creation of fake news and make it a lot easier. It could also make it easier for small snippets of truth to be exaggerated.

dont get me started

Unless the AI used to generate the text is extremely advanced, I think any news text generated wouldn't be very effective -- fake or otherwise. I think, at this point in our technology, news of any kind has to be created by humans to effectively educate, manipulate, or make humans react like chickens with their [thinking] heads cut off. If a text generator is going to be effective at writing propaganda, it's skills are going to have to increase a heck of a lot. Ask again in five years.

Experimental Design Overview

Privacy

No personal data collected, except Age range and Occupation. All responses optional. Full GDPR, Privacy and Credits on all pages and in all preliminary material.

Data was stored on the secure University server, GDPR compliant. No personal data collected or saved.

Nonparametric tests are accurate with ordinal data and do not assume a normal distribution.

However, there is a concern that nonparametric tests have a lower probability of detecting an effect that actually exists. The Mann-Whitney test is an example of a nonparametric test.

Website

The various supporting texts around the website such as the full GDPR statement, Privacy Statement, Credits and Help pages for Text Synth generator, Acknowledgments etc. are available to researcher son request.

Stages/Pages

Procedure

The participants are set tasks, which involve writing and compiling text stories in the editor (right side), using their own and the generated text.

They are free to type into the generator or the editor, and copy/paste into the editor. This mimics a normal system and does not restrict what they achieve.

This is the minimal system for testing the ‘stakeholders using text generation in work’ study as writers will know how to copy and paste from one panel to the other.

Video instructions and Help will be are provided online and prior.

See Figure 1. Study flowchart (online pages).

Structure of Study

Begin, Help Video, Start

The first section of the study comprises the Begin page with Introduction, University status, Privacy and GDPR information, a Help video describing the entirety of the test process, and then a Start (...the study) Profile page with an occupation survey to gather basic respondent data.

After viewing the help video, there is a set of background profile (non-personal) questions before they start the experiments.

Participants were asked to provide score Likert scales and give feedback comments at all stages of the process, with, and a series of questions with Likert scores and feedback at the end. They could also use the generator again with no tasks, but still within the experiment design.

Image prompt

All experiments used the same unusual image. This was to prompt the respondent to write a few words or lines into the generator. It was preferred to a text prompt to give more flexibility in response. See Credits for image prompt source.

Minimal system

UI/UX

Use very simple familiar interface, no learning curve.

Basic website to focus on text tasks.

Use a writing prompt – an image, same for all, not selectable. Ambiguous prompt, small image to make respondents use their imagination.

Right side of screen

The generator: Text generation system GPT-2 (OpenAI 2019). Generate new text using a prompt. Many repeats allowed, each with selection and copy/paste into editor panel.

An interface to Text Synth (Bellard 2021) used with permission.

Right side of screen

The editor: custom HTML/CSS/JS. For adding to and editing the generated text (after copy/paste) and saving of new hybrid work, as per the experiment (of 3) set to the participant.

Text editor (HTML form, PHP data filing in CSV format).

Respondents

Mixed group writers from personal and my creativity software mailing list, writers/publishers forums. No control group (non-writers).

Process

The design of the study means any use will provide data, even if the site is abandoned after Begin page. The Start profile page selections provide data on the type of writer, even if they abandon.

First experiment, Caption, provides information on using the generator; second News and third Fiction experiments are extra. This in itself provides data on interest level, although time constraints might also apply as not a monitored study.

Engagement

Levels of engagement are assessed on each section and overall as follows:

Likert scales – positive to negative

Amount of text produced – assume more is positive

Sentiment analysis of feedback texts

Time on writing experiments and feedback

Depth of study – completed 3 experiments, 2, 1 or 0 (start only) - is reflected in overall time and volume of text.

Experiments, Questions and Results

Dates and Location

The experiment ran from 25 June to 28 August 2020. It was online at the Story Live website (Story Live 2021), which is now a public version of the generator/editor, with email as a save method (users email themselves any desired creations).

Respondents

Respondents (82 in total) were stakeholders in various writing areas.

Stakeholder groups, with numbers conducting the study, were:

Friends and Relatives – 39. Selected friends and relatives. Largely professional writers and artists, academics and some professional and amateur writers and musicians.

Story Software users – 32. Geoff Davis design and produced a creativity app for writers and artists. This has a mailing list of active users. These are writers/creatives of various sorts.

