

ARTIFICIAL INTELLIGENCE (AI) WON'T BE WRITING AWARD WINNING SCRIPTS SOON – BUT IT COULD HELP HUMANS WHO DO

*A preliminary story and script development
evaluation framework AI systems' output*

Peter Dodds

Digital Frontiers Pty Ltd, Australia

ABSTRACT

Assessing the quality and relevance of the output of Natural Language Generation (NLG) systems is a challenge. It is difficult to assess empirically. The proposed preliminary Story and Script Evaluation Framework (SSEF) seeks to address this by combining qualitative and quantitative methods to evaluate AI-created material using four criteria – plotting codified story elements on Freytag's Pyramid, measuring emotional connectedness and reactions, assessing if the created scenes or the story flows logically and determining how real or plausible the story is. The planned framework is flexible enough to work with different story genres, though it is primarily designed for scripts and screenplays and short or long-form novels. A key feature of SSEF is that it examines the AI-generated content from the point-of-view of the reader or audience member. It is focused on the impact a story has on the individual and not on the technology or adherence to a particular narrative theory or story genre. Developing techniques to streamline and assess the emotional criteria requires a deep understanding of emotions, emotional connection and emotional responses and the bond between author/writer and an audience. To do this successfully involves recognising the importance of empathy and emotional connection in storytelling. Injecting empathy could also go some way to enabling AI to create contextually correct, emotionally challenging stories. If or when AI achieves high level, intense connections with an audience, then its storytelling will have the ability to be more immersive, more challenging, more compelling and ultimately, more enjoyable. Part of the evolution of NLG is the development of tools like the Story and Script Evaluation Framework which can provide another way to refine the creation of a story and mitigate the issue of AI not knowing the meaning of the sentences it was creating or the implications of decisions made by the generating engine about emotional depth for a story.

INTRODUCTION

This paper proposes a preliminary evaluation framework for story and script development that identifies four criteria - creative, emotional, information flow and realism – as a way to classify and review Artificial Intelligence (AI) authored materials. The evaluation process will be undertaken from the point-of-view of the reader or the audience and not focussed on the technology, dataset or natural language generation characteristics. As an alternative to Untrained Automatic Metrics techniques, SSEF will also allow evaluation of human-authored material, facilitating a side-by-side comparison of, for example, written output from AI and humans created from the same brief from the point of view of the audience. Techniques like Untrained Automatic Metrics look only at the text and not its affect upon a person. To undertake an evaluation using the suggested criteria will require codifying story elements, measuring emotional connectedness and reactions, assessing if the created scenes or the story flows logically and determining how real or plausible the story is.

Using narrative, imagery and drama to communicate, a story's purpose is to evoke an emotional response and move a person in some way. Stories are either descriptive or narrative in form and include poetry, fiction, short stories, scripts and screenplays. Irrespective of the form, these styles are all made up of the same core elements and all have an objective to engage and emotionally affect an audience.

Automatic story creation using AI requires Natural Language Generation (NLG) technologies to create long, coherent passages that realistically express a logical progression of events in the best way possible. Artificial Intelligence (AI) has had some success in writing expository passages that have featured in newspaper stories and as background profiles of, for example, sport stars. At present, the approximate 1500-word limit is a function of the level of development of NLG technology. It is an emerging technology with many unsolved problems and challenges arising from data sparsity and complexity and the dynamic characteristics of data available for the system to use to learn.

There are examples of scripts and short stories created using any one of several natural language generation engines. Some scripts have also been made into short films which are viewable but not (yet) Academy Award winners. The current projects out in the public domain have all found ways to work around shortcomings of AI such as the word limit or the use of less than perfect data sets. The other key limitation is the absence of empathy in AI material. Empathy and developing an emotional connection with an audience are crucial for AI to be a really useful tool to either create stories and scripts or to be used to assist human authors and writers. While cognitive and intellectual empathy can be learnt, emotional empathy must be experienced. This is something AI cannot do – it can learn about but cannot truly understand it because it has not experienced it. This presents a big challenge for AI.

The proposed framework will be suitable and adaptable for use across all five writing genres – expository, descriptive, narrative, persuasive and journals and letters. However, the current focus is on fiction, scripts and screenplays as these areas offer the greatest challenge as well as the widest opportunity to innovate and build a body of knowledge of methods and techniques to evaluate, review, edit and curate AI authored text. There are techniques such as extrinsic or task-based evaluation, subjective human ratings and metrics-based using automated systems such as LEPOR, ROUGE, BLEU and METEOR. The automated solutions are predominantly linguistic-based and summarisation-based and do not evaluate the AI material from the point-of-view of the intended audience. Both task-based and subjective human evaluations are currently manual processes and are time-consuming. The preliminary Story and Script Evaluation Framework (Figure 1) proposed, addresses the time and cost issues by combining the existing understanding of story structure with novel data collection

methods to build a body of knowledge of emotional reaction and intensity married to the story elements and viewed from the point-of-view of the audience.

