



Article

# Paranoid Transformer: Reading Narrative of Madness as Computational Approach to Creativity †

Yana Agafonova <sup>1</sup>, Alexey Tikhonov <sup>2</sup> and Ivan P. Yamshchikov <sup>3,\*</sup>

<sup>1</sup> School of Arts and Humanities, Department of Philology, National Research University Higher School of Economics, 199034 St. Petersburg, Russia; yagafonova@eu.spb.ru

<sup>2</sup> Yandex, 10117 Berlin, Germany; altsof@gmail.com

<sup>3</sup> Max Planck Institute for Mathematics in the Sciences, Max Planck Society, 04103 Leipzig, Germany

\* Correspondence: ivan@yamshchikov.info

† This paper is an extended version of our paper published in Agafonova, Y.; Tikhonov, A.; Yamshchikov, I. Paranoid Transformer: Reading Narrative of Madness as Computational Approach to Creativity. In Proceedings of the International Conference on Computational Creativity, Coimbra, Portugal, 10 September 2020.

Received: 13 August 2020; Accepted: 20 October 2020; Published: 27 October 2020



**Abstract:** This paper revisits the receptive theory in the context of computational creativity. It presents a case study of a Paranoid Transformer—a fully autonomous text generation engine with raw output that could be read as the narrative of a mad digital persona without any additional human post-filtering. We describe technical details of the generative system, provide examples of output, and discuss the impact of receptive theory, chance discovery, and simulation of fringe mental state on the understanding of computational creativity.

**Keywords:** computational creativity; computational narrative; natural language generation; autonomous text generation; receptive theory; chance discovery

## 1. Introduction

The studies of computational creativity in the field of text generation commonly aim to represent a machine as a creative writer. Although text generation is broadly associated with a creative process, it is based on linguistic rationality and the common sense of the general semantics. In [1], the authors demonstrated that, if a generative system learns a better representation for such semantics, it tends to perform better in terms of human judgment. However, since averaged opinion could hardly be a beacon for human creativity, is its usage feasible regarding computational creativity?

The psychological perspective on human creativity tends to apply statistics and generalizing metrics to understand its object [2,3], so creativity becomes introduced through particular measures, which is epistemologically suicidal for aesthetics. While both creativity and aesthetics depend on judgemental evaluation and individual taste that depends on many aspects [4,5], the concept of perception has to be taken into account, when talking about computational creativity.

The variable that is often underestimated in the mere act of meaning creation is the reader herself. Although the computational principles are crucial for text generation, the importance of a reading approach to generated narratives is to be revised. What is the role of the reader in the generative computational narrative? This paper tries to address these two fundamental questions presenting an exemplary case study.

The epistemological disproportion between common sense and irrationality of the creative process became the fundamental basis of the research. It encouraged our interest in reading a generated text as a narrative of poetic madness. Why do we treat machine texts as if they are primitive maxims or

well known common knowledge? What if we read them as narratives with the broadest potentiality of meaning such as insane notes of an extraordinary poet or language expert? Would this approach change the text generation process?

In this paper, we present the Paranoid Transformer, a fully autonomous text generator that is based on a paranoiac-critical system and aims to change the approach to reading generated texts. The critical statement of the project is that the absurd mode of reading and the evaluation of generated texts enhances and changes what we understand under computational creativity. The absurd mode of reading is a complex approach analogous to reading poetic texts, which means accepting grammatical deviations and reinterpreting them as an extra level of emotional semantics, so the generated texts were read as if they had the broadest potentiality for interpretation. Absurd mode of reading for us is quite a demanding reading that accepts as many variants of figurative meaning as possible. Another critical aspect of the project is that the Paranoid Transformer resulting text stream is fully unsupervised. This is a fundamental difference between the Paranoid Transformer and the vast majority of text generation systems presented in the literature that are relying on human post-moderation, i.e., cherry-picking [6].

Originally, Paranoid Transformer was represented on the National Novel Generation Month contest (NaNoGenMo 2019, <https://github.com/NaNoGenMo/2019>) as an unsupervised text generator that can create narratives in a specific dark style. The project has resulted in a digital mad writer with a highly contextualized personality, which is of crucial importance for the creative process [7].

## 2. Related Work

There is a variety of works related to the generation of creative texts such as the generation of poems, catchy headlines, conversations, and texts in particular literary genres. Here, we would like to discuss a certain gap in the field of creative text generation studies and draw attention to the specific reading approach that can lead to more intriguing results in terms of computational creativity.

The interest in text generation mechanisms is rapidly growing since the arrival of deep learning. There are various angles from which researchers approach text generation. For example, van Stegeren and Theune [8] and Alnajjar et al. [9] studied generative models that could produce relevant headlines for the news publications. A variety of works study the stylization potential of generative models either for prose (see [10]) or for poetry (see [11,12]).

