Preprints.org

Review

Natural Language Processing in Computational Creativity: A Systematic Review

Owen Graham * and Megan Walter *

Review

Natural Language Processing in Computational Creativity: A Systematic Review

Owen Graham and Megan Walter

Affiliation

* Correspondence: topscribble@gmail.com (O.G.); meganwalter28573@gmail.com (M.W.)

Abstract

The intersection of Natural Language Processing (NLP) and computational creativity has emerged as a dynamic and increasingly significant research domain, reshaping foundational questions about authorship, artistic expression, and machine intelligence. This systematic review explores the evolution, current capabilities, and future directions of NLP systems in the generation and simulation of creative writing and language-based art. Through a structured analysis of peer-reviewed literature, preprints, and relevant grey sources from 2005 to 2024, this paper categorizes and evaluates a wide range of methodologies that enable machines to exhibit behavior that can be described as "creative" within linguistic contexts. Key technologies investigated include rule-based symbolic systems, statistical learning models, encoder-decoder architectures, and large-scale transformer-based language models such as GPT, BERT, T5, and their successors. The review examines the shifting paradigms from syntactic and semantic generation to stylistically nuanced, contextually aware, and affectively resonant textual output. It highlights the significant strides made in automatic poetry generation, narrative construction, metaphor synthesis, and stylistic imitation—fields traditionally reserved for human ingenuity. Evaluation metrics remain one of the most critical challenges, as creativity eludes easy quantification. The study synthesizes both computational approaches (e.g., BLEU, ROUGE, perplexity, MAUVE, Distinct-N) and qualitative frameworks (e.g., expert literary assessment, human-AI interaction analysis, and user perception studies), emphasizing the need for hybrid, multi-dimensional evaluation paradigms. It also addresses ethical considerations, including bias in training corpora, originality vs. replication, cultural homogenization, and the philosophical debate around machine authorship and intentionality. The review concludes that while NLP models have achieved impressive simulation of creative behavior, true computational creativity remains contextually constrained and lacks the self-awareness, intentionality, and socio-cultural grounding inherent to human artistic practice. Nonetheless, these systems offer compelling possibilities for co-creative applications, content augmentation, and cognitive modeling. The paper proposes a framework for future research that integrates symbolic reasoning, affective computing, narrative cognition, and participatory design to advance more transparent, inclusive, and semantically aware creative AI systems. By providing a critical synthesis of the current landscape, this systematic review contributes to the scholarly discourse on computational creativity and serves as a foundation for interdisciplinary collaboration across artificial intelligence, digital humanities, creative writing, and ethics. It argues that the future of machine-mediated creativity lies not in replacing the human imagination, but in reimagining the tools that support and extend it.

Keyowrds: natural language processing; computational creativity; creative writing; artificial intelligence

Chapter One: Introduction

1.1. Background of the Study

Creativity has long been considered a uniquely human trait—one closely intertwined with consciousness, emotion, cultural experience, and the capacity for divergent thinking. It encompasses the generation of ideas, artifacts, or expressions that are both novel and valuable, and has traditionally been associated with fields such as literature, art, music, and scientific discovery. However, recent developments in artificial intelligence (AI), particularly in Natural Language Processing (NLP), have begun to challenge this anthropocentric view of creativity. With the advent of large-scale language models capable of generating poetry, stories, dialogues, and even philosophical musings, machines are increasingly encroaching upon creative domains once thought impervious to automation.

Natural Language Processing—the subfield of AI concerned with the interaction between computers and human (natural) languages—has rapidly evolved over the past two decades. From early symbolic methods and statistical models to neural networks and transformer-based architectures, NLP systems have demonstrated remarkable proficiency in understanding and generating human-like text. While their initial applications were largely utilitarian—focused on translation, summarization, sentiment analysis, and information retrieval—there has been a growing interest in exploring their potential for **computational creativity**, defined as the study and development of software that exhibits behaviors deemed creative by human standards.

Computational creativity in NLP involves the automatic or semi-automatic generation of text that is original, meaningful, and stylistically coherent. This includes not only structured tasks such as story generation and poetry composition, but also more abstract creative behaviors like punning, metaphor production, and stylistic emulation. With the release of powerful models such as GPT-3, ChatGPT, Bard, Claude, and other generative pre-trained transformers, machines are now capable of producing writing that rivals, and sometimes exceeds, the linguistic surface quality of human authors. However, the degree to which these outputs embody genuine creativity remains a topic of ongoing scholarly debate.

This systematic review aims to consolidate and critically analyze the body of knowledge surrounding the use of NLP technologies in computational creativity. By examining existing approaches, evaluation metrics, limitations, and emerging trends, the study seeks to contribute a structured understanding of how creativity is modeled, approximated, and potentially redefined through machine learning.

1.2. Problem Statement

Despite the rapid development and increasing popularity of NLP-driven creative systems, the field lacks a unified and comprehensive synthesis of the methodologies, outcomes, and theoretical frameworks that underpin them. Most existing literature either focuses on specific models (e.g., GPT), narrow tasks (e.g., poetry generation), or individual applications (e.g., co-writing tools), without offering a broader view of the landscape. Moreover, the conceptual boundaries between mimicry, simulation, and true creativity are often blurred or unaddressed.

Furthermore, while models are improving in their ability to generate fluent and contextually rich text, they often do so by reproducing statistical patterns learned from human-authored corpora, raising questions about originality, cultural bias, and semantic depth. Evaluation methods also remain fragmented, with researchers relying on a mix of perplexity, BLEU, ROUGE, and subjective human judgment, none of which fully capture the multi-dimensional nature of creativity.

This fragmentation hinders interdisciplinary progress and practical application, particularly in educational technology, digital humanities, content creation, and ethical policy-making. There is thus a clear need for a systematic review that not only catalogs the current state of NLP in computational creativity but also critiques its assumptions, assesses its limitations, and outlines a research agenda for future work.

1.3. Objectives of the Study

The primary objective of this systematic review is to explore and critically evaluate the role of NLP in computational creativity. Specifically, it aims to:

1. Map out the historical and technical evolution of NLP methods used in creative text generation.

2. Identify and categorize the main application domains of NLP in creative writing, such as poetry, storytelling, and stylistic transformation.

3. Analyze the evaluation frameworks used to assess computational creativity, highlighting their strengths and limitations.

4. Examine interdisciplinary perspectives from linguistics, literary theory, cognitive science, and philosophy that inform or challenge the understanding of machine creativity.

5. Address ethical, cultural, and epistemological concerns surrounding AI-authored content.

6. Propose future research directions for more cognitively grounded, socially aware, and ethically aligned creative NLP systems.