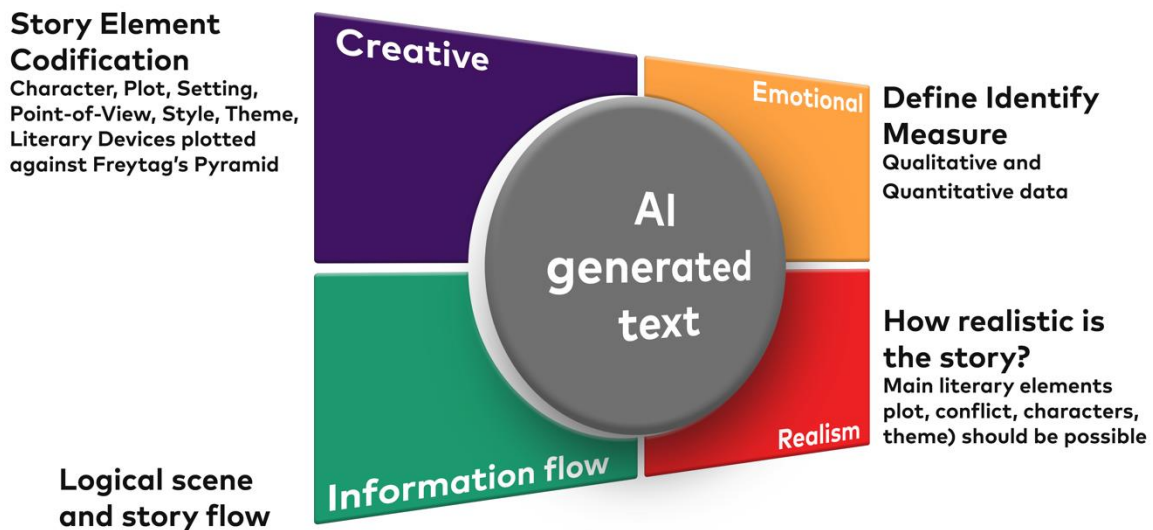


Figure 1. Preliminary Story and Script Evaluation Framework.

Much of the activity in the sphere of AI and storytelling/story creation is around mechanics and structure. Exponents are like builders and landlords, more focussed on the building (what it is) and not who the tenants are (who is creating material) and what they are doing (their creations) and how much people are enjoying the structure (their emotional reaction or attachment). AI can write fiction, poetry, short stories, scripts and screenplays – but it needs human input to outline a premise and basic story elements. AI cannot think on its own. It needs a reference. While Natural Language Generation uses knowledge of the art of how humans communicate it does not know about creativity or the spark of an idea.

THE PRELIMINARY STORY AND SCRIPT EVALUATION FRAMEWORK (SSEF)

The framework combines existing knowledge about story and script creation with proposed new and novel techniques to obtain high-quality data that can be used to evaluate a story (Figure 2).

The Information Flow and Realism criteria are the more straightforward of the chosen evaluation streams. In its simplest sense, the information flow could be expressed as a binary response – yes, it does flow or no, it does not flow logically. However, more complex rankings for story Information Flow, for instance, could be employed where a range of literary or film techniques such as flash forward and flashbacks, contrived jump cuts, dream sequences and hallucinations are used to propel the story forward. While story realism is subjective, it will be examined in the context of the genre and literary style of the story. A sci-fi fantasy story would have a different set of metrics than a love story or an epic adventure for example.

For the Creative criteria, there is a large body of existing research that examines stories, narrative structure and communication theory. However, for the purposes of the preliminary framework, it is the top-level concept of story structure that is the most helpful to the evaluation. The top-level components of character, plot, point-of-view, style, theme and literary device provide sufficient detail for appraisal of the story. The impact of changed circumstance on these components through the evolution of the story also provide specific points to map emotional responses to. There are already solutions that automatically extract data for annotations for

screenplays that would be suitable for use within the preliminary framework for scripts and screenplays [1]. Combining this technique with an advanced version of SSEF could allow scripts to be examined at each draft stage and improved.

| CRITERIA | ACTIVITY | EXISTING | ADAPTATION | NEW |
|-------------------------|--|---|--|--|
| Creative | Codification of story elements. | Use of existing nomenclature. | Ranking of characters, eg. ranging from protagonist (+1) to antagonist (-1). | |
| Emotional | Participant surveys. EDA/GSR & BVP measurement. | Standard survey format with Likert scales and limited option for subjective responses. | Evaluation based on previous survey results correlated with creative Codification rankings. Adaptation of ESA and BVP techniques and measurement to ascertain reactions to dramatic cues. | Top level evaluation of story Calculation of weighted value of particular type of emotional trigger synced to the exact point in the story where the trigger was revealed. Results will also be weighted for age, gender, socioeconomic and educational status. |
| Information Flow | Classify based on story complexity. | Use of existing nomenclature to categorise information flow based on the complexity of story structure and the use of literary devices to tell the story. | Ranking of characters, for example, ranging from protagonist (+1) to antagonist (-1). | Define a matrix that does not penalise simple, well structured and well executed information flow for a love story with a lower score when compared to a more complicated adventure story. |
| Realism | Examined in the context of genre. | Use of existing genre nomenclature and realism expectations to rank created material contextually on 5-point Likert Scale (Poor to Excellent). | | |

Figure 2. Preliminary Story and Script Evaluation Framework Criteria Breakdown.