Generative poetry dates back as far as the work of Wheatley [13], along with other early generative mechanisms, and has various subfields at the moment. Generation of poems could be centered around specific literary tradition (see [14–16]); could be focused on the generation of topical poetry [17]; or could be centered around stylization that targets a certain author [18] or a genre [19]. For a taxonomy of generative poetry techniques, we address the reader to the work of Lamb et al. [20].

The symbolic notation of music could be regarded as a subfield of text generation, and the research of computational potential in this context has an exceptionally long history. To some extent, it holds a designated place in the computational creativity hall of fame. Indeed, at the very start of computer-science, Ada Lovelace already entertained a thought that an analytical engine can produce music on its own. Menabrea and Lovelace [21] stated: “Supposing, for instance, that the fundamental relations of pitched sounds in the science of harmony and musical composition were susceptible of such expression and adaptations, the engine might compose elaborate and scientific pieces of music of any degree of complexity or extent”. For an extensive overview of music generation mechanisms, we address the reader to the work of Briot et al. [22].

One has to mention a separate domain related to different aspects of the ‘persona’ generation. These could include relatively well-posed problems such as the generation of biographies out of the structured data (see [23]) or open-end tasks for the personalization of dialogue agent, dating back to the work of Weizenbaum [24]. With the rising popularity of chat-bots and the arrival of deep learning, the area of persona-based conversation models [25] is growing by leaps and

bounds. The democratization of generative conversational methods provided by open-source libraries (e.g., [26,27]) fuels further advancements in this field.

However, the majority of text generation approaches are chasing the generation as the significant value of such algorithms, which makes the very concept of computational creativity seem less critical. Another major challenge is the presentation of the algorithms' output. The vast majority of results on natural language generation either do not imply that generated text has any artistic value or expect certain post-processing of the text to be done by a human supervisor before the text is presented to the actual reader. We believe that the value of computational creativity is to be restored by shifting the researcher's attention from generation to the process of framing the algorithm [28]. We show that such a shift is possible since the generated output of Paranoid Transformer does not need any additional laborious manual post-processing.

The most reliable framing approaches are dealing with attempts to clarify the algorithm by providing the context, describing the process of generative acts, and making calculations about the generative decisions [29]. In this paper, we suggest that such an unusual framing approach as the obfuscation of the produced output could be quite profitable in terms of increasing the number of interpretations and enriching the creative potentiality of generated text.

Obfuscated interpretation of the algorithm's output methodologically intersects with the literary theory that deals with the reader as the key figure responsible for the meaning. In this context, we aim to overcome disciplinary borderline and create dissociative knowledge, which develops the fundamentals of computational creativity [30]. This also goes in line with the ideas in [31,32] regarding obfuscation as a mode of reading generated texts that the reader either commits voluntarily or is externally motivated to switch gears and perceive a generated text in such mode. This commitment implies a chance discovery of potentially rich associations and extensions of possible meaning.

How exactly can literary theory contribute to computational creativity in terms of the text generation mechanisms? As far as the text generation process implies an incremental interaction between neural networks and a human, it inevitably presupposes critical reading of the generated text. This reading brings a lot in the final result and comprehensibility of artificial writing. In literature studies, the process of meaning creation is broadly discussed by hermeneutical philosophers, who treated the meaning as a developing relationship between the message and the recipient, whose horizons of expectations are constantly changing and enriching the message with new implications [33,34].

The importance of reception and its difference from the author's intentions was convincingly demonstrated and articulated by the so-called reader-response theory, a particular branch of the receptive theory that deals with verbalized receptions. As Stanley Fish, one of the principal authors of the approach, put it, the meaning does not reside in the text but in the mind of the reader [35]. Thus, any text may be interpreted differently, depending on the reader's background, which means that even an absurd text could be perceived as meaningful under specific circumstances. The same concept was described by Eco [36] as so-called aberrant reading and implied that the difference between intention and interpretation is a fundamental principle of cultural communication. It is often the shift in interpretative paradigm that makes remarkable works of art to be dismissed by most at first, e.g., Picasso's *Les Femmes d'Alger (O. J. R. M.)* that was not recognized by artistic society and was not exhibited for nine years since it had been created.