Self Publishing Forum SPF – 9. Professional body for independent publishers. Many of these small publishers have top selling books so it is a professional group.

Unbound Forum – 2. The ‘Unbound’ publisher’s author forum.

The majority were in my own groups Friend and Relatives and Story Software list (87%).

This might introduce bias as the software group are a computer-friendly set, but so is the target audience of creativity tool users.

Some respondents only completed the Start section (leaving profile data but did no experiments and providing no feedback), others part or fully completed the study.

Data

Responses to feedback questions

Questions and comments were processed for volume and sentiments. The actual feedback texts are below. These are included as they are illuminating but difficult to paraphrase.

Indexed data is available to researchers on request.

Not included:

Raw Data – the submitted texts from the three experiments,

Raw Data – the timing, navigation, observed and collected data.

These texts, and other data, are available to researchers on request.

Security

The experimental website was hosted on a top level domain (storylive.com) and with HTTPS (Hypertext Transfer Protocol Secure). This is an internet protocol that protects the integrity and confidentiality of data between the respondent's computer and the website. This ensures security and privacy.

Likert Scales

There is some discussion about how to rank Likert scales. This study used a 5-point scale:

Highly Agree 1 Agree 2 Neutral 3 Disagree 4 Highly Disagree 5.

I have occasionally processed data to show average of Likert scales (so can get an average of 2.5) and a more robust positive/negative summing (responses 1, 2 are positive; 4, 5 are negative), as well as a median value. Average of Likert is not matching data but provides a scaled value for charting.

Experiment Online Design

Stages/Pages

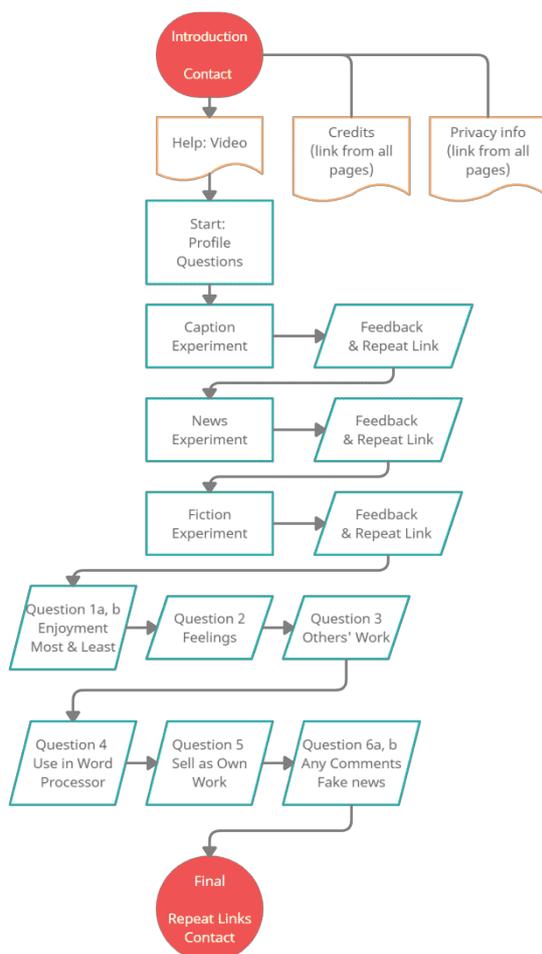
Procedure

The participants are set tasks, which involve writing and compiling text stories in the editor (right side), using their own and the generated text.

They are free to type into the generator or the editor, and copy/paste into the editor. This mimics a normal system and does not restrict what they achieve.

This is the minimal system for testing the ‘stakeholders using text generation in work’ study as writers will know how to copy and paste from one panel to the other.

Video instructions and Help will be are provided online and prior.



Above: Flowchart of Study

Structure of Study

Begin, Help Video, Start

The first section of the study comprises the Begin page with introduction, UAL status, privacy and GDPR information, a Help video describing the entirety of the test process, and then a Start (...the study) Profile page with an occupation survey to gather basic respondent data.

Image prompt

All experiments used the same unusual image. This was to prompt the respondent to write a few words or lines into the generator. It was preferred to a text prompt to give more flexibility in response.

CREATIVE WRITING with help from computer generated text, using this image



In this short experiment, you will:

- look at an image
- then think of key words (or others) to type into a computer text generator
- press 'complete' and the generator produces related sentences (these are not sampled from existing text but freshly generated each time)
- copy the generated text into an editor and
- produce the final piece.