Writing requires an idea and a goal, knowledge of language and grammar, curiosity, an understanding of situations, people and the world. Beyond the foundational elements of a setting, characters, a plot, conflict and resolution there are more complex items such as themes, morals and messages, symbolism and synchronicities and perspectives and points-of-view that bring greater depth to a story and potentially stronger connections between the writer and the reader. A good writer takes the reader on an author-specified emotional journey. By adopting techniques and tools utilised by psychologists, identifying and measuring the intensity of emotional reaction should create a way to place values on story points and their effect on a reader. Much of the recent research about AI created stories focuses on the mechanics of story creation and its structure and not what happens with the reader. To successfully evaluate a story or a script, the focus needs to shift to the point-of-view of the reader or the audience and their relationship with the text.

In terms of what is an effective way to engage a reader, stories developed from knowledge or experiences, which have strong emotional themes, are the most successful [2]. Neuroscientist Dr Paul Zak, leader of the team that discovered oxytocin is synthesised in the human brain, noted stories that capture our attention and involve us emotionally move us to do something –

we react [3]. He observes that stories with a dramatic arc are suitable for high impact narratives which in turn can engage us intellectually and emotionally. Building uncertainty through the arc facilitates character development and keeps us in suspense until the tension is relieved, the conflict resolved and the story concluded. This pattern fits neatly into Freytag's graphical interpretation, known as Freytag's Pyramid (Figure 3), of a story's dramatic structure from exposition to rising action to climax to falling action and then resolution [4].



Figure 3. Based on "Freytag's Technique of the drama"[4]

The importance of empathy and emotional connection in storytelling

Emotional stories resonate with readers. Author Turney [5] writes that "... readers love stories that fuel them with emotion and transport them away from their everyday lives." She further explains that it is emotion, garnered through empathy, that could be the difference between average writing and something that becomes a best seller. A reader should feel that they know the character personally. This key ability possessed by an author fosters the reader-writer connection. Several studies have indicated that empathy is a significant requirement for the audience to enjoy a story [6].

Empathy has largely been looked upon as a necessity for AI to interact directly with humans by recognising and responding appropriately to cues it observes. According to Andrew McStay "... emotions matter" as "... they are the core of human experience, [they] shape our lives" [7]. In his book "Emotional AI" he explains that while technological systems cannot experience emotions, they do have the capacity to recognise and label human emotions and respond accordingly with the pretence of understanding the emotion. Moutinho, writing about dramatist David Košťák's work with overseeing an AI-generated dramatic script, recounts his observation that the script, even partially assisted in places by humans, was difficult for the actors to understand. The play revealed that AI was not very good at the type of drama and emotion needed to make an engaging performance [8].

Empathy in a long-form story or screenplay is more complex and challenging than developing the ability to recognise and respond to human emotions in an empathetic way. In response to

the question “can AI have emotions?” Bharath argues that while it is impossible with current technology for AI to replicate human emotions, it is possible to mimic some emotional characteristics [9]. New AI technologies are learning and identifying human emotions (Figure 4) and using the knowledge gained to improve everything from marketing campaigns, sales pitches through to healthcare and intelligent chatbot business assistants [10].



Figure 4. Human emotional reactions. Source: Shutterstock

A challenge for AI writing with empathy lies with a writer's perception of who their audience is. Writers write to reflect their aspirations. This presents challenges for AI as it does not aspire. A writer chooses courses of action determined by what they see as their ideal readers [11]. In doing this, the accuracy of the guesses a writer makes ultimately affects the value of the finished work. Robinson explains “... if they are able to craft a story that speaks to their intended audience's interests and correctly anticipates its reactions, they can consider it successful. Otherwise, this work might not find any audience at all.” This is a challenge for AI for, in addition to mimicking empathy, it must also have some concept of the audience to whom the material is being presented.

Like a sociopath, AI can be taught cognitive and intellectual empathy. Christov-Moore et al note that a likely reason current AI iterations are not designed to assess context and outcomes coherently is that AI lacks the ability to shift focus to the evolving and unanticipated demands of writing a story and the changing complexity of the emotional relationships between the protagonists within a story [12].

Zhao explores the complexity of character interactions and how an audience reacts to characters' roles as the story develops. When discussing empathy, he explains that an individual needs to put themselves in another person's place or situation to develop empathy with them. [13]. This is as true of characters in stories as it is with real people. This concept of perspective-taking is important in developing an affinity with the narrative and the characters. Traditional media, such as films and television, use empathy to get an audience emotionally involved as it builds an emotional bond between characters, the story and the audience. This helps a person feel the character's emotions vicariously.