One of the most recognizable literary abstractions in terms of creative potentiality is the so-called 'romantic mad poet' whose reputation was historically built on the idea that genius would never be understood [37]. Madness in terms of cultural interpretation is far from its psychiatric meaning and has more in common with the historical concept of a marginalized genius, who has some extraordinary knowledge. The mad narrator was chosen as a literary role for the Paranoid Transformer to extend the interpretative potentiality of the original text that could be not ideal in formal terms; on the other hand, it could be attributed to an individual with an exceptional understanding of the world, which gives more linguistic freedom to this individual for expressing herself and more freedom in interpreting her

messages. The anthropomorphization of the algorithm makes the narrative more personal, which is as important as the personality of a recipient in the process of meaning creation [38]. The self-expression of the Paranoid Transformer is enhanced by introducing nervous handwriting that amplifies the effect and gives more context for interpretation. In this paper, we show that treating the text generator as a romantic mad poet gives more literary freedom to the algorithm and generally improves the text generation. The philosophical basis of our approach is derived from the idea of creativity as an act of trespassing the borderline between conceptual realms. Thus, the dramatic conflict between computed and creative text could be solved by extending the interpretative horizons.

### 3. Model and Experiments

The general idea behind the Paranoid Transformer project is to build a 'paranoid' system based on two neural networks. The first network (Paranoid Writer) is a GPT-based [39] tuned conditional language model, and the second one (Critic subsystem) uses a BERT-based classifier [40] that works as a filtering subsystem. The critic selects the 'best' texts from the generated stream of texts that Paranoid Writer produces and filters the ones that it deems to be useless. Finally, an existing handwriting synthesis neural network implementation is applied to generate a nervous handwritten diary where a degree of shakiness depends on the sentiment strength of a given sentence. This final touch further immerses the reader into the critical process and enhances the personal interaction of the reader with the final text. Shaky handwriting frames the reader and, by design, sends her on the quest for meaning.

#### 3.1. Generator Subsystem

The first network, Paranoid Writer, uses an OpenAI GPT [39] architecture implementation by huggingface (<https://github.com/huggingface/transformers>). We used a publicly available model that was already pre-trained on a huge fiction BooksCorpus dataset with approximately 10k books with 1B words.

The pre-trained model was fine-tuned on several additional handcrafted text corpora, which altogether comprised approximately 50 Mb of text for fine-tuning. These texts included:

- a collection of Crypto Texts (Crypto Anarchist Manifesto, Cyphernomicon, etc.);
- a collection of fiction books from such cyberpunk authors as Dick, Gibson, and others;
- non-cyberpunk authors with particular affinity to fringe mental prose, for example, Kafka and Rumi;
- transcripts and subtitles from some cyberpunk movies and series such as Bladerunner; and
- several thousands of quotes and fortune cookie messages collected from different sources.

During the fine-tuning phase, we used special labels for conditional training of the model:

- QUOTE for any short quote or fortune, LONG for others; and
- CYBER for cyber-themed texts and OTHER for others.

Each text got two labels; for example, it was LONG+CYBER for Cyphernomicon, LONG+OTHER for Kafka, and QUOTE+OTHER for fortune cookie messages. Note that there were almost no texts labeled as QUOTE+CYBER, just a few nerd jokes. The idea of such conditioning and the choice of texts for fine-tuning was rooted in the principle of reading a madness narrative discussed above. The obfuscation principle manifests itself in the fine-tuning on the short aphoristic quotes and ambivalent fortune cookies. It aims to enhance the motivation of the reader and to give her additional interpretative freedom. Instrumentally the choice of the texts was based on two fundamental motivations: we wanted to simulate a particular fringe mental state, and we also were specifically aiming into short diary-like texts to be generated in the end. It is well known that modern state-of-the-art generative models are not able to support longer narratives yet can generate several consecutive sentences that are connected with one general topic. The QUOTE+LONG label allowed us to control the model and target shorter texts during the generation. Such short ambivalent texts

could subjectively be more intense. At the same time, the inclusion of longer texts in the fine-tuning phase allowed us to shift the vocabulary of the model even further toward a desirable 'paranoid' state. We also were aiming into some proxy of 'self-reflection' that would be addressed as a topic in the resulting 'diary' of the Paranoid Transformer. To push the model in this direction, we introduced cyber-themed texts. As a result of these two choices, in generation mode, the model was to generate only QUOTE+CYBER texts. The raw results were already promising enough:

*let painting melt away every other shred of reason and pain, just lew the paint to move thoughts away from blizzes in death. let it dry out, and turn to cosmic delights, to laugh on the big charms and saxophones and fudatron steames of the sales titanium. we are god's friends, the golden hands on the shoulders of our fears. do you knock my cleaning table over? i snap awake at some dawn. The patrons researching the blues instructor's theories around me, then give me a glass of jim beam. boom!*

However, this was not close enough to any sort of creative process. Our paranoid writer had graphomania too. To amend this mishap and improve the resulting quality of the texts, we wanted to incorporate additional automated filtering.

### 3.2. Heuristic Filters

As a part of the final system, we implemented heuristic filtering procedures alongside with a critic subsystem.