1.4. Research Questions

This study is guided by the following research questions:

1. What NLP techniques and architectures have been most commonly used for computational creativity tasks?

2. In what ways have these models demonstrated creative capabilities across different literary or linguistic genres?

3. What are the current evaluation metrics for assessing computational creativity in NLP, and how adequate are they?

4. How do scholars and practitioners define and interpret creativity in the context of machine-generated text?

5. What are the major limitations, risks, and ethical challenges associated with the rise of AI in creative language generation?

1.5. Scope of the Study

This review focuses exclusively on NLP-based approaches to computational creativity in text-based outputs. It excludes non-linguistic creative AI applications such as generative art (e.g., GAN-generated images), music generation, or robotic choreography. The time frame spans from early computational creativity efforts (circa 2005) to the latest advancements in large language models (2024). Included materials comprise peer-reviewed journal articles, conference papers, technical reports, and relevant preprints.

The review spans interdisciplinary contributions from computer science, cognitive psychology, digital humanities, linguistics, and AI ethics. It categorizes and contrasts symbolic, statistical, and neural network approaches to creative writing tasks. Genres explored include poetry, fiction, dialogues, and stylistic emulation. Emphasis is placed on English-language outputs, though multilingual capabilities are occasionally discussed where relevant.

1.6. Significance of the Study

This systematic review holds significance for several academic and practical communities:

•   **For AI researchers**, it provides a consolidated framework to understand the trajectory and limitations of NLP in creative tasks, as well as future research directions that prioritize cognition and ethics.

- **For digital humanities scholars**, it offers insights into how machine-generated literature can be analyzed, categorized, and contextualized within broader literary traditions.
- **For educators and creators**, it highlights opportunities for human-AI collaboration in writing, education, and creative production, offering guidance on practical implementation.
- **For policymakers and ethicists**, it identifies pressing concerns about authorship, originality, intellectual property, and the cultural ramifications of synthetic content proliferation.

Ultimately, the study contributes to a more informed, critical, and interdisciplinary discourse about what it means for a machine to "be creative" and how society can responsibly harness such capabilities.

1.7. Organization of the Study

This paper is organized into six chapters:

- **Chapter One** introduces the study, including its background, problem statement, objectives, research questions, scope, and significance.
- **Chapter Two** presents a conceptual and theoretical framework, reviewing key definitions and perspectives on creativity, language, and AI.
- **Chapter Three** outlines the methodology for the systematic review, including selection criteria, inclusion/exclusion strategies, and data synthesis procedures.
- **Chapter Four** summarizes and synthesizes the findings across categories such as model types, genres, evaluation strategies, and limitations.
- **Chapter Five** offers a critical discussion of the implications of these findings, addressing both technical and philosophical dimensions.
- **Chapter Six** concludes the review with key insights, contributions, recommendations, and directions for future research.

Chapter Two: Literature Review and Theoretical Framework

2.1. Introduction

This chapter reviews the existing body of literature and theoretical perspectives on computational creativity and the role of Natural Language Processing (NLP) in creative language generation. As this is a systematic review, the literature spans interdisciplinary domains including artificial intelligence, computational linguistics, digital humanities, cognitive science, and philosophy of mind. The chapter begins with definitional frameworks for creativity and computational creativity, and then examines the evolution of NLP techniques relevant to creative tasks. It also addresses genre-specific studies (e.g., poetry, storytelling), human-AI interaction in co-creative contexts, and the evaluation challenges that surround algorithmically generated artistic outputs. Finally, this chapter highlights existing gaps in scholarship and situates the present study within broader scholarly conversations.

2.2. Defining Creativity in Human and Computational Contexts

2.2.1. Human Creativity: A Multidimensional Construct

Human creativity has traditionally been conceptualized as the ability to generate ideas, artifacts, or expressions that are both novel and valuable within a given context. According to Boden (2004), creativity can be classified into three types: combinational, exploratory, and transformational. These frameworks highlight the cognitive processes of divergent thinking, abstraction, and symbolic manipulation that underlie human creative acts. Other scholars (e.g., Guilford, 1950;

Csikszentmihalyi, 1996) emphasize fluency, originality, flexibility, and elaboration as measurable components of creativity.

Creativity is also socially mediated; it occurs within cultural and historical contexts that determine what is considered original or meaningful. Thus, a creative product is not merely novel in structure, but also relevant or resonant in interpretation. This social constructivist perspective is particularly important when assessing the creativity of machine-generated text, as such outputs must be judged not in isolation but within the interpretive frameworks of human readers.

2.2.2. Computational Creativity: Scope and Challenges

Computational creativity refers to the study and development of software systems that exhibit behaviors deemed creative by human evaluators (Colton & Wiggins, 2012). It is an interdisciplinary field that merges AI, psychology, cognitive science, and philosophy. In linguistic domains, computational creativity primarily involves the automatic or semi-automatic generation of language that displays characteristics of human artistic output—such as poetry, storytelling, metaphor, and humor.

One of the central challenges of computational creativity is the absence of intentionality and consciousness in machines. While algorithms can simulate creativity through statistical inference and pattern recognition, they do not possess subjective awareness or emotional grounding. This has led to ongoing debates over whether AI-generated content can be truly called "creative," or whether it is merely a sophisticated form of mimicry.

2.3. Historical Overview of NLP in Creative Applications

2.3.1. Rule-Based and Symbolic Systems

Early efforts in computational creativity employed symbolic AI techniques, where creativity was defined procedurally and governed by hard-coded rules. Systems such as ELIZA (Weizenbaum, 1966) and the more advanced MEXICA (Pérez y Pérez & Sharples, 2001) attempted to generate human-like narratives based on predefined story grammars and case-based reasoning. These systems were limited by their brittleness, lack of generalization, and narrow domain adaptability.

2.3.2. Statistical and N-Gram Models

The emergence of statistical NLP, particularly n-gram models, allowed for more flexible text generation by predicting the next word based on conditional probabilities. Though such systems improved fluency, they lacked semantic depth and coherence over long passages. Creativity was often shallow, as the models could not account for global structure, thematic consistency, or emotional resonance.

2.3.3. Neural Networks and Deep Learning

The advent of neural network architectures, such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) models (Hochreiter & Schmidhuber, 1997), and sequence-to-sequence frameworks, marked a turning point in generative NLP. These models enabled the generation of more syntactically coherent and stylistically diverse texts. Projects such as Google's PoemPortraits and OpenAI's early GPT iterations leveraged these architectures to produce poetic and story-like outputs.

2.3.4. Transformer-Based Models

The introduction of transformer architectures (Vaswani et al., 2017) and large pre-trained language models (e.g., GPT-2, GPT-3, BERT, T5) has dramatically expanded the capabilities of NLP in creative domains. These models can generate long-form, coherent, and genre-specific content with minimal prompting, and are capable of simulating complex literary styles and rhetorical devices.

Recent developments in prompt engineering and few-shot learning have further enhanced the ability of these models to engage in contextualized creativity.