Measuring emotions

The American Psychological Association definition of “emotion” describes it as a “... *complex reaction pattern, involving experiential, behavioural, and physiological elements, by which an individual attempts to deal with a personally significant matter or event.*” [14].

There are many methods available to measure emotions and emotional connection. However, there is no one gold-standard method [15]. Scherer further explains that no objective way exists to definitively measure a person’s subjective experience. Emotions are real and spontaneous, prompting immediate bodily reactions to a range of circumstances such as threats and confrontations through to happiness and joy. The methods available to capture data include self-reporting, measuring bodily reactions and changes in facial expressions or muscles [16].

The proposed method for measuring emotional responses and connectivity for the research portion of this project is to produce audiobook forms of the AI stories or scripts and use a variety of methods to capture data. Using an audiobook style of delivery of the content removes the effects of level of educational, literacy and environmental variables that could have an impact on the results. The proposed qualitative methods are combined Likert scale question surveys and the use of the Geneva Emotion Wheel. The Geneva Emotion Wheel, developed by the Université de Genève, is a theoretically derived and empirically tested instrument to measure emotional reactions. Quantitative data will be collected using Electrodermal Activity/Galvanic Skin Response (EDA/GSR) Sensors and Blood Volume Pulse (BVP) sensors (Figure 5).



Figure 5.

Applying these techniques to measuring emotional connection and intensity will result in a large pool of qualitative and quantitative data married to the audiobook timeline. The results will have complexity as they will represent an emotion, an intensity, duration of intensity, a

person's pulse and oxygenation plus a description of the type of dramatic episode experienced. EDA/GSR offers no insights into the valence (pleasantness or unpleasantness) of the emotional responses, and this is where there is a role for either an observer to capture and record a response or, preferably, for the participant to accurately complete a self-report or an interview discussing the complete timeline of experiences included in the study.

Abbasi and Beltiukov, in research undertaken in 2019, have investigated a way to categorise emotions within blogs by examining the text only [17]. Once categorised, the emotions are given a weight depending upon their perceived intensity. By multiplying the frequency by the weight assigned to the emotion, the intensity of the identified emotion was determined. Adapting this method of text analysis, by turning Abbasi and Beltiukov approach on its head, provides a way for an EDA/GSR emotional value to be calculated for a chosen point on, for example, a plotting of emotional reactions on Freytag's Pyramid for a particular point in time. In a paper presented at the 2019 Italian Association for Computing Machinery CHIItaly Conference, Warpechowski et al highlights assessing users' emotional reactions are important in modern technological development. [18]. They write that "... *providing easy-to-use and enjoyable emotion tagging tools is of key importance in encouraging users to share their reactions.*" Their research gravitated to using the Geneva Emotion Wheel (GEW) as the instrument to capture and measure emotional reactions. GEW is one of the better known and widely used non-list means to capture levels of emotional reaction used by psychologists (Figure 6). It allows a participant to self-report their reactions by selecting applicable emotion and its magnitude.