Geoff Davis Summer 2020:

Thank you for helping with this research: I am doing this study as part of my research at the [Creative Computer Institute, University of the Arts London](#), into how people use text with computers. This will be used in new creativity software. My Supervisor is Prof. Mick Grierson, Research Leader, UAL CCI. Thanks to [Fabrice Bellard](#) for use of his excellent Text Synth modified and enhanced GPT-2 system.

Powered by [textsynth](#).

HELP VIDEO - WATCH FIRST

[More Notes and Privacy and GDPR information](#). No personal details are saved during the research.

Where this information is unrelated to the business of University of the Arts London the opinions expressed in it are the opinions of the researcher and do not necessarily constitute those of University of the Arts London.

[Read details of UAL and Study ethical and privacy policy here.](#)

[Privacy](#) [Credits](#) [Start](#)

Begin page:

After viewing the help video, there is a set of background profile (non-personal) questions before they start the experiments.

Participants are asked to provide score Likert scales and give feedback comments at all stages of the process, with, and a series of questions with Likert scores and feedback at the end.

They also get a chance to use the generator again with no tasks, but still within the experiment design.

Minimal system

UI/UX

Use very simple familiar interface, so no distractions. Basic website to focus on text tasks.

Use a writing prompt – an image, same for all, not selectable. Ambiguous, small image prompt to make respondents use their imagination *ref*

Caption Experiment page (first)

WRITE A CAPTION by adding words about the image to left side, 'Complete Text', then edit on right side.

Text Synth

Text completion using the GPT-2 language model. It is a neural network of up to 1.5 billion parameters. Type a text and let the neural network complete it. Each try returns a different randomly chosen completion.

The same model can be used to [compress text messages](#).

Model: top-k: top-p: temperature: seed:

Type your text here. It will be automatically completed with a (hopefully) realistic text.

© 2019-2020 [Fabrice Bellard](#) - [Technical notes](#)

Look at the photo. Get ideas for words to start the generator.

To generate text, ignore 'Select an example', type some words in, then press Complete Text

Generate another, if you want, then copy and paste into the editor below and edit or add new text

Paste generated text into box, edit if you want, click here to go to next

[Privacy](#) [Credits](#) [First experiment](#)

2 x iframes, left and right; plus writing prompt image (same for all)

Left of screen:

The generator: Text generation system GPT-2 (OpenAI 2019)

Interface Text Synth (Bellard 2021)

Right side of screen:

The editor. Custom HTML/CSS/JS.

Editor for creation and saving of new hybrid work, as per the experiment (of 3) set to the participant.

Text editor (HTML form, PHP data filing in CSV format)

Respondents

No control group (non writers)

Mixed group writers from personal and my creativity software mailing list, writers/publishers forums.

Privacy

No personal data collected, except Age range and Occupation. All responses optional. Full GDPR, Privacy and Credits on all pages and in all preliminary material.

Process

The design of the study means any use will provide data, even if the site is abandoned after Begin page. The Start profile page selections provide data on the type of writer, even if they abandon.

First experiment, Caption, provides information on using the generator; second News and third Fiction experiments are extra. This in itself provides data on interest level, although time constraints might also apply as not a monitored study.

Engagement

Levels of engagement are assessed on each section and overall as follows:

Likert scales – positive to negative

Amount of text produced – assume more is positive

Sentiment analysis of feedback texts

Time on writing experiments and feedback

Depth of study – completed 3 experiments, 2, 1 or 0 (start only) - is reflected in overall time and volume of text.

Experiments, Questions and Results

Dates and Location

The experiment ran from 25 June to 28 August 2020. It was online at the Storylive website (Storylive 2021), which is now a public version of the generator/editor, with email as a save method (users email themselves any desired creations).

Respondents

Respondents (82 in total) were stakeholders in writing fields. Stakeholder groups, with numbers conducting the study, contacted were:

FR – 39 – Friends and Relatives - selected friends and relatives – a group of professional writers and artists, academics and some professional and amateur writers; musicians.

SS – 32 – Story Software users – I have a creativity app for writers and artists, this has a mailing list of purchasers. These are all writers/creatives of various sorts.

SPF – 9 – Self Publishing Forum – this is a professional body for independent publishers. Many of these small publishers have best sellers so it is a professional group.

UB – 2 – Unbound Forum– the publisher’s author forum.