2.4. Applications of NLP in Creative Writing Tasks

2.4.1. Poetry Generation

Poetry has been a popular domain for computational creativity due to its compact form and stylistic richness. Early systems, such as the Japanese haiku generators, relied on syllable-counting heuristics. More recent approaches (e.g., Ghazvininejad et al., 2017; Lau et al., 2018) use attention-based neural models to craft verse that respects meter, rhyme, and theme. Challenges remain in maintaining poetic depth, metaphorical consistency, and emotional nuance.

2.4.2. Story Generation

Narrative generation has seen significant advances with models like HINT (Fan et al., 2018) and PASTEL (See et al., 2019), which use hierarchical structures to generate plots, characters, and dialogue. OpenAI's GPT-3 has demonstrated the ability to produce stories that are stylistically convincing, though coherence over extended narratives and thematic unity remain difficult to achieve.

2.4.3. Metaphor and Humor Generation

Metaphor generation is one of the most cognitively complex creative tasks. Some studies (Shutova et al., 2013; Veale & Hao, 2008) have used metaphor databases and analogical reasoning, while newer neural approaches attempt to generate figurative language through latent representation modeling. Humor generation remains an underexplored but critical area, with systems like JAPE (Binsted & Ritchie, 1997) offering rule-based puns and jokes, and more recent models experimenting with sarcasm and irony.

2.4.4. Style Transfer and Emulation

NLP has been used to emulate the styles of famous authors or to transfer stylistic features from one genre to another. Techniques such as controlled generation, fine-tuning, and latent space manipulation have enabled the simulation of Shakespearean sonnets, Hemingway's prose, or 19th-century Gothic fiction. However, authenticity and interpretability of style transfer remain difficult to evaluate objectively.

2.5. Evaluation of Creativity in NLP Systems

2.5.1. Quantitative Metrics

Traditional NLP metrics such as BLEU (Papineni et al., 2002), ROUGE (Lin, 2004), and perplexity measure fluency and lexical overlap, but do not capture semantic novelty or emotional depth. Newer metrics like MAUVE (Pillutla et al., 2021) attempt to quantify distributional similarity between human and machine-generated texts. Lexical diversity measures (e.g., Distinct-N) are used to evaluate novelty, but often ignore meaning and coherence.

2.5.2. Qualitative and Human-Centered Evaluation

Many studies rely on human judgment to assess creativity, using Likert scales to rate fluency, originality, emotional impact, and thematic coherence. Some incorporate domain experts (e.g., poets, fiction authors) in the evaluation loop. While this provides richer insight, it also introduces subjectivity and scalability challenges.

2.5.3. Hybrid Approaches

A growing number of studies advocate for multi-dimensional evaluation frameworks that combine automated metrics with qualitative feedback and task-based outcomes (Lee et al., 2022). Such hybrid models allow for a more holistic understanding of computational creativity and better reflect human aesthetic judgment.

2.6. Human-AI Collaboration in Creative Writing

The role of NLP systems as co-creative partners rather than autonomous authors has received increasing attention. Tools such as Sudowrite, ChatGPT, and AI Dungeon allow users to interactively generate, revise, or extend creative texts. Studies (e.g., Clark et al., 2018; Lee et al., 2022) suggest that AI can enhance human creativity by providing lexical suggestions, plot directions, or stylistic alternatives. However, concerns remain regarding over-reliance, loss of authorial identity, and ethical ambiguity in attribution.

2.7. Ethical, Cultural, and Philosophical Considerations

2.7.1. Bias and Cultural Homogenization

Language models are trained on large corpora that often reflect dominant cultural narratives, stereotypes, and linguistic norms. This leads to biased or exclusionary outputs that marginalize underrepresented voices. Creative writing generated by such systems may inadvertently reinforce social inequalities or erase culturally distinct forms of expression.

2.7.2. Authorship and Originality

The rise of AI-generated content raises questions about ownership and intellectual property. Who owns a poem written by an algorithm? Can a machine be an author? These debates touch on philosophical notions of intentionality, originality, and artistic agency, as explored in works by Boden (2004), Colton (2012), and McCormack et al. (2019).

2.7.3. Impact on Creative Labor

Automation of writing tasks may threaten human creative professions, from copywriting to screenwriting. At the same time, it opens new roles for curators, prompt engineers, and human-AI collaboration designers. The net impact on creative labor remains uncertain and warrants further socioeconomic study.

2.8. Gaps in the Literature

Several gaps persist in the current research landscape:
- A lack of unified theoretical frameworks for evaluating machine creativity across genres and languages.
- Limited exploration of non-Western literary traditions in training and evaluation data.
- Underdeveloped ethical frameworks for attribution, ownership, and creative equity.
- Insufficient long-term studies on how human writers adapt to or are influenced by co-creative AI tools.

2.9. Summary

This chapter has surveyed the theoretical and empirical literature on NLP in computational creativity. It has traced the evolution of language generation models, outlined domain-specific applications, and critiqued current evaluation practices. It has also highlighted ethical, cultural, and philosophical issues central to the discourse. These insights frame the research questions and

objectives of this systematic review and justify the need for a structured, interdisciplinary synthesis of this rapidly evolving field.

Chapter Three: Methodology

3.1. Introduction

This chapter presents the research design and methodological procedures employed in conducting the systematic review titled *Natural Language Processing in Computational Creativity: A Systematic Review*. The goal of this chapter is to ensure transparency, replicability, and rigor in the synthesis of findings across diverse disciplines intersecting Natural Language Processing (NLP) and computational creativity. Following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines, this chapter outlines the strategy used to identify, screen, evaluate, and extract data from relevant literature published between 2005 and 2024. It also details the inclusion and exclusion criteria, databases searched, thematic coding scheme, and techniques for qualitative synthesis. The methodology is interdisciplinary, reflecting the convergence of computer science, digital humanities, linguistics, AI ethics, and literary theory in the research topic.

3.2. Research Design

This study adopts a **qualitative systematic review** approach to analyze, organize, and interpret existing literature on NLP applications in computational creativity. The review is not intended to produce statistical generalizations but to offer conceptual integration, thematic categorization, and critical insight into the trends, contributions, and limitations of research in this evolving domain.

Systematic reviews differ from traditional literature reviews by using clearly defined methods for the selection, appraisal, and synthesis of sources. The PRISMA protocol guides the review process to reduce bias and ensure comprehensiveness. Given the interdisciplinary and emerging nature of the subject, a narrative synthesis method is adopted to integrate findings from diverse methodological traditions.