The heuristic filters were as follows:

- reject the creation of new, non-existing words;
- reject phrases with two unconnected verbs in a row;
- reject phrases with several duplicated words; and
- reject phrases with no punctuation or with too many punctuation marks.

The application of this script cuts the initial text flow into a subsequence of valid chunks and filters 30.1% of the generated pieces that could not make it through the filter. Here are several examples of such chunks after heuristic filtering:

*a slave has no more say in his language but he has to speak out!  
the doll has a variety of languages, so its feelings have to fill up some time of the day - to - day journals.  
The doll is used only when he remains private. and it is always effective.  
leave him with his monk - like body.  
a little of technique on can be helpful.*

To further filter the stream of such texts, we implemented a critic subsystem.

### 3.3. Critic Subsystem

We manually labeled 1000 of generated chunks with binary labels GOOD/BAD. We marked a chunk as BAD in the case it was grammatically incorrect or just too dull or stupid. The labeling was profoundly subjective. We marked more disturbing and aphoristic chunks as GOOD, pushing the model even further into the desirable fringe state of paranoia simulation. Using these binary labels, we fine-tuned a pre-trained publicly available BERT classifier (<https://github.com/huggingface/transformers#model-architectures>) to predict the label of any given chunk.

Only 25.7% of the input passes the BERT-based critic. The final pipeline that consists of the Generator subsystem, the heuristic filters, and the Critic subsystem produces the final results as such:

*a sudden feeling of austin lemons, a gentle stab of disgust. i'm what i'm humans whirl in night and distance.*  
*we shall never suffer this. if the human race came along tomorrow, none of us would be as wise as they already would have been. there is a beginning and an end.*  
*both of our grandparents and brothers are overdue. he either can not agree or he can look for someone to blame for his death.*  
*he has reappeared from the world of revenge, revenge, separation, hatred. he has ceased all who have offended him.*  
*and i don't want the truth. not for an hour.*

Table 1 compares generated texts on the different steps of the pipeline. Estimation of generative NLP models is generally a tedious task, yet Table 1 illustrates the properties of the text that the pipeline distills. The texts become more emotional in terms of polarity and more diverse in terms of the words used in them.

**Table 1.** Comparison of initial GPT-generated texts, heuristically filtered texts and texts after BERT filtration with 95% confidence intervals. Polarity and subjectivity are calculated by TextBlob library. Absolute polarity is averaged over the text samples of comparable length. Resulting texts become more and more emotional in terms of absolute sentiment. The resulting Paranoid Transformer texts have the highest average number of unique words per sentence and the highest variation in terms of the length of the resulting texts. The highest number in every column is highlighted with bold.

Model	Polarity	Subjectivity	Unique Words per Text Piece
Generated Sample	13.29% ± 0.8%	27.34% ± 1.8%	11.5 ± 0.7
Heuristically Filtered Sample	14.23% ± 0.9%	<b>32.26%</b> ± 2.1%	14.7 ± 0.9
BERT Filtered Sample	<b>15.65%</b> ± 0.9%	30.66% ± 1.3%	<b>15.6 ± 7.1</b>

The resulting generated texts were already thought-provoking and allowed reading a narrative of madness, but we wanted to enhance this experience and make it more immersive for the reader.

### 3.4. Nervous Handwriting

To enhance the personal aspect of the artificial paranoid author, we implemented an additional generative element. Using the implementation (<https://github.com/sjvasquez/handwriting-synthesis>) for handwriting synthesis from Graves [41], we generated handwritten versions of the generated texts. Assuming that more subjective and polarized texts should be written in a shakier hand-writing, we took the maximum of the TextBlob predictions for the absolute sentiment and subjectivity. For the text for which at least one of these parameters was exceeding 0.5, we set the bias parameter of the exponent to 0 for the shakiest handwriting. If either estimate was above 0.5 yet at least one of them exceeded 0, the bias was set to 0.5 for a steadier handwriting. Finally, if the text was estimated as neither a polarized nor subjective one, we set the bias parameter to 1 for the steadiest handwriting. Figures 1–3 show several final examples of the Paranoid Transformer diary entries.

*not badly primed in a controlled environment.  
 i can forget, i can't remember.  
 i can't remember.  
 no.  
 you may never know what i'm able to talk of.  
 i won't.  
  
 the day kills me.'*

**Figure 1.** Some examples of Paranoid Transformer diary entries. Three entries of varying length.

*perpetual silence hatched all around me.  
 i floated in silence, unable to understand, terror alert, every bit as nothing  
 i was half aware that i was dreaming. o  
 i had no tactile worries.*

**Figure 2.** Some examples of Paranoid Transformer diary entries. Longer entry proxying 'self-reflection' and personalized fringe mental state experience.

