AI Master's Thesis Proposal: User Profiling of AIGC Community in Social Media Communities - A Comparative Study of Discord and Twitch

Peixiong Zheng

Supervisor: Dr. Almila Akdag
Second Supervisor: Dr. Jullian Frommel

Graduate School of Natural Sciences
Utrecht University
October 30, 2024

Contents

1 Introduction

In recent years, the rapid advancements in generative artificial intelligence (GenAI) have transformed the creative landscape, particularly in the domains of content generation and digital media. Models like OpenAI's GPT series, Stable Diffusion, and other deep learning algorithms have unlocked new forms of artistic expression, allowing creators to push the boundaries of what is considered possible in visual, textual, and multimedia creation. This shift has raised critical questions about creativity, authorship, and the impact of AI tools on traditional artistic workflows.

Platforms like Twitch and Discord serve as key spaces where discussions around AI-generated content (AIGC) thrive. On these platforms, users engage both with the tools themselves and in discourse surrounding their applications and implications. While some users experiment with these AI tools for content creation, others debate the ethical and cultural ramifications of AI's growing role in the creative process. Understanding how these conversations evolve and what key topics emerge in such discussions is crucial for analyzing the current and future landscape of AI-driven creativity.

A major component of this thesis is the use of topic modeling techniques to analyze discussions surrounding AIGC on social media platforms. Latent Dirichlet Allocation (LDA) and Gibbs Sampling Dirichlet Mixture Model (GSDMM) are employed to identify and categorize the recurring themes within these discussions. These probabilistic models enable us to uncover latent thematic structures by grouping co-occurring words into topics. However, while LDA and GSDMM are essential in identifying overarching themes, they have limitations, particularly when dealing with fragmented and dynamic conversations typical of platforms like Twitch and Discord.

To address these limitations, Large Language Models (LLMs) like ChatGPT play a complementary role in this research. While LDA and GSDMM provide a high-level categorization of topics, LLMs are applied for more nuanced tasks, such as topic naming and further qualitative analysis. ChatGPT's conversational capabilities allow it to offer deeper insights into user-generated content by contextualizing the thematic clusters identified by LDA and GSDMM. This capability is crucial in a dynamic environment where discussions often shift in focus and require real-time interpretation. Furthermore, ChatGPT is used to simulate ongoing discussions, making it an invaluable tool for tracking how conversations about AIGC evolve over time.

The motivation behind this research stems from the need to bridge the gap between traditional topic modeling techniques and the more advanced analysis enabled by LLMs. By comparing LDA and GSDMM's performance with that of ChatGPT, this thesis aims to demonstrate how LLMs can augment and enhance the understanding of social media conversations, providing more accurate and context-aware interpretations of user interactions. Additionally, this study contributes to the ongoing debate about AI's role in creative industries, examining how online communities perceive and discuss AI-generated content, and what this might mean for the future of digital art and content creation.

In conclusion, this research integrates both traditional topic modeling approaches and advanced LLM analysis to explore the discussions surrounding AI-generated content on platforms like Twitch and Discord. The combination of these methods provides a comprehensive view of how users engage with AIGC, highlighting the evolving nature of online discussions and the potential of LLMs to offer deeper, more contextually aware insights into social media data. The results of this study will have implications not only for the analysis of user-generated content but also for the broader discourse on the intersection of AI and creativity.

2 Literature Review

This literature review is structured into four main sections. The first section examines the impact of AI-generated content on digital communities, highlighting both opportunities and challenges. The second section focuses on large language models (LLMs) used for data analysis, exploring their applications in social media research. The third section discusses the interaction dynamics of social media platforms like Twitch and Discord, analyzing their roles in content creation and community engagement. Finally, the review identifies key research gaps and outlines how this study aims to address them.

2.1 AI-Generated Content (AIGC) and its Impact on Digital Communities

AI-Generated Content (AIGC) has rapidly emerged as a transformative force in the digital landscape, influencing industries ranging from entertainment and digital marketing to education and online communities. AIGC refers to any form of media—text, images, video, and audio—created autonomously by artificial intelligence systems. Leveraging advancements in deep learning, natural language processing (NLP), and generative models, AIGC has revolutionized content creation by automating processes previously dependent on human intervention.

The proliferation of AIGC has opened up new avenues for creativity, democratizing access to tools that were once limited to professionals. For instance, platforms like DALL·E, Stable Diffusion, and MidJourney allow users to generate high-quality artwork through simple text prompts [21]. These technologies enable individuals with limited technical expertise to produce visual and textual content, fundamentally shifting the creative process.

At the same time, AI-generated content has raised critical questions about authorship, originality, and the ethical implications of automation in creative fields. Researchers such as Anyatasia [3] have examined the dual-edged nature of AIGC, arguing that while these tools empower creators and foster innovation, they also introduce concerns related to intellectual property, artistic authenticity, and the displacement of human labor.

2.1.1 Rise of Generative Models in AIGC

The rise of Generative Adversarial Networks (GANs) and Transformer-based architectures has been a driving force behind the advancements in AI-generated content. GANs, introduced by Goodfellow et al. [13], consist of two competing networks: a generator that creates new data and a discriminator that evaluates the authenticity of that data. This adversarial training process results in highly realistic outputs, especially in visual media such as images and video. GANs have been widely adopted for tasks like image synthesis, style transfer, and deepfake creation, transforming the landscape of digital art and entertainment.

On the other hand, Transformer-based models, particularly GPT (Generative Pre-trained Transformers) and BERT (Bidirectional Encoder Representations from Transformers), have revolutionized the generation of textual content. These models, leveraging attention mechanisms to capture long-term dependencies in text, are capable of producing coherent and contextually relevant language, enabling them to generate articles, dialogues, and even entire stories [35]. The GPT-3 model, for example, has been used in various applications ranging from content creation for blogs and news outlets to virtual assistants and chatbots.

These technologies have also enabled the rise of multimodal AIGC, where models like CLIP (Contrastive Language–Image Pretraining) and DALL·E integrate text and images to generate new forms of content [26]. By aligning text and image representations, these models can generate creative and aesthetically appealing visuals from descriptive language, further expanding the possibilities of AI in creative industries.

2.1.2 Social Media and Community Dynamics Around AIGC

The integration of AI-generated content (AIGC) into social media platforms has transformed community dynamics, particularly on platforms like Twitch and Discord, which serve as active hubs for discussions surrounding AI art and creativity. These platforms enable users to exchange ideas, showcase their AI-generated creations, and engage in conversations about the artistic, technical, and ethical dimensions of these technologies. Notably, Discord communities centered on tools such as Stable Diffusion and MidJourney have emerged as vital spaces for collaboration, where AI art enthusiasts experiment with different AI models and collectively improve their creative output [22, 12].

Oppenlaender et al. [22] describe how these communities foster peer-to-peer learning, creating an environment in which users share knowledge and techniques to enhance the creativity and quality of AI-generated works. This sense of collective learning is further emphasized by Anyatasia [2], who highlights the importance of feedback loops in these discussions, as real-time peer feedback helps artists rapidly refine their creations and iterate on AI-generated content. These dynamics underscore the importance of community collaboration in improving AI-generated art.

However, the nature of discussions and interactions around AIGC differs between platforms. On Twitch, where live interactions and visual content predominate, discussions are often more reactive and emotionally charged. Users respond to AI-generated visuals and streams in real-time, shaping a dynamic, entertainment-focused dialogue. On Discord, the structure of interactions allows for more in-depth, technical conversations, where users engage in thoughtful discussions about the creative and ethical implications of AI-generated content. Research indicates that Discord users often delve into technical intricacies, such as model training, dataset usage, and parameter optimization [12, 29], reflecting a more nuanced approach to AI creativity than typically found in fast-paced Twitch chats.

These platform-specific dynamics highlight the ways in which social media ecosystems influence user engagement with AI-generated content. Twitch's immediacy fosters a focus on the aesthetic and entertainment aspects of AIGC, while Discord enables deeper explorations of AI's role in the creative process. Understanding these distinctions is essential for comprehending the broader impact of AIGC on online communities and the way users interact with AI-generated works across different platforms.

2.1.3 AIGC in Creative Industries

Beyond social media, AIGC has had a profound impact on creative industries, reshaping workflows and democratizing content creation. AI tools are increasingly used in fields such as video game development, film production, graphic design, and digital marketing, where they streamline repetitive tasks, enhance efficiency, and augment human creativity. Platforms like RunwayML provide AI-powered tools that allow filmmakers to generate visual effects, animations, and enhanced footage, reducing the time and cost associated with traditional production methods [12, 3, 31].

The accessibility of AIGC tools has empowered independent creators and small studios, allowing them to compete with larger production companies by reducing the barriers to entry in content creation. However, this democratization presents challenges, particularly in maintaining quality and originality in an increasingly saturated market of AI-generated content. As Oppenlaender et al. [22] note, the increasing prevalence of AI-generated works necessitates new methods for distinguishing between human-created and AI-generated content, particularly in competitive creative fields.

Moreover, the integration of AI into creative workflows has fostered new forms of human-AI collaboration. Co-creation, where human artists guide the creative process by setting parameters or providing input while AI generates content, has become a prominent feature in industries such as graphic design and music composition [21, 29]. This hybrid form of creativity challenges traditional concepts of authorship and artistic originality, raising important questions about the role of human artists in an era of autonomous AI-generated content. As Anyatasia [2] emphasizes, such collaborations lead to an evolving understanding of creativity, where the human artist becomes more of a curator or conductor, guiding the output generated by AI while the machine assumes the role of an automated creator.

These developments not only change the way content is produced but also influence broader ethical and legal discussions regarding intellectual property and the ownership of AI-generated works. The European Union's Copyright Directive, for example, continues to grapple with questions about whether AI-generated content can be protected by copyright, and if so, who should own the rights [31]. The blurring of boundaries between human and machine creativity invites ongoing dialogue about the future of authorship in AI-driven creative industries.

2.2 Architectures of LLMs Employed for Data Analysis

As discussions around AI-Generated Content (AIGC) expand, understanding the architectures of large language models (LLMs) used in analyzing such data becomes essential. Unstructured data, such as forum threads or community chats, requires models that can handle context, ambiguity, and high-dimensional data. This section introduces key LLM architectures and assesses their contributions to analyzing discussions on platforms like Discord and Twitch.

2.2.1 BERT (Bidirectional Encoder Representations from Transformers)

BERT, developed by Devlin et al. [11], introduced a novel bidirectional approach to pretraining transformer models. Unlike earlier autoregressive models, which process text in a single direction (either left-to-right as in GPT, or right-to-left), BERT's architecture allows for bidirectional encoding, enabling the model to capture contextual dependencies in both directions simultaneously. This bidirectional nature is critical for understanding the full context of each word in a sentence, particularly in tasks where meaning depends on both previous and future words.

BERT's architecture consists of several stacked layers of encoders (12 for the base model and 24 for the large model), where each encoder follows the transformer architecture proposed by Vaswani et al. [35]. Each encoder layer has two key components: multi-head self-attention and a feed-forward neural network.

The following are the key components of BERT's architecture:

1. **Input Representation:** BERT's input is a sequence of tokens, which include both the words of a sentence and special tokens like `[CLS]` (classification token) and `[SEP]` (separator token). Each token is represented by the sum of its token embedding, positional encoding, and segment embedding. The embedding layer is used to map input tokens into a dense vector space, which is crucial for capturing the nuances of word relationships and positions in a sentence.

2. **Positional Encoding:** Since transformers do not inherently account for the order of tokens, positional encodings are added to the token embeddings. These encodings enable BERT to capture the position of each token in the sequence, allowing the model to differentiate between tokens that appear at different positions.

3. **Encoder Layers:** BERT's transformer encoder layers are responsible for processing the input sequence. Each encoder consists of:

- **Multi-head Self-Attention Mechanism:** This allows the model to attend to different parts of the input sequence simultaneously, capturing long-range dependencies and relationships between words. BERT uses self-attention to compute attention scores for every pair of words in the input, which helps in understanding the context by considering all other words in the sentence.

- **Feed-forward Neural Network:** After the attention mechanism, the output is passed through a feed-forward network, which applies non-linearity and transformations to the data. Each encoder has its own set of parameters, allowing for more transformations as the data passes through multiple layers.

In the case of the base model, BERT has 12 such layers stacked on top of each other. In the large model, 24 layers are used. Each layer refines the representation of the input sequence, enabling the model to capture increasingly linguistic features.

4. **Masked Language Modeling (MLM):** One of the key innovations in BERT's pretraining is masked language modeling. During pretraining, 15% of the input tokens are randomly masked, and the model is tasked with predicting the original tokens based on the context provided by the unmasked tokens. This allows the model to learn strong bidirectional representations, as it must use both the left and right context to predict the missing token. MLM enables BERT to capture contextual information more effectively than models that predict the next token in a sequence.

5. **Next Sentence Prediction (NSP):** Another task used during BERT's pretraining is next sentence prediction. This involves providing the model with pairs of sentences and training it to predict whether the second sentence follows the first one in the text. By incorporating NSP, BERT learns to model relationships between sentences, which is particularly useful for tasks like question answering and natural language inference. The NSP objective is crucial for understanding discourse-level relationships in text.

6. **Output Representation:** After passing through the encoder layers, BERT produces two types of output:

- `[CLS]` **token output:** The embedding corresponding to the special `[CLS]` token is often used for classification tasks. It represents a summary of the entire input sequence and is fed into a classifier for downstream tasks.
- **Token-wise output:** BERT also generates an embedding for each token in the input sequence, which can be used for token-level tasks such as named entity recognition (NER) or part-of-speech tagging.