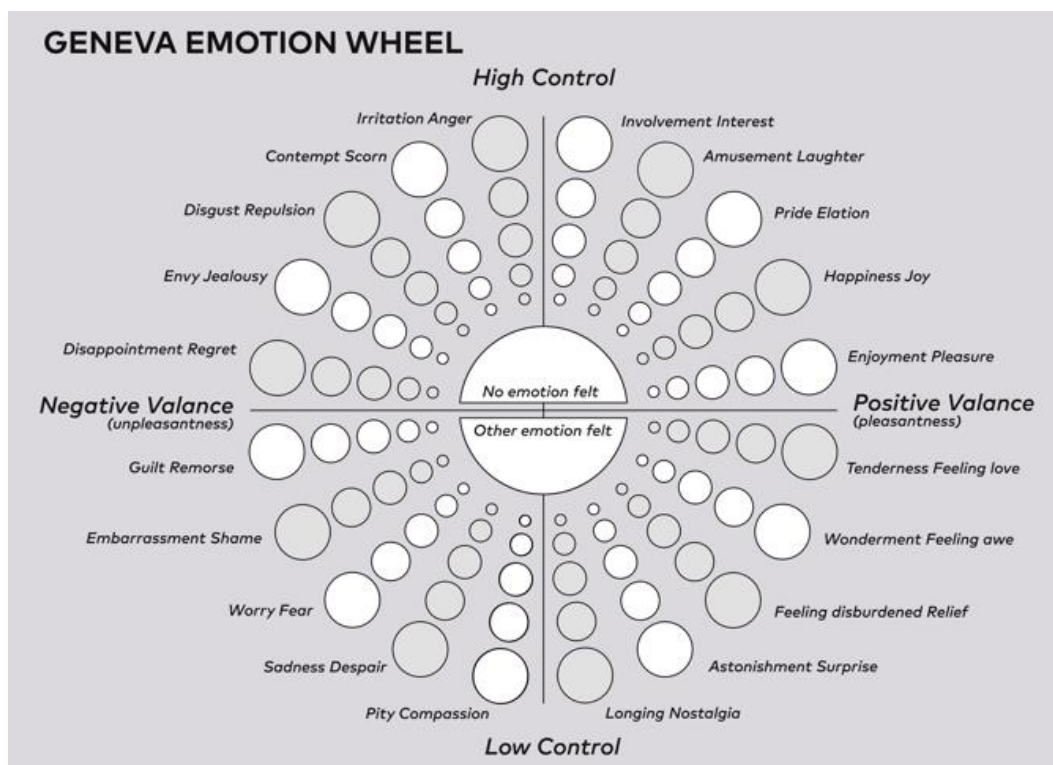


Figure 6. Geneva Emotion Wheel (GEW) (Referenced from <https://positivepsychology.com/emotion-wheel/>)

Self-reporting through surveys and using the Geneva Emotion Wheel has the advantage of acknowledging that emotions are essentially perceptual and self-referential. It presents a collection of twenty emotions which are classified against Valence (Unpleasantness or

Pleasantness) and Control/Power (High/Low) criteria. GEW has numerous characteristics that give it an advantage over other reporting methods. These include a reduction in the importance of word labels, incorporation of emotional intensity, and coverage of 'no emotion' states.

Limitations of Natural Language Generation

In a 2021 Forbes article, writer Rob Toews discusses the limitations of current AI technology and the opportunities addressing these limitations present for the future. He makes the point that for critics of AI, they see the limitations and shortcomings as recognition that the pursuit of AI has failed or is misguided. Toews counters this argument, stating that the identified limitations are an inventory of what must be solved to further advance AI [19].

His article identifies four things that AI cannot do – use common sense, learn continuously and adapt on the fly, understand cause and effect and reason ethically. Humans use a vast, generally unspoken body of common-sense everyday knowledge in their day-to-day activities. Toews uses the example of a person going to a restaurant, ordering a steak and leaving a big tip. In response to the question “what did the person eat?”, a human would answer a steak, but AI would be unable to respond to the question as it would not be capable of fully understanding context unless it learnt about the elements in the sentence – what happens at a restaurant, what ordering is and that a tip is part of the payment (this confirms that the item ordered was supplied) (Figure 7).

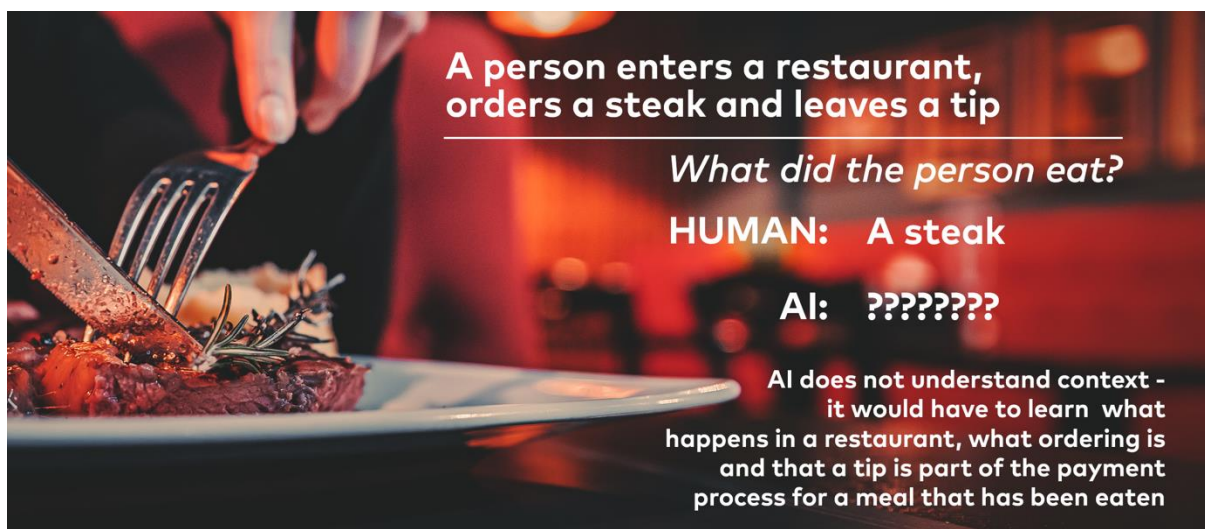


Figure 7. AI does not understand context.