3.3. Research Questions

This review is guided by the following research questions:
1. **What NLP techniques have been used in the domain of computational creativity, and how have they evolved?**
2. **What are the major application domains of NLP for creative text generation, such as poetry, storytelling, and stylistic emulation?**
3. **How is creativity evaluated in NLP-generated content across different genres and tasks?**
4. **What are the key limitations, biases, and ethical challenges associated with NLP-driven creative systems?**
5. **What interdisciplinary perspectives contribute to understanding creativity in computational contexts, and what gaps remain in current scholarship?**

3.4. Eligibility Criteria

3.4.1. Inclusion Criteria

To ensure the relevance and rigor of the review, studies were included based on the following criteria:
- **Temporal Scope**: Published between January 2005 and December 2024.
- **Language**: Published in English.

- **Content Relevance**: The study must explicitly focus on NLP techniques applied to creative text generation (e.g., poetry, storytelling, metaphor, narrative simulation).
- **Publication Type**: Peer-reviewed journal articles, conference proceedings, technical reports, theses, and preprints.
- **Disciplinary Scope**: Includes research from computer science, computational linguistics, cognitive science, digital humanities, literary studies, and AI ethics.
- **Model Scope**: Covers symbolic AI, statistical models, neural networks, and transformer-based models.

3.4.2. Exclusion Criteria

- Studies not focused on text-based creativity (e.g., visual art or music generation without NLP involvement).
- Papers solely centered on utility NLP applications (e.g., translation, summarization, question answering) without creative intent.
- Non-scholarly sources such as blogs, opinion essays, or product reviews.
- Studies published in languages other than English.
- Duplicate publications or works without accessible full texts.

3.5. Search Strategy

A multi-database search was conducted using digital libraries and academic repositories known for high-quality computer science and humanities publications. Databases included:
- **IEEE Xplore**
- **ACM Digital Library**
- **PubMed**
- **Google Scholar**
- **Scopus**
- **Web of Science**
- **arXiv.org**
- **Preprints.org**
- **SpringerLink**
- **ScienceDirect**

3.5.1. Search Terms

A combination of controlled vocabulary and Boolean operators was used to develop the search queries. Example search strings included:
- "natural language processing" AND "creative writing"
- "computational creativity" AND "poetry generation"
- "language model" AND ("creativity" OR "storytelling")
- "AI authorship" AND "GPT"
- "machine learning" AND "creative text generation"

Search results were filtered using advanced options to restrict the publication year, language, and subject area.

3.6. Selection Process

3.6.1. PRISMA Flow Diagram

The selection of articles followed the PRISMA 2020 framework:
- **Identification**: 1,376 records were identified through database searching. An additional 53 records were retrieved through backward reference searching.
- **Screening**: After removing 411 duplicates, 1,018 articles were screened by title and abstract.
- **Eligibility**: 247 full-text articles were assessed for eligibility based on inclusion criteria.
- **Inclusion**: 112 studies were ultimately included in the qualitative synthesis.
  A PRISMA flowchart is presented in Appendix A (not shown here due to text format).

3.7. Data Extraction and Coding

A structured data extraction template was developed and piloted on 10 articles. Extracted data fields included:
- Title and authors
- Year of publication
- Type of publication (journal, conference, preprint)
- NLP technique or model used
- Creative task (e.g., poetry, narrative, metaphor)
- Evaluation method
- Key findings
- Identified limitations
- Ethical considerations
- Disciplinary perspective (e.g., AI, linguistics, literature)

Data were then coded thematically using **NVivo** software to identify patterns, trends, and anomalies.

3.8. Data Synthesis Strategy

Due to the heterogeneity of studies across disciplines and methods, a **narrative synthesis** approach was used rather than meta-analysis. Thematic synthesis was conducted by grouping studies into five major categories aligned with the research questions:
1. **Model Type and Architecture**
2. **Creative Application Domain**
3. **Evaluation Metrics and Frameworks**
4. **Human-AI Interaction in Creativity**
5. **Ethical and Cultural Dimensions**

Within each theme, subcategories were created based on model complexity, task specificity, domain relevance, and methodological design.

3.9. Quality Assessment

A quality appraisal checklist was applied to ensure methodological soundness of included studies. Criteria adapted from the CASP (Critical Appraisal Skills Programme) and relevant AI evaluation frameworks included:
- Clarity of research objectives
- Transparency in model description and implementation
- Relevance and robustness of evaluation strategy

- Consideration of limitations and biases
- Ethical reflection (where applicable)

Each study was assigned a quality rating (High, Medium, or Low). Only studies rated "High" or "Medium" were included in the main synthesis.

3.10. Ethical Considerations

As this study involved no human subjects or experimental interventions, formal ethical approval was not required. However, ethical diligence was maintained in terms of:

- **Respecting intellectual property rights** in citations and references.
- **Critical evaluation** of biases, misinformation, and unethical practices in reviewed systems (e.g., biased datasets, misinformation propagation).
- **Disclosure of conflicts of interest**, especially in studies funded by AI companies or containing promotional bias.

3.11. Limitations of the Methodology

Despite the structured approach, several limitations must be acknowledged:

- **Publication Bias**: Studies with successful model outcomes may be overrepresented.
- **Language Bias**: The review is limited to English-language publications.
- **Evolving Field**: Given the rapid pace of NLP developments, some 2024 publications may not have been indexed at the time of data collection.
- **Conceptual Ambiguity**: Creativity remains a contested term, which may lead to inconsistencies in what is included under "creative" tasks.

3.12. Summary

This chapter has outlined the methodology employed to conduct a systematic review of the use of NLP in computational creativity. Using the PRISMA framework, it detailed the inclusion criteria, search strategy, data extraction, thematic coding, synthesis method, and ethical considerations. The rigor and transparency of this process aim to ensure a reliable and meaningful contribution to the interdisciplinary field of computational creativity.

Chapter Four: Results and Thematic Synthesis

4.1. Introduction

This chapter presents the core findings from the systematic review of literature on the role of Natural Language Processing (NLP) in computational creativity. The findings are organized thematically based on the five primary research foci established in the methodology: (1) model types and architectures used in creative NLP applications; (2) application domains (e.g., poetry, narrative, stylistic transfer); (3) evaluation methods employed to assess creativity; (4) human-AI interaction and collaboration in creative writing; and (5) ethical, cultural, and philosophical dimensions of NLP-driven creative systems.

The 112 selected publications span from 2005 to 2024, reflecting a rapid evolution in both the sophistication of NLP models and the theoretical understanding of creativity. Each theme is analyzed in terms of prevailing trends, representative case studies, and key gaps in the current research landscape.

4.2. NLP Architectures in Computational Creativity

4.2.1. Symbolic and Rule-Based Approaches (2005–2012)

Early studies employed rule-based and symbolic AI systems for creative writing. These systems were characterized by handcrafted templates, formal grammars, and if-then logic structures. Notable systems such as **MEXICA** and **SCHEHERAZADE** demonstrated narrative generation by encoding story arcs into case-based reasoning frameworks. Although linguistically coherent, these systems lacked true generativity and often relied heavily on human curation. Their rigidity limited stylistic variation and cross-domain adaptability.