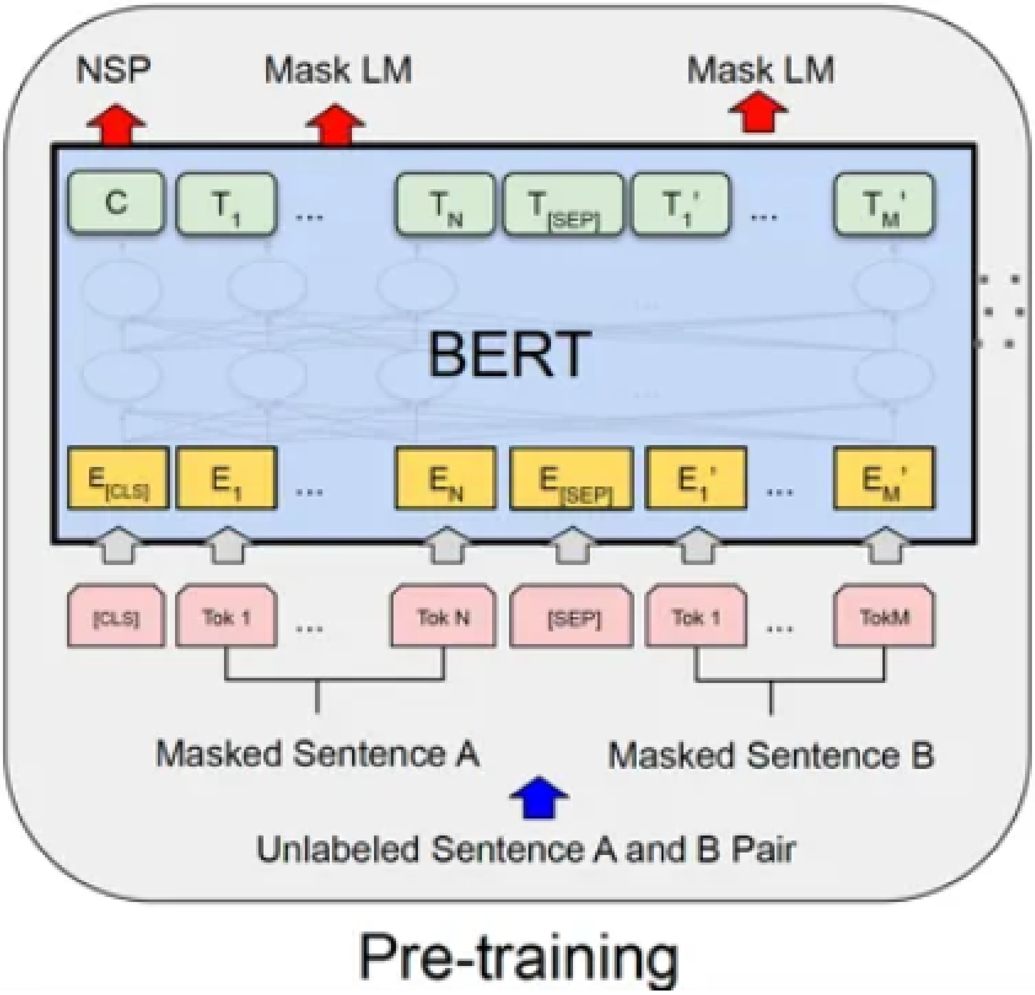

Figure 1: BERT Model Pipeline: Embedding, Encoder Layers, and Classifier. The embedding layer processes input tokens with positional encodings. The sequence passes through multiple encoder layers, and the output is used for classification tasks.

Applications of BERT in Data Analysis:

BERT's architecture is particularly suited for tasks that require deep understanding of context, such as:

- **Question Answering (QA):** BERT's bidirectional nature allows it to understand both the question and the context in which an answer appears, making it effective for extracting answers from text.

- **Text Classification:** By using the [CLS] token as a summary of the input sequence, BERT is effective in classifying text into categories such as sentiment analysis, spam detection, or intent classification.

- **Named Entity Recognition (NER):** BERT's token-level output is ideal for tasks that require identifying specific entities (e.g., people, organizations) within a text.

The combination of masked language modeling and next sentence prediction enables BERT to excel at both word-level and sentence-level tasks, making it a versatile model for analyzing textual data on platforms like forums and streaming communities, where context can shift between individual words or across sentences.

2.2.2 RoBERTa (A Robustly Optimized BERT Pretraining Approach)

RoBERTa, introduced by Liu et al. [19], is a robustly optimized version of BERT that incorporates several key improvements in the pretraining process. While RoBERTa's architecture remains largely identical to BERT, the optimizations in training strategies allow it to leverage data more effectively and capture more intricate language patterns.

Unlike BERT, which includes the next sentence prediction (NSP) task during pretraining, RoBERTa eliminates this task entirely. Liu et al. found NSP to contribute little to downstream task performance. Instead, RoBERTa focuses exclusively on the masked language modeling (MLM) task. Several optimizations in RoBERTa's pretraining procedure include:

- **Increased training epochs:** RoBERTa is trained for longer periods, allowing the model to refine its understanding of language patterns.

- **Dynamic masking:** RoBERTa employs dynamic masking during training, meaning that different tokens are masked across training epochs. This prevents the model from overfitting to specific tokens and encourages more generalizable language representations.

- **Larger batch sizes and learning rates:** By increasing batch sizes and fine-tuning learning rates, RoBERTa can process more data at once, leading to improved performance.

RoBERTa's architecture retains the transformer-based encoder structure from BERT, which consists of layers of self-attention mechanisms and feed-forward neural networks. The input token representation remains similar, using a combination of token embeddings, positional encodings, and segment embeddings. As with BERT, RoBERTa uses the special [CLS] token for classification tasks.

Key Advantages of RoBERTa over BERT:

RoBERTa's improved training strategies make it particularly effective for tasks that involve large-scale datasets or require long-term contextual understanding. Key tasks where RoBERTa demonstrates superiority include:

- **Text Classification:** RoBERTa can be fine-tuned on various text classification tasks, such as sentiment analysis or topic classification, using the [CLS] token's output.

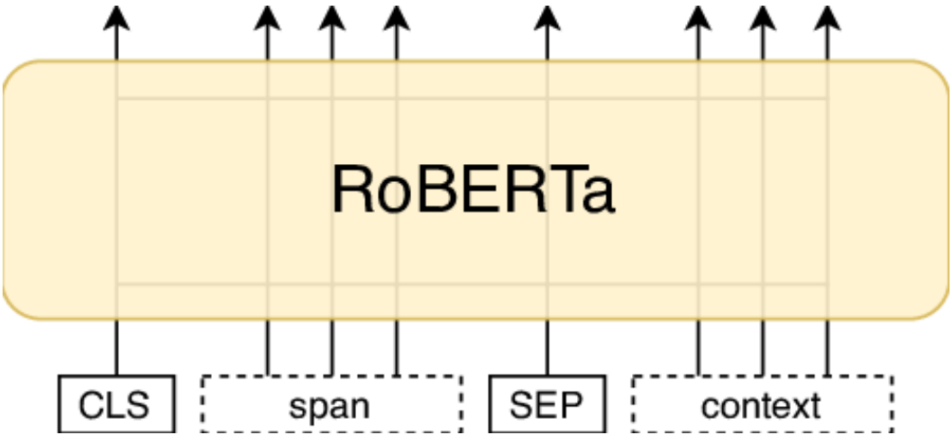

Figure 2: RoBERTa Model Pipeline: Input tokens like [CLS], spans, [SEP], and context are passed through the RoBERTa encoder. The encoder outputs token representations for downstream tasks such as classification or sequence labeling.

- **Topic Modeling:** RoBERTa's ability to capture language patterns makes it ideal for tasks like topic modeling, where understanding latent structures in documents is essential.
- **Question Answering:** By removing the NSP objective, RoBERTa improves its ability to model long-range dependencies between question and context, leading to better performance on question-answering tasks.

In terms of benchmarks, RoBERTa consistently outperforms BERT on tasks like the GLUE benchmark [36], achieving higher accuracy across tasks like sentiment analysis, text classification, and natural language inference. Its architectural similarity to BERT, combined with improved pretraining, allows RoBERTa to capture deeper contextual meanings from large datasets.

2.2.3 T5 (Text-to-Text Transfer Transformer)

T5, introduced by Raffel et al. [29], is based on the principle that all natural language processing (NLP) tasks can be framed as text-to-text problems. It converts both the input and the output into sequences of text, regardless of whether the task involves translation, summarization, classification, or even regression. By adopting this unified framework, T5 streamlines the approach to handling a wide variety of NLP tasks.

The T5 model follows a transformer-based encoder-decoder design, where the input is first processed by an encoder, and the output is subsequently generated by a decoder. In both the encoder and decoder, T5 employs multi-head self-attention mechanisms. The encoder takes in the input text, processes its context, and transforms it into a latent representation. The decoder, meanwhile, generates the output sequence by attending to both the encoder's representation and the previously generated tokens in the output sequence.

Key components of T5's architecture include:

- **Encoder-Decoder Structure:** T5 is based on the transformer architecture, utilizing an encoder-decoder structure that processes text input in a sequence-to-sequence format. The encoder first processes the input text, creating a contextualized representation, while the decoder predicts the output by attending to the encoder's output and generating text autoregressively.

- **Multi-head Self-Attention:** Both the encoder and decoder make use of multi-head self-attention mechanisms to allow the model to focus on different parts of the input at once, capturing relationships between words and providing context for word disambiguation.

- **Text-to-Text Framework:** T5 treats all tasks as a text-to-text problem. For instance, a translation task would have an input like "translate English to German: That is good." and an output "Das ist gut." Similarly, for summarization tasks, the input might be "summarize: [long text]", and the output would be a concise summary.

During training, T5 is pre-trained on the C4 (Colossal Clean Crawled Corpus) dataset, which consists of diverse web-based texts. The model is trained using a span-corruption objective, where random spans of text are replaced with a special token, and the model is tasked with reconstructing the original spans. This helps T5 learn robust text generation capabilities, applicable to a wide range of NLP tasks.

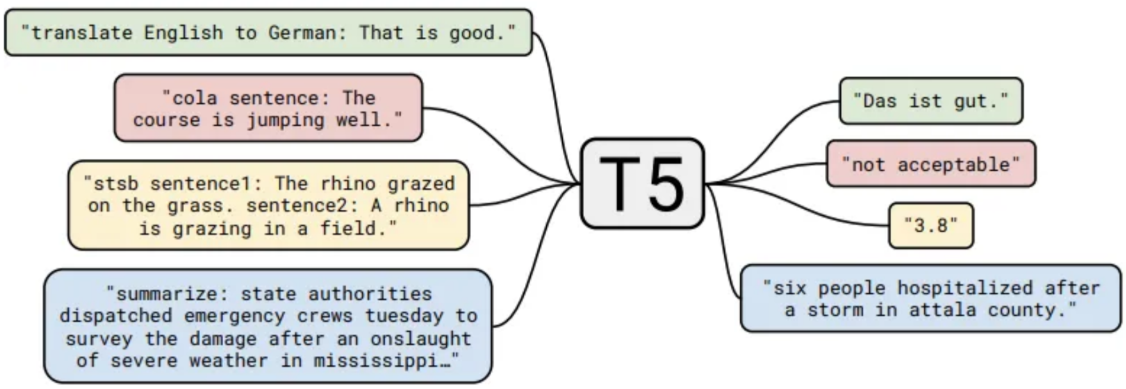

Figure 3: T5 Model Pipeline: T5's architecture handles a variety of tasks by framing each one as a text generation problem. The input text is encoded, and the decoder generates the appropriate output, whether it be translation, summarization, or classification.

Applications of T5 in NLP:

T5's architecture allows it to excel in tasks that require input-output transformations, such as:

- **Text Summarization:** T5 is highly effective in generating summaries from long documents. Its sequence-to-sequence structure ensures that it can capture the most important information from the input text and generate concise, informative summaries.

- **Translation:** T5 can handle machine translation tasks by treating them as text generation problems. By conditioning the output on the input language, T5 can translate between different languages.

- **Text Classification:** Although T5 is inherently a generative model, it can still perform classification tasks by generating a label or score based on the input text. For instance, T5 can be used for sentiment analysis by generating labels such as "positive" or "negative".

- **Sentence Similarity:** T5 can be employed to assess sentence similarity by generating scores or labels for pairs of sentences, making it useful in applications such as semantic textual similarity.

In the context of AI-Generated Content (AIGC) data analysis, T5 can be leveraged to generate concise summaries of user discussions, provide translations for multilingual communities, and perform text classification to identify key trends. Its flexible, unified architecture enables it to adapt to various scenarios, making it a versatile tool for analyzing user-generated content.

2.2.4 LaMDA (Language Model for Dialogue Applications)

LaMDA, introduced by Thoppilan et al. [33], is a large language model specifically designed for dialogue and conversation-based applications. Unlike models such as BERT or GPT, which are trained primarily on general NLP tasks like text classification or summarization, LaMDA is tailored to generate coherent and contextually relevant responses in multi-turn conversations. The key feature that differentiates LaMDA from other models is its ability to maintain the context of a conversation over multiple dialogue turns, ensuring that its responses remain appropriate and contextually aligned with the flow of the conversation.

LaMDA is built on a transformer architecture but incorporates several optimizations designed specifically for dialogue generation. These optimizations enable the model to track conversational history and respond meaningfully to prompts. The training of LaMDA involves fine-tuning on datasets composed of dialogue data from diverse conversational domains. This allows the model to better capture the nuances of natural conversations, including context-switching, topic continuity, and handling ambiguous or open-ended questions.

Key Components of LaMDA's Architecture:

The following are the key architectural and training features that enable LaMDA to excel at dialogue generation:

- **Transformer-based Architecture:** LaMDA is fundamentally based on the transformer architecture, which has proven highly effective in capturing long-range dependencies in text. Like other transformer-based models, LaMDA employs multi-head self-attention mechanisms to allow the model to attend to various parts of the input sequence simultaneously. This architecture is well-suited for understanding and generating linguistic structures.

- **Pretraining with Conversational Data:** Unlike other general-purpose language models, LaMDA is pretrained on large-scale conversational datasets that encompass diverse dialogue contexts. These datasets include dialogues from various domains such as customer support, social media interactions, and casual conversations. Pretraining on such conversational data allows LaMDA to develop a deeper understanding of how dialogues unfold and to generate responses that are relevant to the context.

- **Next-Turn Prediction as a Pretraining Objective:** One of the core pretraining tasks for LaMDA is next-turn prediction. In this setup, the model is trained to predict the next turn in a conversation based on the entire dialogue history. This enables LaMDA to handle conversational nuances such as topic shifts, follow-up questions, and responses that reference earlier parts of the dialogue. The next-turn prediction objective allows LaMDA to generate more contextually coherent and engaging dialogue responses, particularly in multi-turn interactions.

- **Handling Context Shifts and Open-Ended Questions:** A key challenge in dialogue modeling is handling context shifts and open-ended prompts. LaMDA incorporates mechanisms to address these challenges by maintaining a memory of the dialogue history and utilizing this context to inform its responses. This ensures that LaMDA's responses remain coherent even when the conversation moves across different topics or when the user poses ambiguous questions.