Attempts at AI narrative such as Ross Godwin and Kenric McDowell's “1 The Road” and “Dinner Depression” were developed entirely from deep learning. “The Day a Computer wrote a Novel” created by the Kimagure Artificial Intelligence Writer Project and Nick Monfort's “World Clock” are interesting but are not Tolstoy or Jane Austen. Apocalyptic author Mike Pearl has co-written material with the GPT2 engine. In his blog, Pearl describes a process where he writes the introduction, and the AI completes the following paragraphs. He draws attention to the inherent creepiness of the AI-generated material that flows from his use of melancholic imagery in his introduction [20]. It is also easy to confuse the AI. In another test, he writes about the dissection of a lung in the style of a food writer or a person preparing a recipe. The AI continues in this style with gruesome instructions on cutting the human lung into 1cm cubes and dicing it with other ingredients.

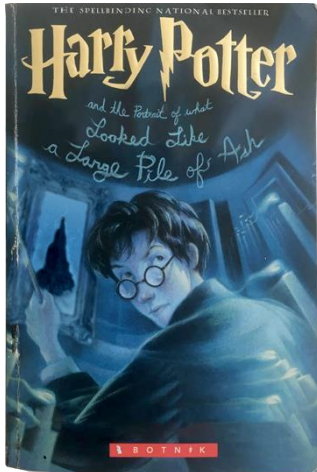


Figure 8. Source: Botnik
<https://botnik.org/content/harry-potter.html>

Pennsylvanian company Botnik, known for their creative language remixing keyboard, used their remixing software to create a Harry Potter chapter. Using predictive technology, Botnik allowed the software to write a three-page chapter. While the chapter does share some of the linguistic DNA of J. K Rowling's writing, it does descend into nonsensical text. The Botnik team fed all seven Harry Potter books through their predictive text keyboard and then allowed it to create a book chapter: "Harry Potter and the Portrait of What Looked Like a Large Pile of Ash". This 13th chapter of a "long lost" Harry Potter novel, called "The Handsome One," follows Harry, Ron, and Hermione as they climb Hogwarts castle to battle a group of Death Eaters (Figure 8). The chapter started a trend of people recording YouTube videos and reading the material. In 2017, the UK edition of The Guardian declared the trend so popular that it became the 4th best internet moment of 2017 [21].

The story almost makes sense. It has some examples of silly dialogue between the characters. In one example Harry, Ron and Hermione are confronted by a closed door. "*They looked at the door, screaming about how closed it was and asking it to be replaced with a small orb. The password was 'BEEF WOMEN,' Hermione cried.*" Some phrases sound real and contextually J.K. Rowling-ish while others are absurd. Examples include "*Ron's Ron shirt was just as bad as Ron himself*", "*Harry tore his eyes from his head and threw them into the forest*", "*He saw Harry and immediately began to eat Hermione's family*" and "*Not so handsome now,' thought Harry as he dipped Hermione in hot sauce*".

Moutinho writes about the creation of a one hour play titled "AI - When a robot writes a play" where the protagonist, a robot, goes out into the world to learn about people, emotions, life and death. Created using GPT-2 and working within the roughly 1,000-word text output limitation, the theatre production team built the play by entering two sentences of dialogue where one of the two characters chat about human feelings and experiences to start each of the play's eight scenes [8].

Luke Dormehl, writing in Digital Trends, notes that AI creativity is improving fast. There are many examples online featuring AI written films. "Solicitors", directed and filmed by Jacob Vaus, used a human decided film topic as its inciting incident. [22; 23]. The resultant film with its twists is watchable and hilarious. Calamity AI has produced other AI written films including "Date Night" which ends with the cast discussing how they are in a movie. The film also has the memorable line "You slapped a volcano" in the hypnotising scene. Parts of the movie are believable but there is a point midway through the movie where AI lost the plot – literally [24; 25].

If not writing stories and scripts, how can AI be used to help writers?

At the current time, AI is behind solutions ranging from applications that check grammar and plagiarism to apps that can write entire articles or stories. It can be a resource for writers who seek inspiration. Suffering from writer's block, for example, a writer can access a wide range of software resources that provide story "triggers" to assist them. Copywriters have become big users of AI content creators as the software can recommend the best structure and vocabulary to employ. These products simplify the writing process and enhance content quality to maximise reader engagement.

AI writing assistants can help at all stages of the writing process from ideation, to organising narrative structure, to deciding tone and refining style. It can also help with determining the readability of a writer's work using tests such as the Flesch-Kincaid Readability test or the Gunning Fog Index. Microsoft Word uses natural language AI to assist dyslexic suffers by suggesting sentence corrections rather than word by word corrections. Other features within Word include the replacement of gender-specific terms, such as fireman, with the gender-neutral term firefighter. These types of software features are being integrated into a wide range of writing assistant tools and grammar correction services. They support the process of creative writing, freeing up the writer to concentrate on the development of their idea while reducing the cognitive strain by allowing them to be better supported (for instance, by working out where to put the commas in sentences) and better organised (by building a history or index of resources collated for a project).

There are also research teams working on collaborative AI tools where a writer and the AI work interactively to develop a story. In an article for the UNSW Sydney Newsroom site, Knight describes the work of PhD student Rodolfo Ocampo who has created the GPT-3 Narrative Device application. Ocampo does not see AI replacing the human creative process but rather a tool to assist in the process [26]. AI Software has been developed to further assist writers to paraphrase their work using Natural Language Processing. Natural Language Processing has also been employed to organise research materials for writers.