The majority were in my own groups FR and SS (87%). This might introduce bias as the software group are a computer-friendly educated set, but so is the target audience of creativity tool users.

Some respondents only completed the Start section (leaving profile data but did no experiments and providing no feedback), others part or fully completed the study.

Start profile page

SHORT QUESTIONS

No personal details are saved. These are optional, but is helpful if you answer these, as there might be a relationship between them and how you do the experiment. Thanks!

If you would like a copy of **research findings** please email me to go on the list. This is still anonymous as no personal information is collected.

Quick questions

Have you used a text generator before?

Yes

Can you recall the name of the system, or any comments?

Age: under 18 18-29 30-39 40-64 65 plus

How would you describe your writing?

Amateur or occasional

Professional

What are your main writing activities (as many as you like):

Academic, teaching

Artist

Copywriter, Non-fiction

Fiction

Journalist

Memoir, History

Plays, scripts

Poet

Report writer

Scientist

Scribbler, for fun

Student

Other

Save and first experiment

[Privacy](#) [Credits](#) [Start](#)

Experiments

News Experiment page (second)

WRITE A NEWS ITEM. Longer text, one or two paragraphs.

Text Synth

Text completion using the [GPT-2](#) language model. It is a neural network of up to 1.5 billion parameters. Type a text and let the neural network complete it. Each try returns a different randomly chosen completion.

The same model can be used to [compress text messages](#).

Model: top-k: top-p: temperature: seed:

Select an example

Type your text here. It will be automatically completed with a (hopefully) realistic text.

© 2019-2020 Fabrice Bellard - Technical notes

Look at the photo. Get ideas for words to start the generator.

To generate text, ignore 'Select an example', type some words in, then press



Generate another, if you want, then copy and paste into the editor below and edit or add new text

Paste generated text into box, edit if you want, click here to go to next

Chance to repeat experiment

[GO BACK to last page to do another news item](#)
[or go to NEXT page](#)

News feedback Likert and comments page (second experiment)

NEWS

This second time, did you like using the text generator more or less?:

A lot A little Neutral Slightly disliked Strongly disliked

What did you think about using the text generator?
Did it help or hinder or no effect?
How did it help you?

[Privacy](#) [Credits](#) [First experiment](#)

Feedback Questions 1-6

The 6 Likert questions with feedback comments, which come after the 3 experiments, are designed to capture different aspects of using a computer system to create actual text, rather than using a text and document enabler.

Scales 1-5 (one needed reversing to keep scales Y positive)

WRITING EXPERIMENTS STUDY

THANKS FOR HELPING. PLEASE ANSWER 6 SHORT QUESTIONS

Please add any comments, your feedback is important.

Q1/6: Overall, did you enjoy using the text generator?:

A lot A little Neutral Slightly disliked Strongly disliked

What was **most interesting**? Please explain.

What was **least interesting**? Please explain.

Save

[Privacy](#) [Credits](#) [First experiment](#)

WRITING EXPERIMENTS STUDY

Q2/6: I felt emotional when the generated text appeared:

Strongly agree Agree Neither Disagree Strongly disagree

Which emotions did you feel, please describe?

Save

[Privacy](#) [Credits](#) [First experiment](#)

WRITING EXPERIMENTS STUDY

Q3/6: Do you feel that you have used somebody else's work?

Strongly agree Agree Neither Disagree Strongly disagree

Does this seem relevant and why?

Save

[Privacy](#) [Credits](#) [First experiment](#)

WRITING EXPERIMENTS STUDY

Q4/6: Do you think you would use this, or a similar feature, in an ideas editor or word processor?

Strongly agree Agree Neither Disagree Strongly disagree

Imagine a generator in your ideas editor or word processor. What extra features would you like?

Save

[Privacy](#) [Credits](#) [First experiment](#)

WRITING EXPERIMENTS STUDY

Q5/6: Do you think you could sell this as your own work?

Strongly agree Agree Neither Disagree Strongly disagree

Do you have any comments on notions of authorship, plagiarism, ownership and intellectual property?

Save

[Privacy](#) [Credits](#) [First experiment](#)

WRITING EXPERIMENTS STUDY

Please comment on anything else you noticed.

If you have heard of 'fake news' do you think this is relevant?

Save

[Privacy](#) [Credits](#) [First experiment](#)

After each question there is a chance to repeat or progress.

Question 1 repeat choice

[TO REPEAT QUESTION, GO BACK](#)

or [GO TO NEXT QUESTION 2](#)