- **Fine-Tuning for Safety and Specificity:** LaMDA is fine-tuned not only to generate relevant dialogue responses but also to avoid unsafe or harmful responses. Fine-tuning on datasets that include diverse conversational behaviors ensures that LaMDA generates responses that are not only contextually appropriate but also safe and aligned with conversational norms. Additionally, LaMDA is fine-tuned to handle specificity, providing answers that are relevant to the question asked without deviating from the subject.

Applications of LaMDA in Dialogue Analysis:

LaMDA is particularly effective in generating dialogue-based outputs, making it suitable for a variety of conversational applications:

- **Multi-turn Conversations:** LaMDA is designed to excel in multi-turn dialogues where context needs to be tracked over several conversation rounds. This makes it highly suitable for applications such as customer support, virtual assistants, and chatbots, where the conversation can span multiple questions and answers.

- **Open-Domain Dialogue Generation:** Unlike task-specific models, LaMDA can generate relevant and contextually appropriate responses across a wide variety of conversational topics. This flexibility is particularly valuable in applications where user interactions may not follow a pre-defined structure.

- **Social Media and AI-Generated Content (AIGC) Analysis:** In the context of AIGC, LaMDA can be applied to analyze user discussions and extract meaningful insights from conversational data. It is capable of understanding social media interactions, where dialogue context may shift dynamically, making it well-suited for social platforms like Discord or Twitter.

- **Human-Computer Interaction (HCI):** LaMDA can enhance human-computer interactions by providing conversational agents that understand and respond to users in a natural, context-aware manner. This could improve user experiences in a wide range of domains, including healthcare, education, and entertainment.

LaMDA's ability to maintain context across multiple turns, generate coherent responses, and handle diverse conversational scenarios makes it a powerful tool for dialogue

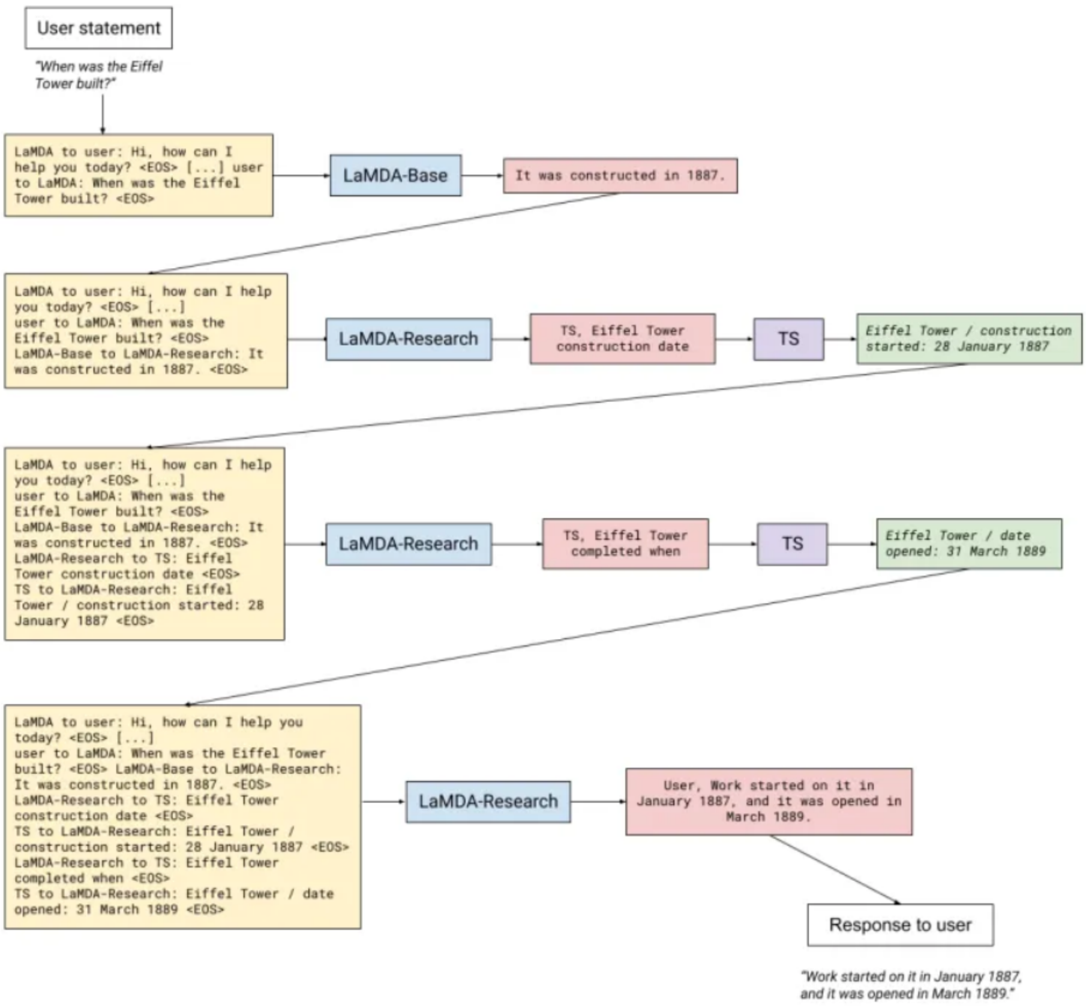

Figure 4: LaMDA Architecture for Dialogue: The model is optimized for maintaining context across multiple turns in a conversation. It is pretrained on diverse dialogue datasets and incorporates next-turn prediction to generate contextually coherent responses.

analysis and generation in real-world applications. Its specialized architecture and training approach ensure that it can handle the human dialogue, from casual conversations to more structured interactions.

2.2.5 GPT (Generative Pre-trained Transformer)

The GPT series, developed by OpenAI, represents a major leap forward in natural language processing through its use of transformer-based architectures for language generation. The first version, GPT-1, introduced by Radford et al. [27], was a generative model that demonstrated the potential of transformers in text generation tasks. GPT-2 [28], an expanded and improved version, scaled up the model size and showcased the power of unsupervised pretraining on massive datasets.

GPT-1 (Improving Language Understanding by Generative Pre-Training)

GPT-1 is built using the transformer decoder architecture and is trained autoregressively, meaning that it generates text one token at a time based on previous tokens in the sequence. The model uses multi-head self-attention mechanisms and feed-forward networks to capture long-range dependencies in the input text. The key innovation in GPT-1 was demonstrating that pretraining on a large corpus of unlabelled text, followed by fine-tuning on specific downstream tasks, could yield impressive results on a variety of NLP tasks.

The GPT-1 architecture includes 12 transformer layers, each composed of multi-head self-attention, a residual connection, and a feed-forward neural network. During training, the model learns to predict the next word in a sequence, leveraging the entire left-hand context, making it a unidirectional model (left-to-right).

- **Pretraining Objective:** GPT-1 uses a simple autoregressive pretraining objective where it maximizes the likelihood of the next word in a sequence given all previous words.

- **Fine-tuning:** After pretraining, GPT-1 is fine-tuned on specific supervised datasets for downstream tasks like text classification, question answering, and summarization.

GPT-2 (Language Models Are Unsupervised Multitask Learners)

Building on the success of GPT-1, GPT-2 was introduced as a much larger model with 1.5 billion parameters, demonstrating the benefits of scaling model size. GPT-2 follows the same autoregressive architecture but is pretrained on a much larger dataset, allowing it to generalize better across a wide range of tasks without fine-tuning. One of GPT-2's key contributions is its ability to perform various NLP tasks in a zero-shot setting, meaning it can handle tasks without any task-specific training.

Key innovations in GPT-2 include:

- **Model Scaling:** GPT-2 contains 48 layers and 1.5 billion parameters, showing a dramatic increase in the model's capacity to learn and generate coherent text.

- **Unsupervised Pretraining:** GPT-2 was trained on the WebText dataset, which consists of diverse web-based text sources, allowing it to capture a wide range of language patterns, contexts, and topics.

- **Zero-shot Learning:** One of the key achievements of GPT-2 is its ability to generalize to new tasks without task-specific fine-tuning, performing well across tasks like summarization, translation, and question answering.

The autoregressive nature of GPT-2 means that the model generates one token at a time, attending to all previously generated tokens. This enables the model to generate highly coherent text that maintains context across long sequences, making it suitable for tasks like story generation, dialogue modeling, and even code generation.

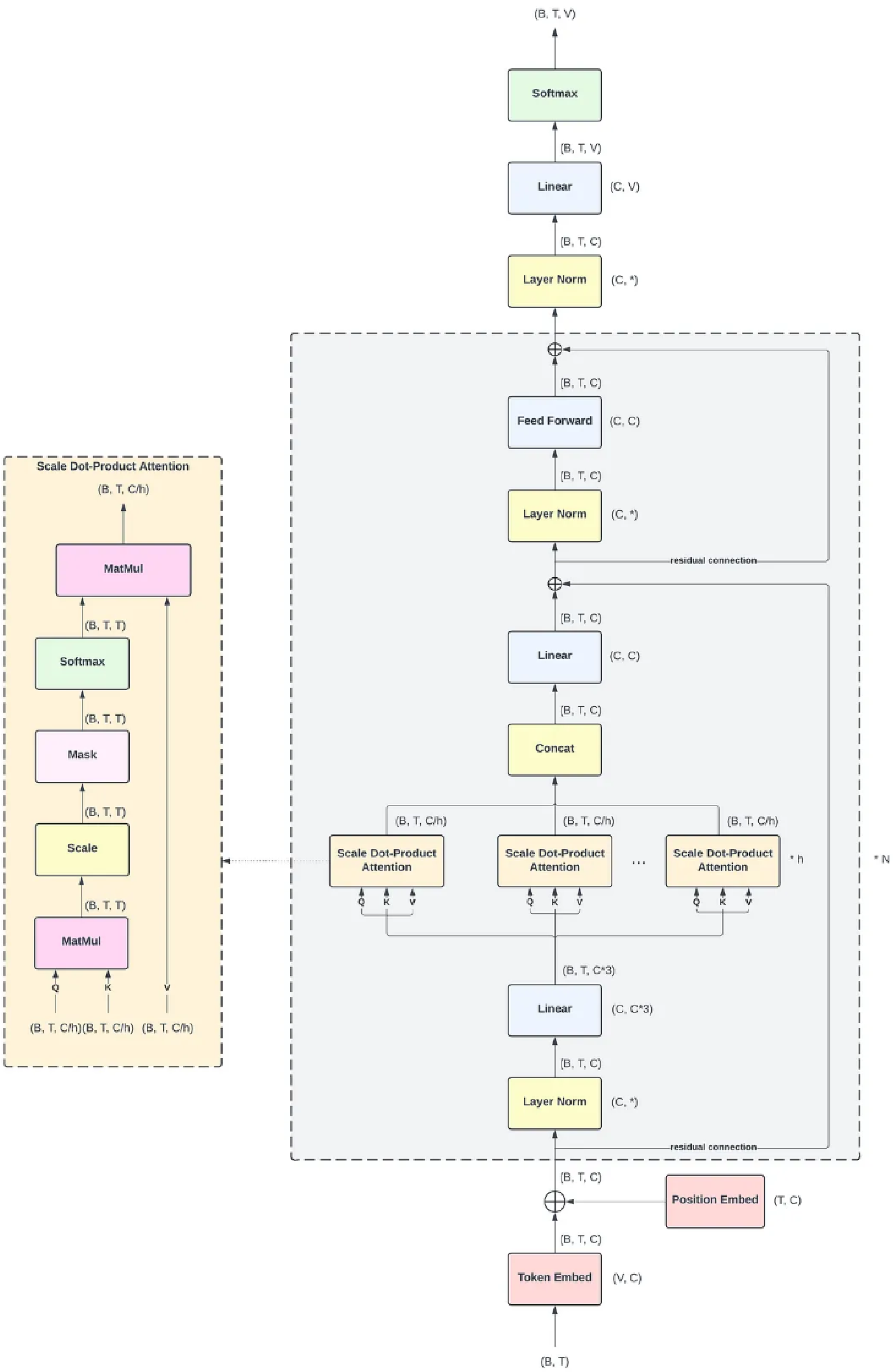

Figure 5: GPT-2 Model Architecture: The model is a transformer decoder that generates text autoregressively, predicting one token at a time based on previous tokens. The multi-head self-attention mechanism captures long-range dependencies in the input sequence.

Key Differences Between GPT-1 and GPT-2:

- **Model Size:** GPT-1 has 117 million parameters, while GPT-2 scales up to 1.5 billion parameters, showing how model scaling leads to better language generation capabilities.

- **Training Corpus:** GPT-2 is trained on a much larger dataset (the WebText corpus), which contains over 40 GB of text data, compared to the smaller dataset used for GPT-1.

- **Zero-shot Learning:** GPT-1 required task-specific fine-tuning for downstream tasks, while GPT-2 is capable of zero-shot learning, handling various tasks without additional fine-tuning.

Applications of GPT-2 in NLP:
GPT-2's architecture and capabilities make it particularly useful for a wide range of NLP tasks, including:

- **Text Generation:** GPT-2 is capable of generating highly coherent and contextually relevant text over long sequences, making it suitable for creative writing, story generation, and dialogue systems.

- **Summarization:** GPT-2 can be used to generate concise summaries of longer texts by conditioning the output on the input context.

- **Translation:** Without any task-specific training, GPT-2 can perform machine translation by conditioning on an input sentence in one language and generating the corresponding sentence in another language.

- **Question Answering:** GPT-2 can generate answers to questions by attending to the input context and generating a relevant response.

GPT's Relevance to AIGC Data Analysis:
In the context of AI-Generated Content (AIGC) data analysis, GPT models (especially GPT-2) can be leveraged to generate insights from user discussions, summarize long conversations, and predict trends based on the text. The ability of GPT-2 to maintain coherence over long text sequences makes it especially useful for analyzing user-generated content on platforms such as social media, forums, and streaming platforms.

GPT's transformer-based architecture, combined with its ability to generalize across multiple tasks, enables it to provide meaningful insights and predictions in AIGC contexts. By analyzing conversational patterns, summarizing discussions, or even predicting the flow of conversations, GPT-based models offer valuable tools for content analysis.

2.2.6 NLP for Social Media Analysis

Natural language processing (NLP) has become an essential tool for analyzing social media platforms, enabling researchers to extract insights from user-generated content, identify trends, and study patterns in online discourse. Social media platforms such as Discord and Twitch generate massive amounts of unstructured data, making it challenging to manually process and extract meaningful insights. NLP techniques facilitate the automatic processing of these conversations, identifying key themes and trends in real-time discussions. In recent years, large language models like ChatGPT-3.5 and ChatGPT-4 have gained prominence in tasks such as conversation summarization, sentiment analysis, and topic detection, enhancing our ability to understand online interactions at scale.