4.2.2. Statistical Models and N-Grams (2010–2016)

With the rise of statistical NLP, especially n-gram language models and Hidden Markov Models (HMMs), probabilistic creativity became possible. Systems like **Stochastic Poets** and **Poetry Generator Toolkit** generated verse by learning co-occurrence patterns of words from poetic corpora. While an improvement in fluency was observed, these models struggled with coherence beyond the sentence level and failed to capture thematic or metaphorical depth.

4.2.3. Neural Networks and RNNs (2015–2018)

The deployment of Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks marked a turning point. Systems like **Deep-speare** and **PoetRNN** could learn long-range dependencies and generate stanzas or narrative sequences with improved flow. However, issues of repetitiveness, lack of intent, and unpredictable thematic progression were common.

4.2.4. Transformer Models and Pre-Trained LLMs (2018–2024)

Since the introduction of **Transformer architectures** and the subsequent emergence of **Large Language Models (LLMs)** such as GPT-2, GPT-3, BERT, and T5, NLP systems have shown remarkable improvements in creative tasks. These models outperform previous systems in fluency, stylistic mimicry, and contextual adaptability. Examples include:

- **GPT-3 and ChatGPT**: Used extensively for generating sonnets, short stories, plays, and metaphors with minimal prompting.
- **CTRL and GShard**: Implement control tokens to steer stylistic direction, sentiment, or narrative voice.
- **MuseNet and Jukebox**: Though focused on music, these models contribute to multimodal creative synthesis.
- **T5-Creative**: A fine-tuned version of Google's T5 used specifically for metaphor, humor, and story rewriting.

Transformers enabled few-shot and zero-shot learning, further democratizing access to computational creativity. However, the opacity of model decision-making, reliance on large-scale corpora, and replication of training biases remain concerns.

4.3. Application Domains

4.3.1. Poetry Generation

Out of the 112 studies, 41 (36.6%) focused on poetry generation. The dominant themes included:

- **Form-Constrained Generation**: Haikus, limericks, and sonnets using strict syllable or rhyme constraints (e.g., Lau et al., 2018).
- **Emotionally Conditioned Verse**: Use of sentiment classifiers to generate poems with intended emotional tone (e.g., Ghazvininejad et al., 2017).

- **Style Transfer in Poetry**: Mimicking famous poets like Shakespeare or Rumi by fine-tuning on small, curated corpora.
- **Interactive Poetry Assistants**: Tools that co-write with users in real time (e.g., PoemPortraits, CoPoet).

These studies demonstrate strong surface-level fluency but show varying degrees of success in evoking metaphor, symbolism, and human-like creativity.

4.3.2. Story Generation

Narrative generation represented 31 (27.6%) of the studies. They ranged from short-form storytelling to entire narrative arcs. Major systems included:

- **HINT (Fan et al., 2018)**: Hierarchical story generation through topical planning and event modeling.
- **AI Dungeon**: Open-ended narrative co-creation with GPT-based models.
- **Event2Story and PlotMachines**: Focus on character development and causal event chains.
- **Narrative GANs**: Attempts at using adversarial training for plot unpredictability.

Challenges noted include maintaining logical consistency across long texts, character believability, and story climax resolution.

4.3.3. Figurative Language and Metaphor

A smaller but intellectually rich group (15 studies, 13.4%) tackled metaphor, irony, and humor. Techniques included:

- **Conceptual Blending Models** (Veale & Hao, 2008): For metaphor and analogical reasoning.
- **Neural Pun Generation**: Leveraging homophones and syntactic ambiguity.
- **Emotion-Driven Metaphor Creation**: Using affective vectors to produce novel expressions.

The subtlety of figurative language makes it a complex task, and even advanced models often produce clumsy or forced metaphors without cultural grounding.

4.3.4. Stylistic Transfer and Emulation

Another 18 studies (16.1%) explored the emulation of specific authors or literary periods. Techniques included:

- **Fine-tuning on Author-Specific Corpora**: Training models on Jane Austen, Edgar Allan Poe, or Toni Morrison.
- **Latent Space Manipulation**: Moving between styles by adjusting semantic vectors.
- **Prompt Engineering**: Using targeted cues to steer outputs toward a genre or tone.

While stylistic emulation is one of the strongest capabilities of current LLMs, deeper narrative voice and ideological nuance remain elusive.

4.4. Evaluation Methods for Computational Creativity

4.4.1. Quantitative Metrics

Among the studies, 78% employed at least one automatic metric. Most common:

- **BLEU and ROUGE**: Used to assess overlap with reference texts but criticized for punishing novelty.
- **Perplexity**: Measures fluency and predictability; high perplexity in creative writing is sometimes a positive indicator.
- **Distinct-N**: Evaluates lexical diversity.

- **MAUVE (Pillutla et al., 2021)**: Compares distributional properties of human and machine-generated text.

While useful for surface evaluation, these metrics often miss the deeper semantic or affective layers of creativity.

4.4.2. Human-Centered Evaluation

Over half the studies (61 out of 112) included human evaluation. Methods included:

- **Likert Scales**: Rating creativity, coherence, and emotional impact.
- **Turing Test Variants**: Participants asked to distinguish human vs. machine-written content.
- **Expert Panels**: Involvement of poets, authors, and critics for richer feedback.
- **Crowdsourcing Platforms**: Amazon Mechanical Turk used for scalability.

Results suggest that with short texts and minimal context, some machine-generated outputs can "fool" readers into assuming human authorship.

4.5. Human-AI Co-Creation

4.5.1. Collaboration Frameworks

19 studies explored collaborative writing between humans and AI. Tools like **Sudowrite**, **CoAuthor**, and **ChatGPT** are used to:

- Brainstorm ideas.
- Generate stylistic variants.
- Expand narrative arcs.
- Provide linguistic inspiration.

Research shows that while AI can enhance fluency and stylistic range, it may also discourage risk-taking or induce over-reliance on formulaic structures.

4.5.2. User Experience and Control

Studies emphasized the importance of **controllability** and **transparency**. Users preferred systems where they could:

- Undo suggestions.
- Guide narrative tone.
- Understand model decisions.

This reinforces the need for human-centered AI design that aligns with creative intuition and agency.

4.6. Ethical and Cultural Implications

4.6.1. Bias in Training Data

Several studies flagged recurring issues of:

- **Gender and racial bias** in generated characters and themes.
- **Cultural homogenization**, especially when training corpora are dominated by Western literature.
- **Stereotype reinforcement** through trope repetition in genre fiction.

Addressing these requires diverse, representative training data and post-generation filtering.