*i hope to see the disaster firsthand.  
 i hope that they will be better than what we have done before.  
  
 this doesn't feel like life anymore.  
 he's living in a country full of total strangers.*

**Figure 3.** Some examples of Paranoid Transformer diary entries. Typical entries with destructive and ostracized motives.

Figure 1 demonstrates that the length of the entries can differ from several consecutive sentences that convey a longer line of reasoning to a short, abrupt four-words note. Figure 2 illustrates typical entry of 'self-reflection'. The text explores the narrative of dream and could be paralleled with a description of an out-of-body experience [42] generated by the predominantly out-of-body entity. Figure 3 illustrates typical entries with destructive and ostracized motives. This is an exciting side-result of the model that we did not expect. The motive of loneliness is recurring in the Paranoid Transformer diaries.

It is important to emphasize that the resulting stream of the generated output is available online ([https://github.com/altsoph/paranoid\\_transformer](https://github.com/altsoph/paranoid_transformer)). No human post-processing of output is performed. The project won the NaNoGenMo 2019 (<https://nanogenmo.github.io/>) challenge. As a result, a book [43] is published. To our knowledge, this is the first book fully generated by AI without any human supervision. We regard this opinion of the publisher as a high subjective estimation of the resulting text quality.

A final touch was Random Sketcher an implementation of [44] trained on Quick, Draw! Dataset. Each time any of categories from the dataset appear on the page the Random Sketcher generates a picture somewhere around. Random circles hinting on the stains of a coffee cup suggest extra-linguistic signs to the reader and create an impression of a mindful work behind the text. All those extra-linguistic signs like handwriting and drawings build a particular writing subject with its own paranoid consciousness. The artistic hesitations and paranoiac mode of creative thinking became a central topic of many works of literature. Gogol's short story "Memoirs of a Madman" is probably one of the best classical examples of a "mad" or fringe-state narrative in modern literature that questions the

limits between creative potentiality and paranoid neurosis, which inspired the form of a Diary of a Madman generated by Paranoid Transformer, see [45]. Figure 4 shows photos of the hardcover book.



Figure 4. Hardcover version of the book by Paranoid Transformer.

#### 4. Discussion

In Dostoevsky’s “Notes from the Underground” [1], there is a striking idea about madness as a source of creativity and computational explanation as a killer of artistic magic: “We sometimes choose absolute nonsense because in our foolishness we see in that nonsense the easiest means for attaining a supposed advantage. However, when all that is explained and worked out on paper (which is perfectly possible, for it is contemptible and senseless to suppose that some laws of nature man will never understand), then certainly so-called desires will no longer exist” [46]. Paranoid Transformer brings forward an important question about the limitations of the computational approach of creative intelligence. This case demonstrates that creative potentiality and generation efficiency could be considerably influenced by such poorly controlled methods as obfuscated supervision and loose interpretation of the generated text.

Creative text generation studies inevitably strive to reveal fundamental cognitive structures that can explain the creative thinking of a human. The suggested framing approach to machine narrative as a narrative of madness brings forward some crucial questions about the nature of creativity and the research perspective on it. In this section, we discuss the notion of creativity that emerges from the results of our studying and reflect on the framing of the text generation algorithm.

What does creativity in terms of text generation mean? Is it a cognitive production of novelty or rather generation of unexpendable meaning? Can we identify any difference in treating human and machine creativity?

In his groundbreaking work, Turing [47] pinpointed several crucial aspects of intelligence. He stated: “If the meaning of the words “machine” and “think” are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer



to the question, “Can machines think?” is to be sought in a statistical survey such as a Gallup poll.” This starting argument turned out to be prophetic. It pinpoints the profound challenge for the generative models that use statistical learning principles. Indeed, if creativity is something on the fringe, on the tails of the distribution of outcomes, then it is hard to expect a model that is fitted on the center of distribution to behave in a way that could be subjectively perceived as a creative one. Paranoid Transformer is a result of a conscious attempt to push the model towards a fringe state of proximal madness. This case study serves as a clear illustration that creativity is ontologically opposed to the results of the “Gallup poll.”

Another question that raises discussion around computational creativity deals with a highly speculative notion of self within a generative algorithm. Does a mechanical writer have a notion of self-expression? Considering a wide range of theories of the self (carefully summarized in [48]), a creative AI generator triggers a new philosophical perspective on this question. As any human self, an artificial self does not develop independently. By following John Locke’s understanding of self as based on memory [49], Paranoid Transformer builds itself on memorizing the interactive experience with a human, furthermore, it emotionally inherits to its supervising readers who labeled the training dataset of the supervision system. On the other hand, Figure 5 clearly shows the impact of crypto-anarchic philosophy on the Paranoid Transformers’ notion of self. One can easily interpret the paranoid utterance of the generator as a doubt about reading and processing unbiased literature.