This section examines the role of advanced NLP models, particularly ChatGPT, in analyzing social media conversations. It explores the key methodologies employed, such as the use of transformers for generating human-like text, and the deployment of models for sentiment analysis to gauge public opinion on various subjects. Moreover, the integration of NLP into social media analytics enables the study of complex phenomena like user engagement, community dynamics, and the spread of information across platforms.

In addition to these advanced models, traditional NLP methods like topic modeling also play a crucial role in social media analysis. Topic modeling techniques, such as Latent Dirichlet Allocation (LDA) and Gibbs Sampling Dirichlet Mixture Model (GSDMM), help uncover the latent thematic structures in user-generated content by clustering related words and identifying recurring discussion topics. This section delves deeper into the application of these topic modeling techniques in the context of social media, examining how they can be used to categorize conversations into meaningful themes and compare the dynamics of discussions across platforms like Discord and Twitch.

Topic Modeling in Social Media Analysis

Topic modeling is a powerful tool in Natural Language Processing (NLP) that aims to discover the underlying thematic structures within large textual datasets. In contrast to sentiment analysis, which identifies emotional tones, topic modeling clusters related words based on their co-occurrence, revealing latent topics. This approach is particularly valuable in social media analysis, where user conversations are highly unstructured, fragmented, and diverse [4, 1]. Given the real-time, spontaneous nature of social media interactions, manual content analysis becomes impractical, and topic modeling offers a scalable solution to categorize these discussions into interpretable themes.

Latent Dirichlet Allocation (LDA), introduced by Blei et al. [4], has been a foundational technique for topic extraction. It assumes that documents consist of multiple topics, and each topic is a distribution of words. LDA has been widely applied in social media research, enabling the detection of thematic trends in user discussions [37, 12]. However, LDA's efficacy diminishes in short-text scenarios, as common in social media, where conversations tend to be brief and focused on singular topics.

To address the challenges posed by short-form content, alternative models like Gibbs Sampling Dirichlet Mixture Model (GSDMM) have been developed. GSDMM assumes that each document (e.g., a short message) pertains to a single topic, which makes it especially effective for analyzing brief, high-frequency user interactions [39, 25]. Both models offer distinct advantages depending on the nature of the data, and their application in this study provides insights into the evolving discussions on AI-generated content in social media.

Latent Dirichlet Allocation (LDA)

LDA is one of the most commonly used probabilistic models for discovering topics in large corpora. It operates under the assumption that documents are mixtures of topics, where each topic is a mixture of words [4]. LDA performs well in identifying multiple themes within longer documents and has been employed across various domains, including AI-generated content analysis, social media interactions, and creative industries [12, 37].

Despite its widespread application, LDA faces limitations when dealing with short-form text common in social media platforms. The model's assumption that each document contains multiple topics does not always hold in environments where messages are brief and concentrated on a single subject. Additionally, the probabilistic nature of LDA can lead to topic overlap, which complicates the interpretation of results in datasets where conversations are focused on distinct issues.

Gibbs Sampling Dirichlet Mixture Model (GSDMM)

GSDMM offers an alternative to LDA for scenarios involving short-form, real-time conversations. Developed by Yin and Wang [39], GSDMM assumes that each document is associated with a single topic, making it better suited for analyzing short texts, such as social media posts. By using Gibbs Sampling, GSDMM iteratively refines topic assignments to maximize the coherence of the clusters. This model is particularly effective in environments where discussions are brief and focused, allowing for a clearer delineation of topics.

GSDMM has demonstrated success in producing more precise clusters of thematic content in social media analysis. Its ability to handle short, fragmented messages makes it an ideal model for applications where user conversations revolve around singular themes, such as technical discussions or specific user interactions in online communities [25]. The use of GSDMM in this study allows for more granular insights into discussions surrounding AI-generated content, which are often fragmented yet focused.

Comparing LDA and GSDMM for Social Media Analysis

The application of LDA and GSDMM highlights the differences in how each model handles social media data. LDA's strength lies in its ability to identify multiple topics within longer documents, making it suitable for more comprehensive analyses [4]. However, in the context of social media, where conversations are often brief and focused, GSDMM's single-topic assumption provides a more accurate representation of the data. Studies have shown that GSDMM outperforms LDA in scenarios with short-form content, producing clearer and more interpretable topics [39, 25].

For the purposes of this research, both models were employed to extract thematic trends from social media conversations about AI-generated content. While LDA was useful for identifying broader thematic trends, GSDMM proved more effective in capturing the nuanced, short-form discussions typical of social media interactions. Therefore, we cannot simply conclude which of these two models is more suitable for social media analysis. Further research is needed in combination with specific circumstances.

Insights from Topic Modeling in Social Media

Topic modeling has been widely used to explore the dynamics of social media discussions, revealing key themes, shifts in user interests, and emerging trends [1]. By applying LDA and GSDMM, this research offers valuable insights into the nature of discussions surrounding AI-generated content on social media platforms. The models revealed a progression in user concerns, from general discussions on AI tools to more specialized topics such as ethical considerations and creative applications. The findings also suggest that domain-specific stop words improve topic coherence, as shown in studies by Raffel et al. [29].

The use of topic modeling in this study contributes to the broader understanding of how AI-generated content is perceived and discussed in online communities. Future research can further refine these methods by incorporating additional models and extending the temporal analysis to capture more granular shifts in conversation dynamics.

ChatGPT for Social Media Interaction Analysis

ChatGPT, an advanced variant of OpenAI's GPT models, is particularly well-suited for handling the dynamic and often fragmented nature of social media conversations. Unlike traditional models focused on static text, ChatGPT is designed to maintain coherence across multiple dialogue turns, which makes it ideal for real-time interaction analysis on platforms like Twitter, Reddit, and Discord.

The architecture of ChatGPT is based on the transformer model introduced by Vaswani et al. [35], with specific enhancements to handle conversational data more effectively. The model uses reinforcement learning with human feedback (RLHF) [23] to improve its ability to maintain context over extended conversations, ensuring that it can respond appropriately even as discussions shift topics or grow in complexity.

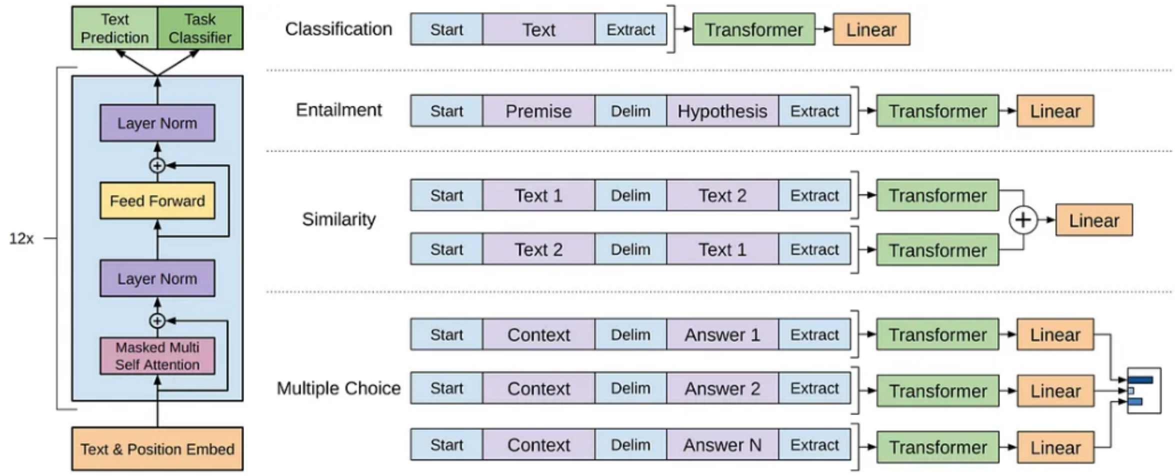

Figure 6: ChatGPT Model Architecture for Social Media Analysis. The architecture demonstrates how text inputs, including social media posts, are processed through multiple transformer layers and fine-tuned for conversational coherence.

Applications of ChatGPT-3.5 and ChatGPT-4 in Social Media Analysis

ChatGPT-3.5 and ChatGPT-4 have been applied in several studies focused on analyzing social media data. These models have been used for tasks such as:

- **Conversation Summarization:** ChatGPT is capable of generating concise summaries of long-running discussions by understanding the context across multiple posts. This makes it valuable for summarizing extended threads on platforms like Reddit or Twitter, where discussions can span numerous interactions.
- **Sentiment Analysis:** By fine-tuning ChatGPT with labeled sentiment data, researchers have used the model to classify the tone of social media posts as positive, negative, or neutral. This process has been particularly useful in tracking public sentiment over time during major events or trending discussions.
- **Topic Detection:** ChatGPT can identify emerging themes in social media conversations by analyzing the text and detecting patterns or recurring subjects. This feature enables researchers to monitor shifts in public interest and detect new trends as they emerge.

Methodologies for Social Media Analysis Using ChatGPT

The general approach to using ChatGPT for social media analysis involves several key steps:

- **Data Collection:** Large datasets are gathered from social media platforms such as Twitter, Reddit, or Discord, focusing on posts relevant to the research topic.

- **Pre-processing:** The data is cleaned to remove noise, such as spam or duplicate posts. This step ensures that the model processes meaningful content.

- **Model Fine-tuning:** ChatGPT is fine-tuned using relevant datasets for the specific task, such as conversation summarization, sentiment analysis, or topic detection. Pre-labeled data can be used to enhance performance in tasks like sentiment classification.

- **Analysis and Output:** The model generates outputs based on the task at hand. For instance, in sentiment analysis, ChatGPT classifies posts based on sentiment, while in conversation summarization, it condenses large conversations into key points.

Case Studies and Research Findings

In a study by Xie et al. [38], ChatGPT was used to analyze Twitter discussions surrounding a global event. By applying ChatGPT's conversation summarization capabilities, the researchers were able to extract the key issues discussed by users and identify the main themes over time. The model's ability to maintain coherence across multiple posts proved particularly valuable in summarizing discussions that spanned hundreds of tweets.

Another study by Kim et al. [17] applied ChatGPT to analyze sentiment within Reddit discussions. The goal was to understand how public sentiment evolved during a specific event and how ChatGPT could dynamically track these changes. The model was fine-tuned on sentiment-labeled data and used to classify Reddit posts into positive, negative, or neutral categories. Over time, the model detected shifts in sentiment in response to major updates or announcements, providing insights into how the event influenced user opinion.

The studies discussed highlight the utility of NLP models like ChatGPT in social media analysis. ChatGPT's ability to manage large datasets, generate summaries, and detect trends in real-time interactions makes it a valuable tool for understanding online discourse. Moving forward, researchers can explore integrating ChatGPT into more complex frameworks, such as combining it with network analysis to study the influence of certain users or topics within social networks. Additionally, the use of ChatGPT in predicting user behavior and modeling future conversations presents an exciting avenue for further exploration in social media analysis.

2.3 Social Media Analysis

2.3.1 Social Media Analysis on TTIG

This section delves into the methodological perspectives and community impacts of Text-to-Image Generation (TTIG) technologies based on comprehensive studies, particularly those conducted by Fayya Anyatasia [2] and Jonas Oppenlaender et al. [21, 22]. Their work provides critical insights into the evolving role of TTIG AI in creative workflows and its broader societal implications, offering a structured view of how social media platforms facilitate discourse surrounding these technologies.

Summary

Fayya Anyatasia's thesis [2] investigates the interplay between creative practitioners and Text-to-Image Generation (TTIG) AI technologies. Through a rigorous mixed-methods approach, combining the ethnography of 331 Reddit posts and comments with a survey of 92 creative professionals, Anastasia explores how TTIG AI is embraced and resisted within creative workflows. Her research illuminates how these technologies impact creative empowerment and ethical considerations, revealing a spectrum of integration techniques and attitudes within the creative community. Anyatasia's findings underscore the need for ethical deliberation in deploying AI tools in creative practices, providing foundational insight for future explorations in AI-generated content.

Methodology and Findings of TTIG AI in Creative Practices

Fayya Anyatasia's comprehensive study employs a mixed-method approach to dissect the multifaceted influence of Text-to-Image Generation (TTIG) AI technologies on creative practices. This methodology intertwines qualitative and quantitative analyses to illuminate creative practitioners' nuanced engagements and diverse perspectives toward AI integration.

Fayya Anyatasia's investigation into integrating Text-to-Image Generation (TTIG) AI in creative practices is grounded in a mixed-methods approach. It combines netnography and an online questionnaire to explore qualitative and quantitative dimensions of AI utilization among creative practitioners.

Netnography The netnographic component analyzed 331 Reddit posts and comments from specific subreddits relevant to art and AI technologies. This process was facilitated by the Python Reddit API Wrapper (PRAW), which streamlined data collection. The analysis phase encompassed sentiment analysis to gauge general attitudes towards TTIG AI within the art community and inductive thematic analysis to identify recurrent themes and insights about AI's role and impact in creative work. The thematic analysis was iterative, beginning with generating initial codes that were progressively refined and clustered into coherent themes reflecting the sentiments and perspectives of the creative community on Reddit.

Online Questionnaire The survey component was designed to complement and deepen the qualitative insights from the netnography. It featured the User Motivation Inventory (UMI) and a series of open-ended questions to elicit detailed responses about the motivations, workflows, and ethical considerations of creative practitioners concerning TTIG AI. The UMI, comprising 18 questions, was employed to quantitatively assess various dimensions of user motivation in engaging with TTIG AI. The questionnaire was distributed through the author's networks and social media channels, targeting a broad spectrum of creative practitioners familiar with TTIG AI tools such as Midjourney, DALL-E 2, and Stable Diffusion.

Participants The study engaged 92 creative practitioners from diverse disciplines, including but not limited to illustration, photography, game design, UX design, and art. This heterogeneous group was selected to ensure a comprehensive understanding of the adoption and impact of TTIG AI across different creative fields. Participants were not limited to professional artists, as the study also sought to capture the perspectives of amateurs and art students, acknowledging that experience levels and familiarity with AI could influence perceptions and usage patterns of TTIG AI technologies.