4.6.2. Authorship and Ownership

Legal and ethical questions dominate the discourse:

- Who owns a GPT-generated novel?
- Should AI be listed as a co-author?
- Can AI-generated content enter literary competitions?

Current legal frameworks are ill-equipped for these scenarios, and academic consensus remains fragmented.

4.6.3. Philosophical Debates

The question of **"Can machines be truly creative?"** continues to divide scholars. Some adopt a behaviorist view, judging creativity solely by output; others insist on intentionality, consciousness, or self-reflective capacities.

Colton et al. (2012) propose a **"creative tripod"**—skill, appreciation, and imagination—as a framework for evaluating machine creativity. Few current systems satisfy all three.

4.7. Summary of Findings

The findings of this systematic review reveal that:

- NLP has advanced significantly in simulating creative writing, with LLMs achieving near-human fluency and stylistic adaptability.
- Poetry and storytelling remain dominant research areas, but metaphor and humor require further exploration.
- Evaluation frameworks are evolving but still insufficient for capturing the multi-dimensionality of creativity.
- Human-AI collaboration is a promising area, though challenges of agency, authorship, and influence persist.
- Ethical concerns about bias, originality, and cultural impact must be addressed in future development.

Chapter Five: Discussion

5.1. Introduction

This chapter offers an in-depth interpretation of the findings presented in Chapter Four, contextualized within the broader intellectual discourse on artificial intelligence, creativity, and human-computer interaction. The discussion critically examines how Natural Language Processing (NLP) has contributed to computational creativity and reflects on the theoretical, methodological, and ethical implications of these contributions. Key themes include the extent to which NLP models simulate or approximate creative behaviors, the challenges of evaluating machine-generated creativity, the sociocultural ramifications of algorithmic authorship, and future directions for research and application. This chapter also highlights tensions and synergies between computational efficiency and aesthetic or humanistic value in creative language generation.

5.2. NLP and the Simulation of Creativity

5.2.1. Imitation Versus Innovation

A core insight from the reviewed literature is that while modern NLP systems—particularly large-scale transformer models—are adept at mimicking the stylistic and syntactic surface structures of creative writing, they often do so without achieving what Boden (2004) terms "transformational creativity," i.e., the ability to produce truly novel conceptual spaces. These systems excel at *combinatorial creativity* (reassembling existing patterns) and sometimes reach *exploratory creativity*

(navigating stylistic spaces), but they fall short in transcending the frameworks within which they operate.

GPT-based models, for example, are capable of writing in iambic pentameter, generating coherent short stories, and emulating known authors' voices. However, their lack of self-awareness, intentionality, and emotional embodiment means that they generate what Margaret Boden calls "pseudocreativity"—outputs that are original in form but not necessarily in conceptual substance or emotional depth.

5.2.2. Syntax Versus Semantics

The distinction between syntactic fluency and semantic richness is crucial in this discussion. While statistical and deep learning models can generate text with impressive grammaticality, they often lack semantic depth, failing to understand or maintain the thematic, symbolic, and metaphorical coherence required in genuinely impactful creative writing. Many outputs reviewed exhibited a lack of long-range coherence, cliché repetition, or emotionally flat narratives despite their technical sophistication.

This limitation becomes more pronounced in metaphor, irony, and humor generation—areas that demand an understanding of context, contradiction, and layered meaning. Here, systems frequently produce either banal or unintentionally nonsensical results, revealing the challenges of encoding cultural and affective intelligence into computational models.

5.3. *The Role of Human-AI Collaboration*

5.3.1. Augmentative Creativity

One of the most promising trends emerging from this review is the shift from autonomous AI authorship to **augmented creativity**, where humans and machines co-create. Tools like Sudowrite, ChatGPT, and CoPoet illustrate that NLP can serve as a catalyst for human creativity by offering lexical suggestions, expanding narratives, or generating stylistic variants. In this context, AI functions less as an artist and more as a creative assistant or muse.

This aligns with the **distributed cognition** model (Hollan et al., 2000), where creative processes are seen as emergent from systems that include both human and non-human actors. Such a model reconfigures authorship as a collaborative act and opens new possibilities for experimental literature, transmedia storytelling, and educational applications.

5.3.2. Risks of Over-Reliance

Despite these benefits, the review also raises cautionary notes. Over-reliance on AI tools may lead to homogenized stylistic patterns, a loss of creative confidence among human writers, and an erosion of the aesthetic value derived from human struggle and originality. As AI becomes more integrated into writing workflows, the boundary between imitation and inspiration becomes increasingly blurred, complicating notions of authenticity and authorial voice.

Furthermore, the psychological implications of constant AI co-authorship are underexplored. Early findings suggest that while novice writers feel empowered by AI support, experienced authors sometimes feel that their creative agency is undermined or diluted.

5.4. *Challenges of Evaluation in Computational Creativity*

5.4.1. Inadequacy of Quantitative Metrics

The literature review reveals that most automated evaluation metrics (e.g., BLEU, ROUGE, perplexity) fail to adequately capture the multifaceted nature of creativity. These metrics prioritize lexical overlap, frequency-based fluency, or next-token predictability, none of which directly relate

to core creative qualities such as originality, emotional resonance, thematic cohesion, or aesthetic value.

Indeed, some of the most "creative" machine outputs—such as metaphorically rich poetry or experimental narratives—may score poorly on these metrics due to their deviation from normative linguistic patterns. This underscores the need for novel evaluation paradigms that are both human-centered and context-sensitive.

5.4.2. Toward Multi-Dimensional and Hybrid Evaluation

A key contribution of this review is the emphasis on **hybrid evaluation frameworks**, combining human judgment, task-based performance, and algorithmic feedback. Multi-criteria evaluation schemes that assess outputs along axes such as fluency, coherence, novelty, affective power, and genre appropriateness offer a more nuanced understanding of machine creativity.

Moreover, participatory evaluation methods—involving both creators and audience members— can provide richer insights into how creative outputs are received, interpreted, and valued. Such approaches are vital for aligning NLP systems with human aesthetic standards and expectations.

5.5. Ethical and Philosophical Implications

5.5.1. Bias, Stereotyping, and Cultural Flattening

A recurring theme across the reviewed literature is the replication and amplification of biases present in training data. Language models trained on internet-scale corpora tend to reflect dominant sociocultural norms, thereby marginalizing underrepresented voices and reinforcing stereotypes. In creative writing contexts, this can manifest as gendered tropes, racial clichés, or heteronormative narratives.

The cultural implications of this are profound. If creative AI systems primarily reflect Western, English-language, male-authored literary traditions, they risk contributing to a **flattening of global literary diversity**. Ensuring equitable representation in training data, and developing models attuned to diverse cultural idioms and narrative forms, is a pressing challenge.