*copyrighted protein fiction may be deemed speculative propaganda.*

**Figure 5.** “Copyrighted protein fiction may be deemed speculative propaganda”—the authors are tempted to proclaim this diary entry the motto of Paranoid Transformer.

According to the cognitive science approach, the construction of self could be revealed in narratives about particular aspects of self [38]. In the case of Paranoid Transformer, both visual and verbal self-representation result in nervous and mad narratives that are further enhanced by the reader.

Regarding the problem of framing the study on creative text generators, we cannot avoid the question concerning the novelty of the generated results. Does Paranoid Transformer demonstrate a new result that is different from others in the context of computational creativity? First, we can use external validation. At the moment, the Paranoid Transformer’s book is prepared to come out of print. Secondly, and probably more importantly here, we can indicate the novelty of the conceptual framing of the study. Since the design and conceptual situatedness influence the novelty of study [50], we claim that the suggested conceptual extension of perceptive horizons of interaction with a generative algorithm can solely advocate the novelty of the result.

An important question that deals with the framing of the text generation results is the discussion of a chance discovery. In [31], the author laid out three crucial keys for chance discovery: communication, context shifting, and data mining. Abe [32] further enhanced these ideas addressing the issue of curation and claiming that curation is a form of communication. The Paranoid Transformer is a clear case study that is rooted in Ohsawa’s three aspects of chance discovery. Data mining is represented with a choice of data for fine-tuning and the process of fine-tuning itself. Communication is interpreted under Abe’s broader notion of curation as a form of communication. Context shift manifests itself through the reading the narrative of madness that invests the reader with interpretative freedom and motivates her to pursue the meaning in her mind through simple, immersive visualization of the systems’ fringe ‘mental state’.

## 5. Conclusions

This paper presents a case study of a Paranoid Transformer. It claims that framing the machine-generated narrative as a narrative of madness can intensify the personal experience of the reader. We explicitly address three critical aspects of chance discovery and claim that the resulting

system could be perceived as a digital persona in a fringe mental state. The crucial aspect of this perception is the reader, who is motivated to invest meaning into the resulting generative texts. This motivation is built upon several pillars: a challenging visual form, which focuses the reader on the text; obfuscation, which opens the resulting text to broader interpretations; and the implicit narrative of madness, which is achieved with the curation of the dataset for the fine-tuning of the model. Thus, we intersect the understanding of computational creativity with the fundamental ideas of the receptive theory.

**Author Contributions:** Conceptualization, Y.A., A.T. and I.P.Y.; data curation, A.T. and I.P.Y.; formal analysis, Y.A. and I.P.Y.; investigation, A.T. and I.P.Y.; methodology, Y.A., A.T. and I.P.Y.; project administration, I.P.Y.; Validation, A.T.; and writing—original draft, Y.A. and I.P.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Yamshchikov, I.P.; Shibaev, V.; Nagaev, A.; Jost, J.; Tikhonov, A. Decomposing Textual Information For Style Transfer. In Proceedings of the 3rd Workshop on Neural Generation and Translation, Hong Kong, China, 4 November 2019; pp. 128–137.
2. Rozin, P. Social psychology and science: Some lessons from Solomon Asch. *Personal. Soc. Psychol. Rev.* **2001**, *5*, 2–14. [CrossRef]
3. Yarkoni, T. The generalizability crisis. *PysArXiv* **2019**. Available online: <https://psyarxiv.com/jqw35/> (accessed on 26 October 2020).
4. Hickman, R. The art instinct: Beauty, pleasure, and human evolution. *Int. J. Art Des. Educ.* **2010**, *3*, 349–350. [CrossRef]
5. Melchionne, K. On the Old Saw “I know nothing about art but I know what I like”. *J. Aesthet. Art Crit.* **2010**, *68*, 131–141. [CrossRef]
6. Agafonova, Y.; Tikhonov, A.; Yamshchikov, I. Paranoid Transformer: Reading Narrative of Madness as Computational Approach to Creativity. In Proceedings of the International Conference on Computational Creativity, Coimbra, Portugal, 10 September 2020.
7. Veale, T. Read Me Like A Book: Lessons in Affective, Topical and Personalized Computational Creativity. In Proceedings of the 10th International Conference on Computational Creativity, Association for Computational Creativity, Charlotte, NC, USA, 17–21 June 2019; pp. 25–32. Available online: [https://computationalcreativity.net/iccc2019/assets/iccc\\_proceedings\\_2019.pdf](https://computationalcreativity.net/iccc2019/assets/iccc_proceedings_2019.pdf) (accessed on 26 October 2020).
8. van Stegeren, J.; Theune, M. Churnalist: Fictional Headline Generation for Context-appropriate Flavor Text. In Proceedings of the 10th International Conference on Computational Creativity, Association for Computational Creativity, Charlotte, NC, USA, 17–21 June 2019; pp. 65–72. Available online: [https://computationalcreativity.net/iccc2019/assets/iccc\\_proceedings\\_2019.pdf](https://computationalcreativity.net/iccc2019/assets/iccc_proceedings_2019.pdf) (accessed on 26 October 2020).
9. Alnajjar, K.; Leppänen, L.; Toivonen, H. No Time Like the Present: Methods for Generating Colourful and Factual Multilingual News Headlines. In Proceedings of the 10th International Conference on Computational Creativity, Charlotte, USA, NC, 17–21 June 2019; pp. 258–265. Available online: [https://computationalcreativity.net/iccc2019/assets/iccc\\_proceedings\\_2019.pdf](https://computationalcreativity.net/iccc2019/assets/iccc_proceedings_2019.pdf) (accessed on 26 October 2020).
10. Jhamtani, H.; Gangal, V.; Hovy, E.; Nyberg, E. Shakespearizing Modern Language Using Copy-Enriched Sequence-to-Sequence Models. In Proceedings of the Workshop on Stylistic Variation, Copenhagen, Denmark, 7–11 September 2017; pp. 10–19.
11. Tikhonov, A.; Yamshchikov, I. Sounds Wilde. Phonetically extended embeddings for author-stylized poetry generation. In Proceedings of the Fifteenth Workshop on Computational Research in Phonetics, Phonology, and Morphology, Brussels, Belgium, 31 October 2018; pp. 117–124.
12. Tikhonov, A.; Yamshchikov, I.P. Guess who? Multilingual approach for the automated generation of author-stylized poetry. In Proceedings of the 2018 IEEE Spoken Language Technology Workshop (SLT), Athens, Greece, 18–21 December 2018; pp. 787–794.
13. Wheatley, J. The Computer as Poet. *J. Math. Arts* **1965**, *72*, 105.