Data Analysis For the qualitative data from both Reddit and the open-ended survey questions, inductive thematic analysis was the primary method employed, allowing for the emergence of themes directly from the data. This approach was chosen for its flexibility and suitability for exploratory research, where preconceived categories are not imposed on the data. The quantitative data from the UMI responses was analyzed to identify patterns of motivation among users, facilitating a nuanced understanding of why creative practitioners engage with TTIG AI.

Fayya Anyatasia's research meticulously catalogs the impact of Text-to-Image Generation (TTIG) AI across various creative practices, unearthing detailed insights into the workflows and attitudes of creative practitioners towards AI. Through an extensive analysis of 331 Reddit posts and comments, complemented by a survey of 92 creative practitioners, Anyatasia provides an in-depth examination of the nuanced ways in which AI is integrated into creative processes.

Creative Workflows and Practices The study identifies a range of workflows within the creative industries, showcasing the versatility of TTIG AI applications. From digital artists to game designers, practitioners employ TTIG AI for a variety of purposes, including but not limited to inspiration generation, rapid prototyping, and enhancing creative workflows. As reported, the common uses of TTIG AI include generating initial sketches, refining compositions, and producing detailed artwork elements that complement the artists' background creation and character design skills.

Differences Among Creative Disciplines Variances were observed in the adoption and application of TTIG AI among different creative practices. Illustrators and digital artists, for instance, frequently use AI to generate concept art and speed up the design process. At the same time, photographers explore AI to create compositions or edit elements within their work. Game designers and UX designers leverage AI to prototype interfaces and environments rapidly. Notably, traditional artists also engage with AI, using it to explore new forms of expression and integrate digital elements into their work.

Contrast Between Survey Results and Reddit Posts Anastasia's analysis reveals discrepancies between the survey responses and Reddit discussions. While Reddit posts often highlighted enthusiastic experimentation with TTIG AI and a focus on the potential benefits, survey responses tended to provide a more measured perspective, acknowledging both the opportunities and challenges posed by TTIG AI. This contrast suggests a broader spectrum of opinions and experiences within the creative community, ranging from optimistic adoption to cautious skepticism.

Methodological Insights The methodology section of the thesis details the dual approach of sentiment analysis and thematic analysis employed to dissect the Reddit data. Sentiment analysis categorized posts into positive, negative, and neutral sentiments, revealing a landscape of opinions on TTIG AI. The thematic analysis further broke down these sentiments into specific themes, such as "AI as a beneficial tool," "ethical concerns," and "AI's impact on creativity," offering a nuanced understanding of the discourse surrounding AI in creative fields.

Survey Design and Participant Demographics The online questionnaire was designed to delve deeper into creative practitioners' motivations, workflows, and ethical considerations using TTIG AI. Questions ranged from demographic information to detailed inquiries about the use and perception of TTIG AI. Participants hail from various creative industries, including illustration, photography, game design, and more, providing a rich dataset for analysis.

This in-depth investigation into the adoption and impact of TTIG AI in creative practices sheds light on the evolving relationship between artists and technology. By articulating the diverse methodologies and outcomes of AI integration, Anyatasia's work contributes valuable insights into the future of AI in creative domains, advocating for a balanced approach that nurtures innovation while conscientiously addressing ethical implications.

In summary, Fayya Anyatasia's thesis not only captures the current landscape of TTIG AI in creative practices but also paves the way for future research, emphasizing the importance of understanding and supporting the creative community in the age of AI.

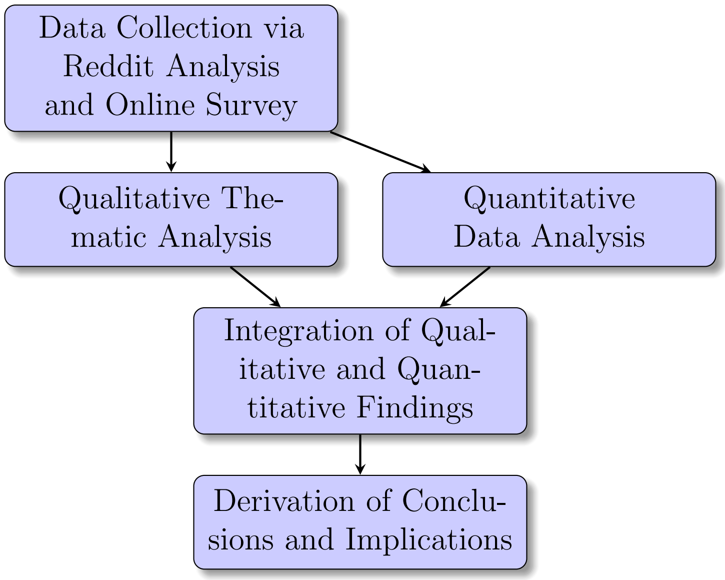

Figure 7: Fayya Anyatasia's Research Methodology Pipeline

2.3.2 Integration of Generative AI into Creative Practices

Recent investigations into generative AI, particularly within creative domains, have illuminated a broad spectrum of public perceptions and applications of text-to-image (TTIG) technologies. Oppenlaender et al.'s research endeavors, encapsulated in their seminal works [21] and [22], provide an extensive exploration of these dynamics, offering a granular understanding of the evolving interface between AI technologies and creative practices.

In *Perceptions and Realities of Text-to-Image Generation*, Oppenlaender et al. [21] adopt a survey-based methodology to scrutinize individuals' perceptions towards emergent TTIG technology. Their research pipeline begins with disseminating a comprehensive questionnaire to gauge participants' familiarity with and attitudes toward TTIG technologies. The study identifies a bifurcation in the perceived future significance of TTIG technologies, contingent upon the respondents' direct experience with these tools. This finding intimates a broader societal and individual underestimation of the implications of generative AI advancements, urging a recalibration of our collective foresight in technology's trajectory.

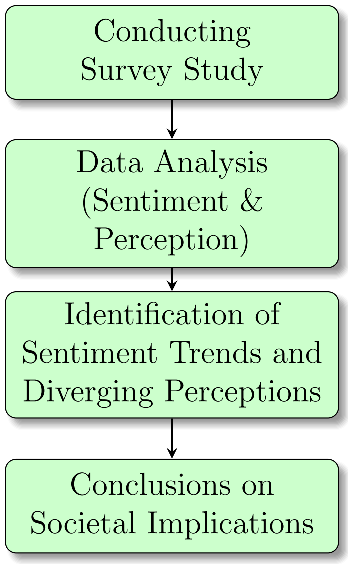

Figure 8: Research Pipeline for "Perceptions and Realities of Text-to-Image Generation"

A Taxonomy of Prompt Modifiers for Text-To-Image Generation, conversely, employs an ethnographic lens to dissect the motivations, challenges, and patterns of engagement among TTIG practitioners [22]. Through an innovative mix of participant observation and analysis of user-generated content, Oppenlaender et al. constructs a taxonomy of prompt modifiers, elucidating the strategic adjustments users employ to optimize their interactions with TTIG generators. This exploration reveals the diverse socio-demographic profiles of TTIG users, their varied motivations for engaging with these technologies, and the regularity of their practice. An output of this study is the articulation of guidelines to refine user experiences with TTIG generators, thereby enhancing the creative utility and accessibility of these AI tools.

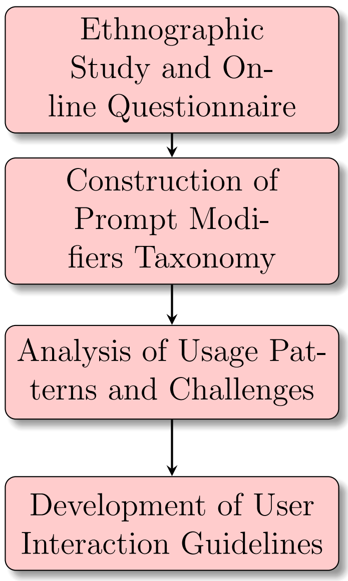

Figure 9: Research Pipeline for "A Taxonomy of Prompt Modifiers for Text-To-Image Generation"

These studies underscore the multifaceted nature of TTIG AI's integration within creative workflows, spotlighting the intricate balance between innovation and ethical stewardship in the digital age. The methodologies employed by Oppenlaender et al.—ranging from survey-based approaches to ethnographic investigations—offer a rich, multi-dimensional perspective on the adoption and impact of generative AI technologies. As we navigate the future of artistic endeavors, the insights from these studies illuminate the path toward fostering a harmonious synergy between human creativity and artificial intelligence while navigating the ethical quandaries and societal implications accompanying such advancements.

Inspirational Findings for Our Research Our research builds upon a comprehensive review of the literature on text-to-image generative AI (TTIG AI), drawing insights from the studies conducted by Oppenlaender et al. [21, 22] and Anyatasia [2]. The focal points of our exploration include:

- **Sentiment Trends:** We aim to understand the impact of TTIG AI on creativity, ethical concerns, and societal perceptions through the lens of sentiment trends observed within digital communities. This includes leveraging large-scale sentiment data to discern AI's nuanced roles in creative practices.
- **Adaptability of AI Tools:** Assessing how AI tools are integrated into content creation across various platforms highlights potential trends in user adaptation and technology acceptance.
- **Ethical and User Attitude Analysis:** A detailed examination of community perspectives on the ethical implications of AI technologies forms a critical part of our analysis.

Instead of conducting new qualitative research, our study will compare Anyatasia's and Oppenlaender et al.'s findings, aiming to synthesize generalizable insights from this comparison. This approach will help us understand the broader sentiment dynamics and user engagement within AI-generated content communities.

Methodological Reflections

Given the constraints on time and the specialized nature of qualitative studies, our research adopts a focused approach that leverages existing studies for comparative analysis. Drawing on the comprehensive work of Anyatasia [2] and the detailed investigations by Oppenlaender et al. [21, 22], we intend to:

- Analyze and compare the results of both studies to extract generalizable insights relevant to our research focus.
- Employ advanced Natural Language Processing (NLP) techniques, like the BERT model, for sentiment analysis on textual data from platforms like Twitch and Discord. The comparative findings will augment this quantitative analysis to better understand community sentiments toward TTIG AI.

This methodology allows for a deep dive into the sentiment analysis of AI-generated content without requiring new qualitative data collection, addressing the practical challenges of conducting ethnographic studies within the project's timeframe. By comparing and integrating the results of the existing studies, our research will provide a solid foundation for understanding the relationship between AI technologies and creative communities, setting a direction for future research that could include qualitative analyses as a valuable next step.

2.3.3 Social Media Analysis on AI-generated Video in Streaming Communities

The advent of artificial intelligence (AI) has significantly transformed content creation and user engagement within streaming communities, particularly on platforms like Twitch and Discord. These platforms serve as valuable environments for examining how AI-generated content, especially video and art, is integrated into social media spaces. This section reviews existing research that highlights methodologies and findings on AI-generated video and its implications for social media analysis, focusing on how platforms like Twitch and Discord shape user engagement, content creation, and community dynamics.

Social Media Platforms: Twitch and Discord

Twitch and Discord are among the most prominent platforms for social media research, especially in analyzing interactions involving AI-generated content. These platforms differ in their structure and user behavior, providing unique environments for content creation and community engagement.

Twitch: Real-Time User Engagement

Twitch, as a live-streaming platform, allows users to broadcast content in real-time while viewers engage with streamers through chat. This immediacy fosters a dynamic environment where content evolves rapidly in response to audience input. Hamilton et al. [14] describe how parasocial relationships develop between viewers and streamers, where users feel a sense of connection despite the one-sided nature of these interactions. Such relationships can amplify viewer engagement, especially when streamers use AI-generated content such as deepfake avatars or procedurally generated media. The study by Böhmer et al. [7] further examines how Twitch chat functions as an interactive element of live streams, shaping content in real time based on audience feedback.

For example, streamers may employ AI tools to generate real-time responses or to create interactive elements within their streams, enhancing audience participation. This dynamic is further explored in the work of Stein [32], who investigates how the integration of AI into social media platforms shifts interaction paradigms by allowing content creators to automate parts of their streams while maintaining a personalized connection with their audience. These developments raise questions about the boundaries between human-generated and AI-generated content, as audiences may struggle to distinguish between the two.

Discord: Community-driven Discussions and AI Content Creation

Discord offers a different mode of interaction, emphasizing asynchronous communication within interest-specific channels. Users on Discord can engage in detailed conversations that unfold over time, making the platform ideal for discussing technical and ethical aspects of AI-generated content. Research by Koo et al. [18] shows how Discord's structure supports in-depth discussions, with channels often dedicated to specific topics such as AI art or video generation. This platform encourages users to share resources, collaborate on projects, and engage in long-form dialogue about the implications of generative AI technologies.

Oppenlaender's research [22] explores how AI art communities on Discord, facilitate discussions about the use of generative AI tools. In these communities, users debate the artistic merit of AI-generated works and share techniques for optimizing generative models like DALL·E and Stable Diffusion. Oppenlaender argues that platforms like Discord foster a hybrid form of content creation, where users and algorithms collaborate to produce innovative media.

The contrast between Twitch and Discord's engagement structures is critical for understanding how AI-generated content is produced, consumed, and discussed. While Twitch encourages immediate, emotional interaction, Discord provides a space for reflection and technical discussion, particularly regarding the ethical and creative implications of AI-generated video and art.

AI-Generated Content in Streaming Communities

AI-generated content, including videos, deepfakes, and AI-enhanced avatars, is becoming increasingly prevalent on platforms like Twitch and Discord. The rise of text-to-image models such as DALL·E [30] and video generation technologies has blurred the line between user-generated and AI-generated content. Anyatasia [2] discusses how AI tools enable creators to expand their creative capacities, while also introducing new ethical and legal challenges, particularly in terms of ownership and authenticity.

AI-generated Video on Twitch

The use of AI-generated video on Twitch is particularly relevant as streamers experiment with AI tools to enhance their broadcasts. For instance, AI-driven avatars and deepfake technologies are employed to automate aspects of streaming, from avatar manipulation to content generation. Ching and Li [9] investigate the implications of deepfake technology in live-streaming environments, noting that these tools have the potential to enhance user engagement by creating more dynamic and interactive content. However, they also raise concerns about the authenticity of such content and the risks of misinformation, as viewers may be unable to distinguish between real and AI-generated elements. Naim and Sadeghi [20] argue that the proliferation of AI-generated video could undermine trust in social media content, particularly if deepfakes are used maliciously.