CONCLUSION

AI written material lacks emotional depth. This is an issue for AI-generated text attempting to develop a connection with a reader or an audience. While AI can be innovative in its response to the trigger statements, the lack of emotional complexity is noticeable in the material created. At times, dialogue and plot are markedly illogical or out of kilter. Evaluating NLG output is necessary to ensure that the material created is of high quality, relevant and captivating. While emotional connection and empathy are important across all literary forms, it is especially important for descriptive and narrative material which must draw a reader or the audience in and engage with them on a deeper and, at times, more rewarding level.

A way to address this shortcoming is to use the knowledge gained through recognising and measuring emotional connectedness within the proposed Story and Script Evaluation Framework to provide input for a new data point within AI story creation systems. The new data point could represent emotional range values and emotions assigned to a character or a plot evolution point along a timeline (or on Freytag's Pyramid) for example. A character's dialogue, for instance, could be better emotionally focused and, with this focus, more able to satisfy a need to develop empathy with the reader. As a larger body of knowledge of emotions and emotional reactions is built, the greater the likelihood of SSEF delivering useful and accurate recommendations. At some point, the human involvement in the emotional evaluation stage may be able to be automated using the large body of data collected from the evaluation of many projects and the patterns identified through analysis of this data as applied to other AI works.

Many of the evaluative techniques currently available are either model specific (they measure the features of the model rather than the story) or require human feedback. SSEF has been designed to address these two shortcomings. Correlating story features and unique elements with judgments based originally on large pools of data from human participants tests the algorithms used in the NLG app and the other characteristics such as engagement, emotional attachment and enjoyment. SSEF aims to be a proxy for the summing of a wide range of reader or audience point-of-view opinions on a particular example of text generated by AI.

Technology now allows the writer to have more knowledge about their audience. With more knowledge, the quality and quantity of their output increase. SSEF has a role to play in this scenario by presenting possibilities for improvement of their creative endeavour. As digital watches and fitness trackers become even more ubiquitous, the more plausible it is for real-time feedback on films, streamed media, audiobooks and TV programs. While the Apple watch is not yet capable of Electrodermal Activity (EDA/GSR) (it can do electrocardiogram (ECG)), Fitbit and other devices have this capability. It is not unrealistic for those people wearing these devices, who opt in, to provide real-time to content creators who could use this pool of data and the other evaluation criteria to determine the success, or lack of success, of their project immediately. Should this eventuate, NLG engines could use this data to constantly improve the algorithms they apply to story and script creation and the tools used to work collaboratively with human writers.

Enhancing existing writer assistance tools with SSEF derived emotional scores for words, phrases, scenarios or plot developments, potentially frees a writer to add the greatest value they can to a script or story by taking the genesis of a great idea, adding creative flair and applying good judgement when integrating the best of AI suggested improvements. This will result in the production of a compelling, immersive and emotionally engaging work.

References

- [1] Winer, D. R. & Young, R. M. (2017). Automated Screenplay Annotation for Extracting Storytelling Knowledge. University of Utah. The AIIDE-17 Workshop on Intelligent Narrative Technologies WS-17-20.
<https://aaai.org/ocs/index.php/AIIDE/AIIDE17/paper/view/15869/15254>
- [2] Dahlstrom, M. F. (2014). Storytelling in science. *Proceedings of the National Academy of Sciences* Sep 2014, 111 (Supplement 4)
https://www.pnas.org/content/111/Supplement_4/13614
- [3] Zak P. J. (2015). Why inspiring stories make us react: the neuroscience of narrative. *Cerebrum: the Dana forum on brain science*, 2015, 2.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4445577/>
- [4] Freytag, G. (1894, 1900) *Freytag's Technique of the drama: an exposition of dramatic composition and art*. An authorized translation from the 6th German ed. by Elias J. MacEwan by Freytag, Gustav, 1816-1895; MacEwan, Elias J.
<https://archive.org/details/freytagstechniqu00freyuoft/page/n9/mode/2up>
- [5] Turney, D. (2021). Emotional Stories Readers Love to Connect With. *Christell*. Article.
<https://www.chistell.com/africanamericanfiction/emotional-stories-readers-love-to-connect-with/>
- [6] Hand, S. & Varan, D. (2009). Interactive Stories and the Audience: Why Empathy is Important. *ACM Computers in Entertainment*, Vol. 7, No. 3.
<https://dl.acm.org/doi/abs/10.1145/1594943.1594951>
- [7] McStay, A. (2018). Emotional AI: The Rise of Empathic Media. 10.4135/9781526451293.
https://www.researchgate.net/publication/326294717_Emotional_AI_The_Rise_of_Empathic_Media
- [8] Moutinho, S. (2021). Kinky and absurd: The first AI-written play isn't Shakespeare—but it has its moments: Artificial intelligence generates a story about a robot trying to understand humanity. *Science* 26 Feb 2021.
<https://www.science.org/content/article/kinky-and-absurd-first-ai-written-play-isn-t-shakespeare-it-has-its-moments>
- [9] Bharath, K. (2021). Can your AI Have Emotions? *Towards Data Science* 18 Feb 2021.
<https://towardsdatascience.com/can-your-ai-have-emotions-7efc35721e12>
- [10] Somers, M. (2019). Emotion AI, explained. MIT Management Sloan School – Ideas Made to Matter 8 Mar 2019. <https://mitsloan.mit.edu/ideas-made-to-matter/emotion-ai-explained>
- [11] Robinson, J. G. (2019). The Audience in the Mind's Eye: How Journalists Imagine Their Readers. *Columbia Journalism Review*. The Tow Center for Digital Journalism at Columbia's Graduate School of Journalism.
https://www.cjr.org/tow_center_reports/how#NAME?nalists-imagine-their-readers.php
- [12] Christov-Moore, L., Reggente, N., Vaccaro, A., Schoeller, F., Pluimer, B., Douglas, P. K., Iacoboni, M. & Kaplan, J. (2022). Are Robots Sociopaths? A Neuroscientific Approach to the Alignment Problem. <https://doi.org/10.31234/osf.io/6tn42>
- [13] Zhao, H. (2012). *Emotion-driven Interactive Storytelling*. Doctoral Thesis. Bournemouth University. <http://eprints.bournemouth.ac.uk/20507/>
- [14] APA. (2022). American Psychological Association definition of “emotion”.
<https://dictionary.apa.org/emotion>