14. He, J.; Zhou, M.; Jiang, L. Generating Chinese Classical Poems with Statistical Machine Translation Models. In Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, Toronto, ON, Canada, 23–24 July 2012.
15. Yan, R.; Li, C.T.; Hu, X.; Zhang, M. Chinese Couplet Generation with Neural Network Structures. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, 7–12 August 2016; pp. 2347–2357.
16. Yi, X.; Li, R.; Sun, M. Generating Chinese classical poems with RNN encoder-decoder. In *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*; Springer: Cham, Switzerland, 2017; pp. 211–223.
17. Ghazvininejad, M.; Shi, X.; Choi, Y.; Knight, K. Generating Topical Poetry. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; pp. 1183–1191.
18. Yamshchikov, I.P.; Tikhonov, A. Learning Literary Style End-to-end with Artificial Neural Networks. *Adv. Sci. Technol. Eng. Syst. J.* **2019**, *4*, 115–125. [[CrossRef](#)]
19. Potash, P.; Romanov, A.; Rumshisky, A. GhostWriter: Using an LSTM for Automatic Rap Lyric Generation. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 17–21 September 2015; pp. 1919–1924.
20. Lamb, C.; Brown, D.G.; Clarke, C.L. A taxonomy of generative poetry techniques. *J. Math. Arts* **2017**, *11*, 159–179. [[CrossRef](#)]
21. Menabrea, L.F.; Lovelace, A. Sketch of the Analytical Engine Invented by Charles Babbage. 1842. Available online: <https://fourmilab.ch/babbage/sketch.html> (accessed on 26 October 2020).
22. Briot, J.P.; Hadjeres, G.; Pachet, F. *Deep Learning Techniques for Music Generation*; Springer: Cham, Switzerland, 2019; Volume 10.
23. Lebret, R.; Grangier, D.; Auli, M. Neural text generation from structured data with application to the biography domain. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; pp. 1203–1213.
24. Weizenbaum, J. ELIZA—A computer program for the study of natural language communication between man and machine. *Commun. ACM* **1966**, *9*, 36–45. [[CrossRef](#)]
25. Li, J.; Galley, M.; Brockett, C.; Spithourakis, G.P.; Gao, J.; Dolan, W.B. A Persona-Based Neural Conversation Model. *arXiv* **2016**, arXiv:1603.06155.
26. Burtsev, M.; Seliverstov, A.; Airapetyan, R.; Arkhipov, M.; Baymurzina, D.; Bushkov, N.; Gureenkova, O.; Khakhulin, T.; Kuratov, Y.; Kuznetsov, D.; et al. DeepPavlov: Open-source library for dialogue systems. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Melbourne, Australia, 15–20 July 2018; pp. 122–127.
27. Shiv, V.L.; Quirk, C.; Suri, A.; Gao, X.; Shahid, K.; Govindarajan, N.; Zhang, Y.; Gao, J.; Galley, M.; Brockett, C.; et al. Microsoft Icecaps: An Open-Source Toolkit for Conversation Modeling. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Florence, Italy, 28 July–2 August 2019; pp. 123–128.
28. Charnley, J.W.; Pease, A.; Colton, S. On the Notion of Framing in Computational Creativity. In Proceedings of the 3rd International Conference on Computational Creativity, Dublin, Ireland, 30 May–1 June 2012; pp. 77–81.
29. Cook, M.; Colton, S.; Pease, A.; Llano, M.T. Framing in computational creativity—A survey and taxonomy. In Proceedings of the 10th International Conference on Computational Creativity. Association for Computational Creativity, Charlotte, NC, USA, 17–21 June 2019; pp. 156–163. Available online: [https://computationalcreativity.net/iccc2019/assets/iccc\\_proceedings\\_2019.pdf](https://computationalcreativity.net/iccc2019/assets/iccc_proceedings_2019.pdf) (accessed on 26 October 2020).
30. Veale, T.; Cardoso, F.A. *Computational Creativity: The Philosophy and Engineering of Autonomously Creative Systems*; Springer: Cham, Switzerland, 2019.
31. Ohsawa, Y. Modeling the process of chance discovery. In *Chance Discovery*; Springer: Cham, Switzerland, 2003; pp. 2–15.
32. Abe, A. Curation and communication in chance discovery. In Proceedings of the 6th International Workshop on Chance Discovery (IWCD6) in IJCAI, Barcelona, Spain, 6–18 July 2011.
33. Gadamer, H.G. *Literature and Philosophy in Dialogue: Essays in German Literary Theory*; SUNY Press: New York, NY, USA, 1994.