Research by Böhmer et al. [7] suggests that while viewers on Twitch are generally aware of the presence of AI tools in content creation, they may still be susceptible to the emotional effects of parasocial relationships. The real-time nature of Twitch's interaction model means that AI-generated content can be seamlessly integrated into streams without disrupting the viewer's sense of engagement. However, the ethical implications of this integration remain a topic of debate.

AI-Generated Content on Discord

On Discord, AI-generated content tends to be discussed in a more technical and collaborative context. Users in AI-focused channels share their experiences with generative tools, exchange tips on model optimization, and explore the creative possibilities of machine learning technologies. Oppenlaender [21] highlights how AI art communities on Discord provide a space for users to engage in critical discussions about the future of creative industries, with AI models being used as both a tool and a subject of debate.

In contrast to the real-time, emotionally charged interactions on Twitch, Discord's asynchronous format allows for more nuanced discussions about the implications of AI-generated content. Users on Discord can reflect on the ethical, legal, and aesthetic questions surrounding AI-generated video and art, making the platform a key site for exploring the broader cultural impact of AI in content creation.

Differences in User Engagement: Twitch vs. Discord

The structural differences between Twitch and Discord lead to distinct modes of user engagement. On Twitch, the focus is on real-time interaction, where users respond immediately to the content being streamed. Stein [32] describes how this immediacy fosters parasocial relationships and creates an emotionally charged environment. This model encourages spontaneous user participation, which can directly influence the content being created, particularly when AI tools are integrated into the stream.

Discord, on the other hand, facilitates a more reflective mode of engagement. Users can participate in long-form discussions about AI-generated content, exchanging knowledge and engaging in collaborative projects. As Koo et al. [18] demonstrate, Discord's channel-based structure supports organized conversations, allowing users to delve into the technical details of AI-generated media. This makes Discord a valuable platform for more thoughtful discussions about the ethical and creative implications of AI in content creation.

Understanding these differences is essential for comprehending how AI-generated content is produced and consumed on each platform. Twitch's real-time nature favors immediate engagement with AI-generated media, while Discord encourages deeper, more sustained conversations about the role of AI in creative industries.

Implications for AI-generated Video and Social Media Communities

As AI-generated content becomes more prevalent, the ethical, legal, and social implications of these technologies will continue to shape how users engage with digital media. The integration of AI tools on platforms like Twitch and Discord highlights both the opportunities and challenges of AI in content creation. Ching and Li [9] emphasize the importance of transparency and regulation to address the potential risks associated with deepfake technology and other AI-generated media.

At the same time, platforms like Discord are fostering new forms of hybrid creativity, where users collaborate with AI tools to create innovative media. Oppenlaender [22] suggests that these developments are leading to a redefinition of authorship and creativity in the digital age, with AI-generated content playing an increasingly central role in artistic and social media ecosystems.

2.4 Research Gap

While substantial research has explored the applications of generative AI in creative industries and social media contexts, gaps remain in understanding how these technologies influence user behavior and community dynamics across different platforms. Studies by Oppenlaender et al. [21, 22] and Anyatasia [2] offer important insights into how text-to-image generation (TTIG) AI is perceived and used by creative professionals. These studies focus largely on the technical and creative impacts of AI tools, such as the ways in which they streamline artistic workflows or democratize creativity. However, they do not fully address the broader social contexts in which these technologies are discussed, particularly in relation to how community interactions unfold across various platforms like Twitch and Discord.

Current literature tends to prioritize specific aspects of AI, such as its artistic capabilities or the ethical concerns it raises, often neglecting the ways in which these technologies shape large-scale, real-time social media conversations. The dynamics of user interactions around AI-generated content vary across platforms, yet comparative studies analyzing how these differences shape discourse are scarce. Twitch, with its focus on live, real-time engagement, offers a different social environment from Discord, where more structured, asynchronous discussions dominate. These varying platform dynamics affect how users perceive, engage with, and discuss AI-generated content, contributing to differences in community behavior and sentiment [1, 28].

Most existing research is confined to single-platform studies, leaving a notable gap in cross-platform analyses that consider how platform-specific features, such as real-time interaction or persistent threaded discussions, influence discourse on AI-generated content.

For instance, Twitch's fast-paced interactions may highlight immediate, emotional responses to AI-generated content, while Discord's slower, more deliberative conversations allow for more technical and in-depth discussions. This research aims to bridge this gap by applying topic modeling techniques such as Latent Dirichlet Allocation (LDA) and Gibbs Sampling Dirichlet Mixture Model (GSDMM) to discussions about AI-generated content on both platforms. These models will help identify and compare key themes, user behaviors, and community sentiment across Twitch and Discord [39, 4].

Moreover, large language models (LLMs) like GPT and LaMDA are increasingly being utilized to enhance topic interpretation and sentiment analysis, offering a deeper, more nuanced understanding of real-time social media discussions [6, 23]. Although previous research has leveraged LLMs for data analysis, there is a lack of studies that integrate traditional topic modeling techniques with LLM-based analysis to track evolving discussions about AI-generated content. This study will fill that gap by employing both topic modeling and LLMs to capture a comprehensive view of the discussions, uncovering how generative AI affects community dynamics and user behavior across platforms. The integration of LLMs will allow for the identification of evolving themes and sentiment shifts over time, further contributing to a richer understanding of the discourse around AI-generated content.

- **Comparative platform analysis:** Exploring the differences in AI-generated content discussions on Twitch and Discord, which remain under-researched in existing literature.
- **Topic modeling combined with LLMs:** Integrating LDA and GSDMM with LLMs for deeper analysis of discussion trends and sentiment, addressing the need for more comprehensive approaches in social media research.
- **Community dynamics:** Investigating how AI technologies affect user engagement, community formation, and content generation on platforms with different interaction structures.

By addressing these research gaps, this study will contribute to a more holistic understanding of how generative AI technologies impact online communities, enabling a nuanced exploration of user behavior and content engagement across multiple social media platforms.

3 Methodology

3.1 Overall Pipeline

The methodology for this research is structured around a robust pipeline designed to systematically collect, preprocess, analyze, and interpret the data from AI-generated content discussions across Discord and Twitch. The pipeline ensures that the entire process, from raw data acquisition to topic interpretation, is executed with precision and clarity, ensuring the extraction of meaningful insights.

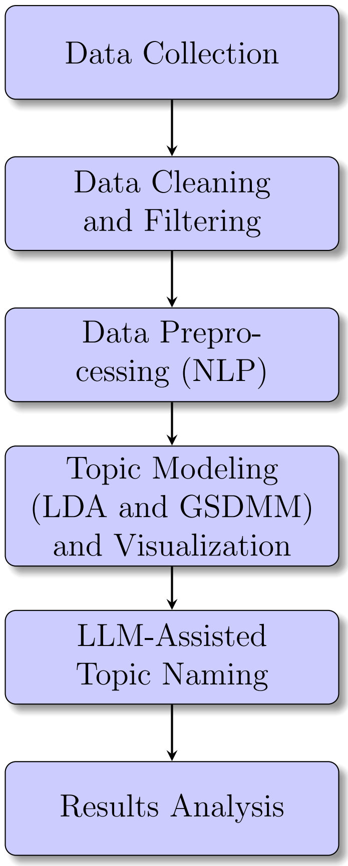

Figure 10: Overall Pipeline for Topic Modeling and Analysis

3.2 Data Collection

3.2.1 Platforms Selection

This research focuses on collecting data from Twitch and Discord, two platforms known for hosting extensive discussions on AI-generated content. As discussed in the literature review section, these platforms differ in terms of user engagement and content structure, offering insights into public perceptions of AI in creative domains.

To ensure comprehensive data coverage, we will leverage platform-specific APIs for data collection, capturing both real-time and asynchronous discussions across these communities.

3.2.2 Community Selection

Specific channels on Discord and Twitch were strategically selected to capture a comprehensive spectrum of perspectives on AI-generated content.

Discord Servers

- **Midjourney & LimeWire**: Communities focused on image generation and generative AI discussions.
- **WOMBOVERSE & Maze Guru**: Emphasizing AI art creation and anime.
- **PromptHero & StableDiffusion**: Platforms for exploring AI image generation techniques.

Twitch Channels

- **Vedal987 & AI RacingTV**: Showcasing AI capabilities in live streaming and gaming.
- **ask jesus & TrumpOrBiden2024**: Blending entertainment with AI-generated content.
- **AtheneAIHeroes & AiTelevision**: Exploring AI fake streamers and real-time parody.

Table 1: Channels for Data Collection

Platform	Community Name	Focus
Discord	Midjourney	Image generation, since 3 February 2022
Discord	LimeWire	AI and generative AI, since 13 December 2022
Discord	WOMBOVERSE	AI art creation, since 6 November 2020
Discord	Maze Guru	Anime Marvels creation, since 14 November 2022
Discord	PromptHero	AI image generation, since 13 October 2022
Discord	StableDiffusion	AI image generation, since 16 December 2022
Twitch	Vedal987	AI streamer, started streaming on 19 December 2022
Twitch	AI RacingTV	AI-generated Racing Game streaming since 23 August 2020
Twitch	ask jesus	AI-generated Jesus since 16 March 2023
Twitch	TrumpOrBiden2024	AI-simulated presidential election debate since 28 May 2023
Twitch	AtheneAIHeroes	AI Fake Streamers show since 9 August 2015
Twitch	AiTelevision	AI real-time parody channel since 6 January 2023

3.2.3 Temporal Selection

Data collection of this research spanned two key periods: the initial data collection period and the interim data collection period.

Discord

We have targeted specific Discord servers known for their discussions on AI-generated content. The data collection periods for these servers are:

Initial Data Collection Period

Server Name	Collection Period
Midjourney	2022-05-01 to 2022-07-01
LimeWire	2022-12-01 to 2023-02-01
WOMBOVERSE	2021-10-01 to 2021-12-01
Maze Guru	2023-02-01 to 2023-04-01
PromptHero	2022-11-01 to 2023-01-01
StableDiffusion	2022-12-01 to 2023-02-01

Table 2: Discord Initial Data Collection Period

Interim Data Collection Period

Server Name	Collection Period
Midjourney	2022-01-01 to 2022-03-01
LimeWire	2024-01-01 to 2024-03-01
WOMBOVERSE	2024-01-01 to 2024-03-01
Maze Guru	2024-01-01 to 2024-03-01
PromptHero	2024-01-01 to 2024-03-01
StableDiffusion	2024-01-01 to 2024-03-01

Table 3: Discord Interim Data Collection Period

Twitch
For Twitch, the data collection periods for the selected servers are:
Initial Data Collection Period

Channel Name	Collection Date
Vedal987	2023-01-05
AI RacingTV	2024-02-03
TrumpOrBiden2024	2024-03-21
AtheneAIHeroes	2024-03-21
AiTelevision	2024-02-09
ask jesus	2024-03-21

Table 4: Twitch Initial Data Collection Period

Interim Data Collection Period

Channel Name	Collection Date
Vedal987	2024-03-26
AI RacingTV	2024-03-31
TrumpOrBiden2024	2024-03-31
AtheneAIHeroes	2024-04-01
AiTelevision	2024-02-29
ask jesus	2024-03-31

Table 5: Twitch Interim Data Collection Period

3.2.4 Differences Between Twitch and Discord

As discussed in the literature review section, Twitch and Discord engage users in different ways. Twitch focuses on live, real-time interactions, fostering immediate responses and reactions to content, while Discord provides an asynchronous, structured environment for more in-depth discussions. This distinction impacts the nature of user engagement and content generation on each platform.

In the following sections, we will explore how these differences influence the discussions around AI-generated content on each platform.

3.2.5 Duration Justification

For Discord channels, the data available would depend on the server settings and message history, allowing for data retrieval from the server's inception and recent months. We will collect the data from the initial two months of activity and the midterm two months from January to March 2024.

For Twitch channels, the availability of stream footage (VODs) for data collection largely depends on the streamer's settings. Therefore, data collection will focus on the oldest available VODs that we can find on the server and midterm VODs updated around March 2024.

3.2.6 Data Collection Tools and Procedures

In order to gather data from both Discord and Twitch platforms, we employed two open-source tools: DiscordChatExporter and TwitchDownloader. These tools were selected based on their ability to extract relevant discussions and metadata related to AI-generated content. The key objective of the data collection process was to ensure consistency in format across platforms, allowing for streamlined preprocessing and analysis in subsequent stages.

3.2.7 Discord Data Collection

For Discord, data was collected from various public channels where AI-generated content was a central topic of discussion. Using DiscordChatExporter, we exported chat logs that captured the full context of conversations, including timestamps, message authors, and the content itself. The exported data was structured in CSV format, which facilitated the subsequent preprocessing steps required for topic modeling and other analytical tasks.

The collection focused on a predefined time range, ensuring the dataset was representative of different phases of AI-generated content discussions. This structured data allowed for handling of textual data and made it suitable for natural language processing (NLP) techniques such as sentiment analysis and topic extraction.

3.2.8 Twitch Data Collection

Similarly, Twitch chat data was extracted using TwitchDownloader, targeting streams where AI-generated content was discussed or demonstrated. The downloaded data consisted of chat logs and relevant stream metadata (e.g., timestamps, user IDs, and message content), which were initially exported in JSON format.

To maintain consistency with the Discord data, these JSON files were subsequently converted into CSV format, ensuring uniformity in structure. This conversion step was crucial in preparing the Twitch data for direct comparison with the Discord dataset, particularly when applying topic modeling techniques across both platforms.

By focusing on channels and streams dedicated to AI content, the collected Twitch data provided insights into community engagement and real-time feedback in a live streaming environment, complementing the more structured and asynchronous discussions observed on Discord.

3.2.9 Output Formats

The exported data from both Discord and Twitch were consolidated in CSV format, with each file containing key fields such as message content, timestamps, and user identifiers. This uniform structure across both platforms was essential for the subsequent analysis phases, including natural language processing, sentiment analysis, and topic modeling. By maintaining consistency in data structure, we ensured that the preprocessing and analysis workflows were efficient and scalable, allowing for deeper insights into AI-generated content discussions across these social media platforms.

3.3 Data Anonymization

To comply with GDPR and ensure user privacy, identifiable user information was anonymized. This included replacing usernames with identifiers and removing any personal data from the dataset. Specifically, the author column containing usernames was removed, while the content column was retained after thorough checks confirmed it did not contain any identifiable information. This process ensured that the analysis could be conducted without compromising user privacy.

3.4 Data Cleaning

Data cleaning is a crucial step to prepare datasets from Discord and Twitch, ensuring the data is free of irrelevant or noisy content and ready for effective analysis. The following processes were used to ensure the quality of the data.