- [15] Scherer, K. R. (2005). What are emotions? And how can they be measured? Sage Publications.
<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.818.2716&rep=rep1&type=pdf>
- [16] Bell, L., Vogt, J., Willemse, C., Routledge, T., Butler, L. T., & Sakaki, M. (2018). Beyond Self-Report: A Review of Physiological and Neuroscientific Methods to Investigate Consumer Behavior. *Frontiers in psychology*, 9, 1655.
<https://doi.org/10.3389/fpsyg.2018.01655>
- [17] Abbasi, M. M & Beltiukov, A. (2019). Summarizing Emotions from Text Using Plutchik's Wheel of Emotions. 7th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2019).
<https://www.semanticscholar.org/paper/Summarizing-Emotions-from-Text-Using-Plutchik's-of-Mohsin-Beltiukov/4ec9848781d9c8df5a970614bbbf6bf731dbdf4>
- [18] Warpechowski, K., Nielek, R., & Orzeszek, D. (2019). Tagging emotions using a wheel user interface. Tagging emotions using a wheel user interface. CHIItaly '19: Proceedings of the 13th Biannual Conference of the Italian SIGCHI Chapter: Designing the next interaction September 2019 Article No.: 23 Pages 1–5
<https://doi.org/10.1145/3351995.3352056>
- [19] Toews, R. (2021). What Artificial Intelligence Still Can't Do. *Forbes*. 1 Jun 2021.
<https://www.forbes.com/sites/robtoews/2021/06/01/what-artificial-intelligence-still-cant-do/?sh=701f3d0f6>
- [20] Pearl, M. (2019). Behold the incredible horror stories an AI will write if you let it - Inside the mind of GPT-2, which can't help but produce really dark stuff.
<https://theoutline.com/post/8302/ai-incredible-horror-stories>
- [21] Botnik. (2016). Harry Potter and the portrait of what Looked Like a Large Pile of Ash.
<https://botnik.org/content/harry-potter.html>
- [22] Dormehl, L. (2020). A.I. creativity is improving fast. This hilarious GPT3-generated film is proof. *Digital Trends*. <https://www.digitaltrends.com/features/solicitors-gpt3-future-of-filmmaking/>
- [23] Calamity AI. (2020), Solicitors | A.I. Written Short Film. Calamity AI.
<https://www.youtube.com/watch?v=AmX3GDJ47wo>
- [24] Calamity AI. (2021), A.I. Goes Meta | Date Night Short Film. Calamity AI.
<https://www.youtube.com/watch?v=xY1VWZPJ5FE>
- [25] Vaus, J. & Weiss, E. (2021). How We Made a Movie by an AI Script Writer. *Built In*.
<https://builtin.com/media-gaming/ai-movie-script>
- [26] Knight, B. (2022). Authors and AI unite: the arrival of AI-augmented writing. *UNSW Newsroom*. <https://newsroom.unsw.edu.au/news/art-architecture-design/authors-and-ai-unite-arrival-ai-augmented-writing>

ACKNOWLEDGEMENTS

The author would like to acknowledge the assistance of Xiaodi Huang PhD and Dr Anwaar Ulhaq of Charles Sturt University, Bathurst NSW Australia, and thanks Botnik.org who provided images for this paper.