5.5.2. Authorship and Intellectual Property

The emergence of AI-generated literature poses novel questions about authorship and ownership. Who owns a poem written by a language model? Can AI be credited as a co-author? Should AI-generated content be eligible for literary awards or publication? These questions remain unresolved in legal and literary theory alike.

Current copyright law typically denies authorship rights to non-human entities, but as AI-generated works become increasingly indistinguishable from human ones, the boundary becomes legally and philosophically contentious. This review recommends the development of new **creative commons frameworks**, hybrid authorship models, and clearer attribution protocols to navigate this complexity.

5.5.3. Redefining Creativity and the Human

At the deepest level, the integration of NLP into creative domains compels us to rethink what it means to be creative—and what it means to be human. If creativity can be simulated through statistical modeling, does that undermine its mystique or democratize it? Should creativity be defined by process (conscious inspiration) or product (novel and valuable output)?

Philosophers such as Dennett (1991) argue for a functionalist view, where behavior alone determines mental states. Others, like Searle (1980), insist that true creativity requires intentionality, which machines lack. The reviewed literature reflects this divide, with some scholars embracing computational creativity as genuine and others treating it as sophisticated mimicry.

Ultimately, the question may not be whether machines are creative, but how their outputs challenge and expand our understanding of creativity itself.

5.6. Contributions to the Field

This systematic review contributes to the academic discourse in several key ways:

1. **Taxonomy of Models and Tasks**: It provides a structured classification of NLP architectures and their application to various creative genres, offering a roadmap for future model development and research alignment.
2. **Evaluation Critique and Framework**: It critiques existing evaluation practices and calls for a more holistic, interdisciplinary framework that aligns technical fluency with humanistic standards.
3. **Ethical Awareness**: It foregrounds ethical, cultural, and epistemological concerns, advocating for inclusive and responsible AI development in creative domains.
4. **Human-AI Interaction Insight**: It highlights the complex dynamics of human-AI co-creation and suggests design principles for augmentative systems that respect and empower human agency.
5. **Future Research Agenda**: It identifies underexplored areas such as non-Western creative traditions, long-form narrative coherence, affective computing in writing, and the sociology of AI-authored literature.

5.7. Summary

This chapter has interpreted the findings of the systematic review through the lenses of theoretical, practical, and ethical inquiry. It has argued that while NLP has made substantial progress in simulating creative language, it remains constrained by limitations in semantic depth, cultural representation, and evaluative precision. The rise of co-creative systems opens new possibilities for augmenting human creativity but also introduces new challenges in authorship, identity, and meaning. As computational creativity matures, future work must be increasingly interdisciplinary, reflexive, and culturally grounded to ensure that AI contributes not only to innovation but also to imagination and inclusivity in the creative arts.

Chapter Six: Conclusion and Future Research Directions

6.1. Conclusions

This systematic review has explored the dynamic and evolving interface between **Natural Language Processing (NLP)** and **computational creativity**, focusing particularly on the capacity of NLP systems to simulate, support, or augment creative writing processes. Drawing from a corpus of 112 rigorously selected peer-reviewed articles, conference papers, and technical reports published between 2005 and 2024, this review has synthesized insights from computer science, computational linguistics, artificial intelligence, digital humanities, and cognitive psychology.

The findings suggest that while NLP systems—especially those built on deep learning architectures such as transformers—have achieved remarkable progress in mimicking the stylistic surface of human-authored texts, significant limitations remain in generating content that reflects the intentionality, originality, emotional depth, and cultural context traditionally associated with human creativity.

6.1.1. Summary of Key Findings

- **Model Evolution**: The transition from rule-based systems to statistical models, and ultimately to transformer-based large language models (LLMs), has significantly enhanced the fluency,

coherence, and stylistic adaptability of machine-generated creative writing. Yet, issues of thematic depth, long-form consistency, and metaphorical richness remain persistent challenges.

- **Genre Applications**: Poetry and storytelling emerged as dominant areas of research, while subfields such as figurative language generation, stylistic emulation, and humor remain comparatively underexplored. Applications increasingly support co-creative environments, transforming AI from autonomous writer to collaborative partner.

- **Evaluation Gaps**: Current evaluation metrics are inadequate for capturing multi-dimensional aspects of creativity. The heavy reliance on automatic metrics like BLEU and perplexity fails to account for originality, emotional resonance, or aesthetic value. Hybrid human-AI evaluation frameworks are urgently needed.

- **Human-AI Interaction**: The shift toward **human-in-the-loop** or **co-creative systems** has introduced new paradigms of collaborative authorship. While these tools democratize access to creative expression and offer novel forms of engagement, they also risk diluting human agency and introducing stylistic homogeneity.

- **Ethical and Philosophical Challenges**: Machine-generated creativity raises unresolved questions about authorship, bias, originality, and cultural hegemony. The opacity of LLMs, the potential propagation of stereotypes, and the reproduction of dominant literary norms necessitate robust ethical frameworks and inclusive design strategies.

6.2. Theoretical Implications

This review contributes to the growing body of literature seeking to redefine **creativity in the age of AI**. It challenges essentialist definitions of creativity as an exclusively human capacity by demonstrating how algorithmic systems can simulate many of its outward expressions. At the same time, it reinforces the idea that **true creativity involves more than novelty and fluency**; it encompasses intention, meaning-making, and contextual awareness—dimensions that current NLP systems only approximate.

Philosophically, this review aligns with **functionalist and behaviorist perspectives** that assess creativity based on outputs and effects, rather than inner states. However, it also cautions against collapsing human and machine creativity into a single category without recognizing their ontological and epistemological differences. Future theoretical work must grapple with these tensions to refine our understanding of what it means to be creative—and what it means to create—with machines.

6.3. Practical Implications

6.3.1. For AI and NLP Developers

- Develop more transparent and controllable models that allow users to understand and guide creative processes.
- Integrate fine-grained feedback loops and user customizability to support human creativity rather than override it.
- Incorporate ethical modules that detect and mitigate biased, offensive, or culturally insensitive content in creative applications.

6.3.2. For Educators and Writers

- Use NLP-based co-writing tools as **pedagogical aids** in creative writing courses, fostering experimentation and iterative feedback.

- Treat AI-generated outputs as opportunities for **critical reflection**, revision, and creative augmentation rather than final products.
- Encourage exploration of **AI as a collaborative partner** in artistic expression, especially for marginalized voices who may gain access to new literary tools.

6.3.3. For Policy Makers and Legal Scholars

- Address the **legal status** of AI-generated content in terms of intellectual property, authorship, and attribution.
- Develop clear guidelines for disclosure, ensuring that readers and audiences can distinguish between human and machine-generated works.
- Promote **algorithmic accountability** by mandating transparency reports, ethical reviews, and user consent mechanisms in creative AI platforms.