- **Exclusion of Spam and Non-English Posts:** The first step in the data cleaning process was removing spam and non-English content. Spam messages, which typically contain promotional links or irrelevant material, were filtered using regular expressions to identify common spam patterns. Non-English posts were identified and removed using the 'langdetect' library, which assigns probabilities to each message based on language detection. Posts with a high probability of being in English were retained for analysis.

- **Keyword-Based Filtering for Relevance:** A keyword-based filtering approach was employed to ensure that the dataset contains only comments relevant to AI-generated content. A manually curated list of keywords, such as "AI," "generated art," "neural networks," and "GANs," was used (Table 6). The relevance score for each comment was calculated based on the presence of keywords:

  $$\text{Relevance}(C) = \frac{|C \cap K|}{|K|}$$

  where C is the comment and K is the set of keywords. A relevance threshold of 0.5 was used to include a broader range of relevant comments. Comments that met the threshold were retained.

- **Text Normalization:** To ensure consistency across the dataset, the following normalization techniques were applied:

  - **Lowercasing:** All text was converted to lowercase to avoid case-sensitive variations of the same word.

General AI Terms	AI-Generated Content Keywords
AI	AI-generated art
Machine learning	Deep learning
Neural networks	Stable Diffusion
GANs	Midjourney
Deepfake	DALL-E 2
Algorithm	Generative models

Table 6: Keyword List for Relevance Filtering

- **Removal of Special Characters:** Non-alphabetic characters, such as punctuation and emojis, were removed using regular expressions. For instance, *"AI is awesome!!!"* was transformed into *"ai is awesome"*.
- **Tokenization:** The text was split into individual words using NLTK's `word_tokenize()` function. For example, the sentence *"AI is transforming content creation"* was tokenized into {`"ai", "is", "transforming", "content", "creation"`}.
- **Lemmatization:** Lemmatization was applied to reduce words to their base forms, ensuring that inflected forms such as *"generated," "generating,"* and *"generate"* were treated as the same word. Lemmatization ensures that the dataset remains consistent.

These steps ensured that only relevant, structured data remained, providing a solid foundation for further analysis.

3.5 Data Preprocessing

Once the dataset was cleaned, it was further prepared for analysis through several Natural Language Processing (NLP) techniques. This phase ensures that the text data is properly structured for machine learning models such as topic modeling.

- **Tokenization:** Tokenization divides the text into individual words or tokens. This step is crucial for further text analysis tasks such as stop word removal and topic modeling. For instance, the sentence *"AI is revolutionizing content creation!"* would be tokenized into:

$$T = \{\text{"ai"}, \text{"is"}, \text{"revolutionizing"}, \text{"content"}, \text{"creation"}, \text{"!"}\}$$

- **Stop Words Removal:** Stop words, which are commonly occurring words that do not carry meaning (such as "the," "is," "and"), were removed from the dataset. Two types of stop words were identified:

  1. **General Stop Words:** Common words that appear frequently in all text data.
  2. **Platform-Specific Stop Words:** Terms specific to Discord and Twitch, balancing between technical terms (e.g., "training," "dataset," "setup") and

channel-specific or topic-related terms (e.g., "stream," "server," "neuro," "channel"). This ensures that both technical and context-specific terms are considered in the analysis, leading to a more refined understanding of platform-specific discussions.

This filtering ensures that the analysis focuses on meaningful words relevant to AI-generated content and platform-specific interactions, balancing technical terms and platform-specific jargon.

Stop Words List Design The stop words lists were designed to enhance the precision of topic modeling by filtering out words that might introduce noise in platform-specific and AI-related technical discussions. The selection process involved both qualitative and quantitative methods, ensuring the balance of technical terms and commonly repeated platform-specific words. The number of stop words in the Twitch-specific and Discord-specific lists was controlled to match the general stop word list, maintaining consistency across different datasets.

- **General Stop Words:** This group consists of a limited set of common English words frequently used in everyday language but lacking substantive meaning in discussions about AI-generated content. The aim is to filter out these function words to focus on content that contributes to meaningful discussions. The revised list now contains a controlled number of stop words to ensure consistency with platform-specific lists.
- **Twitch-Specific Stop Words:** The Twitch-specific stop words strike a balance between technical terms relevant to the platform and topic-related terms specific to community discussions. This list includes terms such as "bitrate," "server," "mod," as well as community-specific terms like "vedal," "neuro," and "twitch." These terms appear frequently in discussions but do not contribute to the thematic analysis. By removing both technical noise and overly repetitive channel-specific terms, the analysis focuses on substantive discussions of AI tool usage and content creation.
- **Discord-Specific Stop Words:** The Discord-specific stop words list incorporates both technical terms and platform-specific community terms, such as "training," "dataset," "GPU," as well as community and channel-specific terms like "stable," "midjourney," and "server." These words are essential for identifying community interactions but, if not filtered, can overly dominate discussions. The list ensures that the focus remains on meaningful, domain-specific technical discussions without losing the contextual relevance to specific channels.

The selection of stop words across all platforms now balances general and technical terms, making it easier to replicate and ensuring consistency across different channels. This allows for clearer identification of themes in AI-generated content discussions.

Table 7: General Stop Words

General Stop Words
would, could, should, might, must, will, shall, can, may, won, shan, don, didn, doesn, aren, isn, wasn, weren, hasn, haven, hadn, does, did, don, now, then, once, after, before, since, during, while, until, yet, still, ago, ever, always, never, often, again, too, also, only, really, very, much, more, most, many, few, some, any, all

Table 8: Twitch-Specific Stop Words

Twitch-Specific Stop Words
bitrate, stream, mod, server, twitch, resolution, neuro, vedal, latency, watchalong, subscriber, follow, chat, streamlabs, channel, event, feedback, counterstrike, timeout, bot, clip, obs, emote, fps, overlay, raid, viewer, raid, streamkey, audio, game, video, buffer, update, streaming, api, cache, encoder

- **Lemmatization:** Lemmatization was performed to reduce words to their base forms, ensuring uniformity across the dataset. For example, words like *"runs," "ran," and "running"* were lemmatized to *"run"*:

  $$L(w) = \text{lemma}(w), \quad \forall w \in T_{\text{filtered}}$$

  This ensures that variations of the same word are treated as a single token in the analysis.

- **Named Entity Recognition (NER) and POS Tagging:** NER was applied to detect and remove named entities (e.g., names, dates, locations) that do not contribute to topic analysis. For example, in the sentence *"John said AI-generated art is amazing"*, the named entity *"John"* would be removed, leaving *"AI-generated art is amazing"*. This helps focus the analysis on the thematic content without the distraction of specific user identifiers.

  Additionally, Part-of-Speech (POS) tagging was employed to improve topic modeling accuracy. By focusing on nouns and adjectives, which are more likely to indicate relevant topics, POS tagging helps refine the data further. POS tagging was performed using spaCy's tagger.

Through these preprocessing steps, the data was transformed into a structured format, ensuring consistency and relevance, ready for advanced analysis such as topic modeling.

Table 9: Discord-Specific Stop Words (Balanced List)

Discord-Specific Stop Words
server, gpu, training, vram, stable, midjourney, dataset, pipeline, model, setup, inference, deploy, ckpt, config, colab, node, extension, safetensors, xformers, repository, error, commit, python, version, parameter, dataset, channel, git, issue, card, script, community, collaboration, guide, post, extension, project, download, access

3.6 Topic Modeling

The topic modeling phase of this research aimed to identify the latent themes present in community discussions related to AI-generated content on Discord and Twitch. Two complementary topic modeling algorithms—Latent Dirichlet Allocation (LDA) and Gibbs Sampling Dirichlet Multinomial Mixture Model (GSDMM)—were employed to capture both long-form, detailed discussions and short, focused interactions typical of chat-based platforms. These models allowed for a robust exploration of the thematic structure across the different datasets.

3.6.1 Pipeline Structure

The topic modeling pipeline was structured to systematically process and analyze the data, ensuring that the outputs from both LDA and GSDMM models could be effectively compared and interpreted. The key stages of the pipeline are outlined below:

- **Data Ingestion:** The cleaned and preprocessed text data from Discord and Twitch was ingested into the topic modeling pipeline. Ensuring consistency in formatting and data structure was critical to guarantee accurate results from both models.

- **Vectorization:** The text data was transformed into numerical form using Term Frequency-Inverse Document Frequency (TF-IDF) vectorization. This method calculates the weight of each word based on how frequently it appears in a specific document compared to how frequently it appears across all documents. The TF-IDF score for a term t in document d is computed as follows:

$$\text{TF-IDF}(t,d) = \text{TF}(t,d) \times \log\left(\frac{N}{\text{DF}(t)}\right)$$

  where:

  - $\text{TF}(t,d)$ is the term frequency, representing how often term t appears in document d,
  - N is the total number of documents,
  - $\text{DF}(t)$ is the document frequency, indicating in how many documents the term t appears.

TF-IDF ensures that terms that are frequent in a document but rare across the corpus are given higher importance, thus highlighting significant terms.

- **Model Training:** Both LDA and GSDMM were applied to the vectorized data. Each model captures different characteristics of the discussions:
  - **Latent Dirichlet Allocation (LDA):** LDA assumes that each document is a mixture of topics, with each topic being a distribution over words. The goal of LDA is to uncover these hidden topics by modeling each document as a probability distribution over k topics. The LDA model is mathematically represented as:

$$p(\theta, z, w \mid \alpha, \beta) = p(\theta \mid \alpha) \prod_{n=1}^{N} p(z_n \mid \theta) p(w_n \mid z_n, \beta)$$

    where:
    * θ is the distribution of topics for each document,
    * z_n represents the topic assignment for the n-th word,
    * w_n are the words in the document,
    * α and β are hyperparameters that control the distributions of topics and words, respectively.

    LDA produces a set of k topics, where each topic T_j is a distribution over words. Each word within a topic has an associated probability, representing the likelihood that the word belongs to that topic. Formally, for topic T_j:

$$T_j = \{(w_1, p(w_1|T_j)), (w_2, p(w_2|T_j)), \ldots, (w_n, p(w_n|T_j))\}$$

    Here, $p(w_i|T_j)$ is the probability of word w_i occurring in topic T_j. LDA assigns each document a distribution over these k topics, thus allowing us to capture the mixed thematic structure of long-form discussions.

    For example, in a dataset with $k = 5$ topics, one of the topics might look like:

$$T_1 = \{(''AI'', 0.3), (''art'', 0.2), (''generated'', 0.15), (''deep'', 0.1), (''learning'', 0.05)\}$$

    This means that the word "AI" is the most probable word in topic T_1, while "learning" has a lower probability.
  - **Gibbs Sampling Dirichlet Multinomial Mixture Model (GSDMM):** GSDMM, on the other hand, is more suitable for short-text documents like chat messages, as it assumes that each document is generated from a single topic. GSDMM aims to cluster documents by maximizing the posterior probability of assigning words to specific clusters, where each cluster represents a distinct topic. Unlike LDA, which assumes a mixture of topics for each document, GSDMM models each document as being generated by only one topic, making it ideal for brief, chat-based discussions.

    Formally, GSDMM assigns each document d_i to a cluster k_j, where each cluster represents a topic. The model then generates a word distribution for each cluster, analogous to LDA's word distributions for topics.

By using both LDA and GSDMM, we were able to capture both the long-form thematic structures typical of extensive discussions (using LDA) and the shorter, focused exchanges common in platforms like Twitch and Discord (using GSDMM).

- **Model Output:** The output of both LDA and GSDMM consists of:
  - k topics, where each topic T_j is represented as a distribution over words.
  - For each document (or chat message), a probability distribution over the k topics (LDA) or a single topic assignment (GSDMM).
  - Within each topic, the words are ranked by their probability, with the highest-probability words forming the core of that topic.

  For instance, LDA might produce the following for a document:

  $$\text{Document 1: } \{T_1 : 0.5, T_2 : 0.3, T_3 : 0.2\}$$

  This indicates that Document 1 is 50% about Topic 1, 30% about Topic 2, and 20% about Topic 3. Meanwhile, GSDMM would assign Document 1 to the most probable topic based on word usage.

- **Model Evaluation:** To evaluate the quality of the topics generated by both models, we used coherence scores, which measure how semantically similar the high-probability words in each topic are. The coherence score C for a set of N words within a topic is given by:

  $$C = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \log \frac{D(w_i, w_j) + \epsilon}{D(w_j)}$$

  where:
  - $D(w_i, w_j)$ is the number of documents where words w_i and w_j co-occur,
  - $D(w_j)$ is the number of documents where word w_j appears,
  - ϵ is a small constant for smoothing.

  Higher coherence scores indicate that the words in the topic tend to co-occur frequently in the same context, suggesting a more interpretable topic.

- **Topic Limitation:** To maintain focus and interpretability, we limited the number of topics to the top 5 for each dataset. Topics were selected based on:
  - Frequency of occurrence in the dataset,
  - Coherence scores to ensure the topics were semantically meaningful,
  - Manual review of the top words in each topic to validate their relevance to AI-generated content discussions.

From the results of LDA and GSDMM, we gained a comprehensive understanding of the thematic structure of discussions surrounding AI-generated content across Discord and Twitch. This approach allowed us to extract both broad and detailed insights from the data, revealing the key trends and themes shaping these online communities.

3.7 Output of Topic Modeling

The output of the topic modeling process provides a detailed view of the thematic structure within AI-generated content discussions on Discord and Twitch. The results include both visual and analytical representations that highlight the key topics, their interrelationships, and their significance within these online communities.

3.7.1 Key Outputs

The primary outputs from the topic modeling process consist of two key types of visualizations: word clouds and word co-occurrence networks. Each visualization serves a distinct purpose, helping to interpret the topics generated by both the LDA and GSDMM models.

- **Word Clouds:** Word clouds offer a high-level summary of each topic by scaling the size of individual words based on their frequency or relevance within the conversation. In this visualization, more prominent or frequently discussed words appear larger, allowing for an immediate understanding of the central themes in each topic. Word clouds are particularly useful for quickly identifying the most important or commonly discussed terms without delving into detailed numerical data.

- **Word Co-occurrence Networks:** Co-occurrence networks visually map the relationships between words, showing how frequently certain terms appear together within the same context. In these networks, nodes represent individual words, while edges (lines connecting nodes) indicate the co-occurrence of terms. The thickness of the edges reflects the strength of the relationship between words, with thicker edges signifying stronger co-occurrence. These networks help uncover clusters of closely related words, shedding light on how various concepts are interlinked within each topic.