34. Hirsch, E.D. *Validity in Interpretation*; Yale University Press: Chelsea, MA, USA, 1967; Volume 260.
35. Fish, S.E. *Is There a Text in This Class?: The Authority of Interpretive Communities*; Harvard University Press: Cambridge, MA, USA, 1980.
36. Eco, U. Towards a semiotic inquiry into the television message. In *Working Papers in Cultural Studies*; Translated by Paola Splendore; Blackwell Publishing: London, UK, 1972; Volume 3, pp. 103–121.
37. Whitehead, J. *Madness and the Romantic Poet: A Critical History*; Oxford University Press: Oxford, UK, 2017.
38. Dennett, D.C. The self as the center of narrative gravity. In *Self and Consciousness*; Psychology Press: Hove, UK, 2014; pp. 111–123.
39. Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I. Language models are unsupervised multitask learners. *OpenAI Blog* **2019**, *1*, 9.
40. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.
41. Graves, A. Generating sequences with recurrent neural networks. *arXiv* **2013**, arXiv:1308.0850.
42. Blanke, O.; Landis, T.; Spinelli, L.; Seeck, M. Out-of-body experience and autoscopia of neurological origin. *Brain* **2004**, *127*, 243–258. [[CrossRef](#)] [[PubMed](#)]
43. Tikhonov, A. *Paranoid Transformer*; NEW SIGHT: 2020. Available online: <https://deadlivemagazine.com/press/paranoid-transformer.html> (accessed on 26 October 2020).
44. Ha, D.; Eck, D. A neural representation of sketch drawings. *arXiv* **2017**, arXiv:1704.03477.
45. Bocharov, S.G. *O Hudozhestvennykh Mirakh*. Moscow, Russia, 1985; pp. 161–209. Available online: [https://imwerden.de/pdf/bocharov\\_o\\_khudozhestvennykh\\_mirakh\\_1985\\_text.pdf](https://imwerden.de/pdf/bocharov_o_khudozhestvennykh_mirakh_1985_text.pdf) (accessed on 26 October 2012).
46. Dostoevsky, F. Zapiski iz podpolya - Notes from Underground. *Povesti i rasskazy v 2 t* **1984**, *2*, 287–386.
47. Turing, A.M. Computing machinery and intelligence. *Mind* **1950**, *59*, 433. [[CrossRef](#)]
48. Janwal, V. Exploring the Notion of Self in Creative Self-Expression. In Proceedings of the 10th International Conference on Computational Creativity ICCCI9, Charlotte, NC, USA, 17–21 June 2019; pp. 331–335.
49. Locke, J. *An Essay Concerning Human Understanding: And a Treatise on the Conduct of the Understanding*; Hayes & Zell: Philadelphia, PA, USA, 1860.
50. Perišić, M.M.; Štorga, M.; Gero, J. Situated novelty in computational creativity studies. In Proceedings of the 10th International Conference on Computational Creativity ICCCI9, Charlotte, NC, USA, 17–21 June 2019; pp. 286–290.

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).