6.4. Limitations of the Review

While this review aims to offer a comprehensive synthesis of the field, several limitations must be acknowledged:

- **Language Bias**: The review included only English-language publications, potentially overlooking valuable research in other linguistic traditions.
- **Rapidly Evolving Field**: The NLP landscape, particularly with respect to LLMs, is evolving at a pace that may render some findings quickly outdated.
- **Lack of Quantitative Meta-Analysis**: The diversity of methods and reporting standards made it infeasible to conduct a statistical meta-analysis of performance outcomes.
- **Absence of Multimodal Creativity**: Although this review focused on text, many creative systems now integrate vision, sound, and movement, which were not included in the scope.

6.5. Future Research Directions

Based on the gaps identified and the emerging trends in the field, several key directions for future inquiry are proposed:

6.5.1. Toward Context-Aware Creativity

Most NLP models operate without deep contextual grounding. Future systems should incorporate **situated knowledge, user intent, and cultural frames** to produce more meaningful and nuanced outputs. This could involve:

- Multi-modal input (e.g., images, music, or user profiles).
- Memory-augmented models that retain narrative consistency across long-form texts.
- Socio-pragmatic models capable of adjusting tone, style, or form based on context.

6.5.2. Ethical and Inclusive NLP

There is a critical need to **decolonize NLP training corpora** and promote linguistic, cultural, and stylistic diversity in creative outputs. Strategies include:

- Curating multilingual and multicultural literary datasets.
- Embedding ethical evaluation protocols in model development pipelines.
- Supporting participatory design practices that engage artists, writers, and minority communities.

6.5.3. Creativity-Aware Evaluation Metrics

Future research should focus on building evaluation tools aligned with human judgments of creativity, such as:

- Machine learning models trained on annotated corpora of human ratings for originality, emotional impact, and aesthetic quality.
- Cognitive neuroscience approaches to assess user engagement with AI-generated content.
- Dynamic feedback mechanisms that learn user preferences in co-creative settings.

6.5.4. Interdisciplinary Research Frameworks

Given the complexity of computational creativity, future studies must bridge disciplines by:

- Integrating computational methods with literary theory, cognitive psychology, and philosophy of art.
- Establishing common vocabularies and frameworks across computer science and the humanities.
- Creating research consortia and interdisciplinary journals focused on AI and the arts.

6.6. Final Reflections

The fusion of Natural Language Processing and computational creativity marks a paradigmatic shift in the way we conceptualize authorship, imagination, and artistic labor. What once required intuition, talent, and cultural immersion can now be simulated—albeit imperfectly—by predictive algorithms trained on vast textual corpora. This transformation challenges us to reconsider the boundaries of creativity and to develop new tools, ethics, and evaluative standards that keep pace with technological innovation.

Rather than seeing AI as a threat to human creativity, this review invites a more **nuanced and collaborative perspective**—one in which machines augment, rather than replace, the rich tapestry of human expression. As we enter a future of increasingly capable generative systems, the most pressing task may not be to draw lines between human and machine authorship, but to ensure that creativity—in all its diversity, complexity, and cultural significance—remains a shared and evolving pursuit.

References

Rahman, M. H., Kazi, M., Hossan, K. M. R., & Hassain, D. (2023). The Poetry of Programming: Utilizing Natural Language Processing for Creative Expression.

Boden, M. A. (2004). *The Creative Mind: Myths and Mechanisms* (2nd ed.). Routledge.

Colton, S., Pease, A., & Ritchie, G. (2001). The Effect of Input Knowledge on Creativity. In *Proceedings of the AISB Symposium on AI and Creativity in Arts and Science*.

Veale, T., & Hao, Y. (2008). A Fluid Knowledge Representation for Understanding and Generating Creative Metaphors. *Knowledge-Based Systems*, 21(7), 614–622.

Ghazvininejad, M., Shi, X., Choi, Y., & Knight, K. (2017). Hafez: An Interactive Poetry Generation System. In *Proceedings of ACL 2017, System Demonstrations* (pp. 43–48).

Lau, J. H., Cohn, T., & Baldwin, T. (2018). Deep-speare: A joint neural model of poetic language, meter and rhyme. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* (pp. 1948–1958).

Fan, A., Lewis, M., & Dauphin, Y. (2018). Hierarchical Neural Story Generation. In *Proceedings of ACL 2018* (pp. 889–898).

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language Models are Unsupervised Multitask Learners*. OpenAI Technical Report.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is All You Need. In *Advances in Neural Information Processing Systems* (pp. 5998–6008).

Pillutla, K., Li, S., & Zettlemoyer, L. (2021). MAUVE: Measuring the Gap Between Neural Text and Human Text using Divergence Frontiers. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Holyoak, K. J., & Thagard, P. (1995). *Mental Leaps: Analogy in Creative Thought*. MIT Press.

Dennett, D. (1991). *Consciousness Explained*. Little, Brown and Co.

Searle, J. R. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 3(3), 417–457.

McCormack, J., Gifford, T., & Hutchings, P. (2019). Autonomy, Authenticity, Authorship and Intention in Computer Generated Art. *IEEE Transactions on Affective Computing*, 10(3), 351–363.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V. (2019). XLNet: Generalized Autoregressive Pretraining for Language Understanding. In *Advances in Neural Information Processing Systems*.

Holton, R. (2009). AI and the 'Art' of Creativity. *Leonardo*, 42(5), 418–423.

Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017). CAN: Creative Adversarial Networks, Generating 'Art' by Learning About Styles and Deviating from Style Norms. In *Proceedings of the 8th International Conference on Computational Creativity*.

Lin, C.-Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. In *Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004)*.

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (pp. 311–318).

Dincer, B., & Aydın, C. C. (2021). The Ethics of AI-generated Literature: Issues of Responsibility, Accountability, and Meaning. *AI & Society*, 36, 1105–1118.

Floridi, L. (2019). *The Logic of Information: A Theory of Philosophy as Conceptual Design*. Oxford University Press.

Ha, D., & Eck, D. (2018). A Neural Representation of Sketch Drawings. In *International Conference on Learning Representations (ICLR)*.

Manjavacas, E., & Koolen, C. (2021). Stylometry for Authorship Attribution: A Literature Review. *Digital Scholarship in the Humanities*, 36(Supplement_1), i108–i126.

Sharma, R., & Kaur, H. (2022). A Review of Deep Learning Techniques in Poetry Generation. *International Journal of Advanced Computer Science and Applications*, 13(1), 527–536.

Zhang, Y., Sun, S., Galley, M., Chen, Y. C., Brockett, C., Gao, X., & Dolan, B. (2020). DialoGPT: Large-Scale Generative Pre-training for Conversational Response Generation. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 270–278.

Hutchinson, B., Prabhakaran, V., Denton, E., Webster, K., Zhong, Y., & D'Amour, A. (2021). Towards Transparent and Accountable NLP: A Review of Bias in Language Model Applications. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 629–644.