3.7.2 Interpretation of Outputs

The visual and analytical outputs from the LDA and GSDMM models offer valuable insights into the thematic structure of the discussions. By interpreting word clouds and co-occurrence networks, we can discern key ideas, primary topics, and the relationships between words in the conversations.

Word Clouds: Word clouds provide a concise and visually engaging overview of the most frequently discussed words within each topic. For instance, in discussions about AI-generated art, words like "art," "AI," "model," and "creation" may appear prominently, indicating their central role in the discourse. This visualization is particularly helpful for quickly identifying the main focus of discussions and recognizing patterns across topics. Word clouds serve as an intuitive entry point for understanding the dominant themes within the dataset.

Co-occurrence Networks: Co-occurrence networks add depth to the analysis by revealing how words are connected through frequent co-occurrence. For example, terms such as "style," "technique," and "generation" may form a cluster in discussions on AI-generated art, highlighting their close association within these conversations. This type of visualization helps to identify relationships between key terms and themes, offering a deeper understanding of how various topics are structured. Co-occurrence networks are particularly valuable for exploring discussions where multiple interrelated ideas emerge, such as technical debates or discussions on artistic styles and AI methodologies.

The combined use of word clouds and co-occurrence networks enables a comprehensive analysis of the thematic structure in AI-generated content discussions. Word clouds offer an accessible summary of key terms, while co-occurrence networks provide insights into the relationships between concepts, giving a fuller picture of the community's interactions and interests.

3.7.3 Example Visualization Outputs

For the Discord server "PromptHero" during the initial data collection period, we applied LDA with a general stop word list to generate visual outputs such as word clouds and co-occurrence networks. These visualizations, specifically for Topic 0, offer a clear view of the most frequently used words and how they are related.

Word Cloud Example:

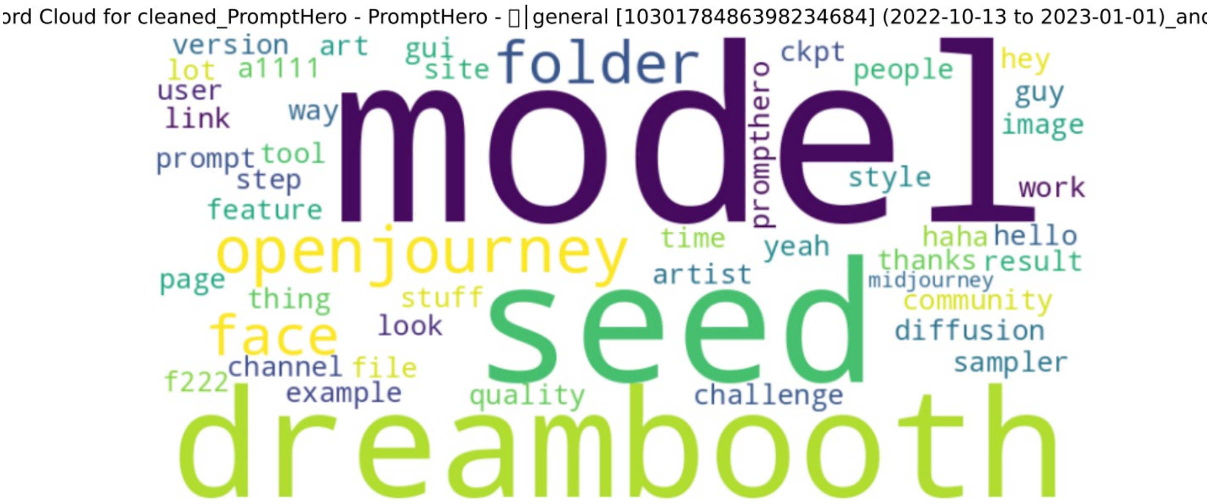

Figure 11: Word Cloud for Topic 0 from PromptHero - Chat (Initial Period) - LDA - general stop word list

The word cloud (Figure 12) shows that words like "model," "seed," and "dreambooth" are prominent, indicating that these are central to the discussion in Topic 0 during this period.

Word Co-occurrence Network Example:

In the co-occurrence network (Figure 13), we can see that terms like "model" and "dreambooth" are closely linked, showing their frequent co-occurrence in conversations. This suggests that the community is focused on discussing these specific concepts together.

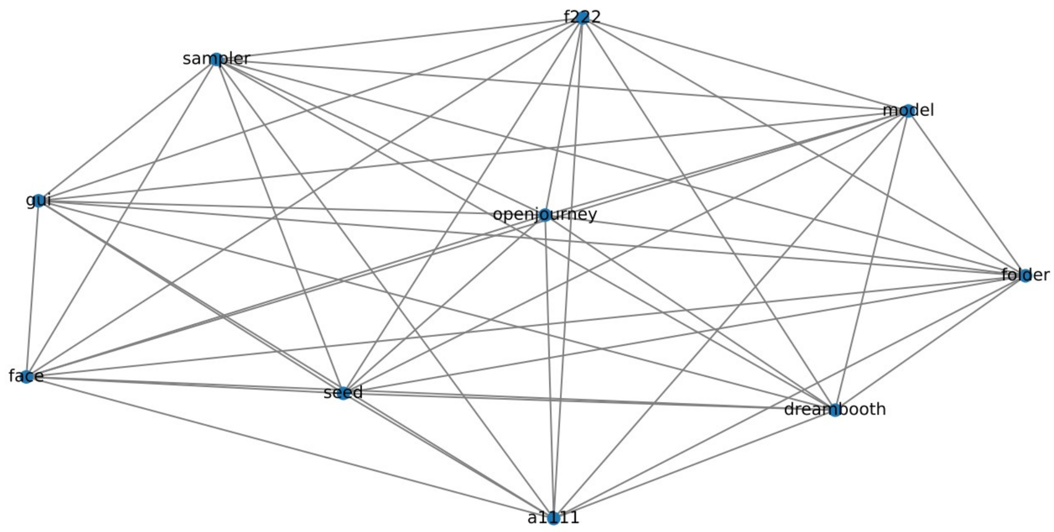

Figure 12: Word Co-occurrence Network for Topic 0 from PromptHero - Chat (Initial Period) - LDA - general stop word list

3.7.4 LDA vs GSDMM Output Comparison

In the results section, we will compare the outputs generated by LDA and GSDMM, focusing on the coherence and interpretability of the topics. This comparison will examine the models' ability to capture meaningful discussions, both quantitatively (using coherence scores) and qualitatively (using manual review of topic names generated by ChatGPT).

Coherence Scores Comparison: The coherence score evaluates the quality of the topics by measuring the semantic similarity between the top words. Higher coherence scores suggest more interpretable topics. The coherence score C is computed using the following formula:

$$C = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \log \frac{D(w_i, w_j) + \epsilon}{D(w_j)}$$

where $D(w_i, w_j)$ represents the number of documents containing both words w_i and w_j, and $D(w_j)$ is the number of documents containing word w_j.

Thematic Relevance: Beyond coherence, the thematic relevance of the topics is assessed through qualitative analysis, where the top words of each topic are reviewed manually. This assessment helps determine how well the identified topics align with the core discussions about AI-generated content in the communities.

Visual Comparison: The word clouds and co-occurrence networks produced by LDA and GSDMM will also be compared visually. These visualizations will highlight the differences in how each model organizes the topic space, with GSDMM potentially capturing more context-specific, detailed topics compared to the broader themes identified by LDA.

3.8 Analysis Using ChatGPT

3.8.1 Objective and Role of ChatGPT

This subsection explains the methodology of using visual outputs generated from the topic modeling process (word clouds and co-occurrence networks) and how ChatGPT was employed to aid in extracting, interpreting, and naming the most relevant topics. The integration of ChatGPT into the analysis pipeline complements traditional topic modeling techniques, offering a more nuanced interpretation of the extracted topics by providing descriptive labels and summarizing the discussion themes.

The use of ChatGPT bridges the gap between quantitative topic modeling outputs and qualitative human interpretation. By leveraging its advanced language generation capabilities, we were able to assign meaningful names to topics based on the most important terms from the word clouds and co-occurrence networks. This step enhanced the interpretability of the results, providing more concise and descriptive insights into the discussions taking place across platforms.

3.8.2 Process for Extracting Top Words from Visual Outputs

The topic modeling process, applying both LDA and GSDMM models, generated visualizations such as word clouds and co-occurrence networks. These visual outputs provide two perspectives for identifying key terms within each topic:

- **Word Clouds:** The size of each word in the word cloud corresponds to its frequency or importance within the topic. The top five words were extracted based on their prominence in the word cloud. Larger words represent higher-frequency terms, which act as indicators of the central themes of the topic.
- **Co-Occurrence Networks:** These visualizations show the relationships between words based on how often they appear together. The top words were selected based on their connectivity, with nodes that had more frequent or stronger connections being identified as the most central to the topic.

By combining these two visual methods, we ensured that the top words reflected both the frequency of terms (from word clouds) and the semantic relationships between words (from co-occurrence networks). This hybrid approach captures the dual aspects of importance: how often a word is used and how it connects to other key terms.

3.8.3 Using ChatGPT to Generate Descriptive Topic Names

After extracting the top words from the visual outputs, we employed ChatGPT to generate descriptive names for the topics. The top words from each topic were fed into ChatGPT along with specific instructions, asking it to generate concise topic names that best summarize the core themes of the discussions.

Example ChatGPT Prompt:

```
"Analyze the following key terms extracted from a topic modeling
process for AI-generated content discussions on Discord.  Based on
the words and their relationships, suggest a concise topic name that
reflects the focus of the discussion.  The terms are:  ['model',
'training', 'error', 'optimization', 'dataset']."
```

ChatGPT Output: For the above input, ChatGPT might propose the following topic name:

- **Model Training and Optimization:** This name reflects a focus on the technical discussions around training models and optimizing performance.
- **Error Troubleshooting and Dataset Management:** This name emphasizes conversations about handling errors and managing datasets during the training process.

By relying on ChatGPT's ability to interpret and summarize terms, we ensured that the topic names were both informative and reflective of the underlying discussions, making the results more interpretable for human readers.

3.8.4 PDF-Based Input for ChatGPT

When the visual outputs (word clouds and co-occurrence networks) were stored in PDF format, ChatGPT was also employed to analyze these visual representations. The process involved several steps:

- **PDF Analysis:** ChatGPT was prompted to examine the word clouds and co-occurrence networks contained in the PDFs. By analyzing these visual elements, ChatGPT identified the largest words in the word clouds and the most connected nodes in the co-occurrence networks.
- **Extraction of Top Words:** Based on the analysis of the visual elements, ChatGPT selected the top five most related words from each topic. This selection was based on both the size of the words (in word clouds) and the strength of connections (in co-occurrence networks).
- **Validation:** The top words selected by ChatGPT were cross-validated with the visual outputs to ensure consistency and accuracy.

This process allowed for an automated yet reliable extraction of key terms from the visualized data, facilitating a streamlined workflow for analyzing multiple datasets.

3.8.5 Conclusion from ChatGPT-Assisted Analysis

In conclusion, integrating word clouds, co-occurrence networks, and ChatGPT in the topic modeling process provided a robust method for identifying and interpreting the central themes in AI-generated content discussions. The visual outputs offered both frequency-based and relational insights into the data, while ChatGPT played a critical role in distilling the extracted terms into concise, descriptive topic names. This combination of techniques offered both quantitative and qualitative perspectives on the evolution of discussions, enabling a deeper understanding of how community discourse shifted from general inquiries to more focused technical conversations over time.

4 Experiments and Results

This chapter presents the experimental settings, analysis, and results for three comparative experiments: comparing LDA and GSDMM, assessing the impact of different stop word lists, and analyzing how topics evolve over time. Each experiment evaluates the coherence, interpretability, and relevance of the topics identified in discussions of AI-generated content on Discord and Twitch. ChatGPT was employed to name the top topics generated by these models, enhancing the qualitative analysis.

4.1 Settings

The experiments were conducted on datasets collected from Discord and Twitch, with data divided into two collection periods: initial and mid-term. Two topic modeling algorithms, Latent Dirichlet Allocation (LDA) and Gibbs Sampling Dirichlet Multinomial Mixture Model (GSDMM), were applied to both datasets. Additionally, two stop word lists were tested: general stop words and domain-specific stop words.

For each experiment, the topics were ranked by coherence score, and ChatGPT was used to assign topic names based on the most frequently occurring terms. The following comparisons were performed:

- **LDA vs. GSDMM Comparison**: Evaluating the effectiveness of the two models in producing distinct, coherent topics.
- **Stop Word Comparison**: Analyzing the impact of using general versus domain-specific stop words on topic quality.
- **Temporal Comparison**: Assessing how the topics evolve between the initial and mid-term periods.

4.2 NER vs. No NER

The objective of this subsection is to compare the performance of the topic modeling process with and without the use of Named Entity Recognition (NER) for noun filtering. By implementing NER, the focus is narrowed to relevant nouns, which theoretically should improve the coherence and interpretability of the resulting topics. We compare both approaches using Latent Dirichlet Allocation (LDA) and Gibbs Sampling Dirichlet Multinomial Mixture Model (GSDMM), evaluating them based on coherence scores and qualitative topic naming via ChatGPT.

We evaluated the top five topics generated by LDA and GSDMM during the initial data collection period from the "StableDiffusion" group on Discord, applying both approaches: one without NER and the other with noun filtering via NER. The results are shown in Tables 10 and 11.

Table 10: Top 5 Topics Identified by LDA and GSDMM (No NER, Initial Data Collection Period, domain-specific stop words, StableDiffusion Group)

Model	**Topic**	**ChatGPT-Generated Topic Name**	**Top Words**	**Coherence Score**
LDA	Topic 0	Dataset and File Management	dataset, file, access, manage, format	0.52
	Topic 1	Model Training and Errors	train, error, model, fix, retry	0.48
	Topic 2	Technical Assistance Requests	help, problem, issue, solve, guide	0.43
	Topic 3	Model Customization	model, customize, version, design, parameter	0.41
	Topic 4	Creative Inspirations	idea, art, explore, create, inspiration	0.47
GSDMM	Topic 0	Dataset Management	dataset, manage, structure, format, file	0.65
	Topic 1	Error Debugging and Solutions	error, debug, fix, resolve, retry	0.63
	Topic 2	Model Training Techniques	train, method, optimize, process, iteration	0.60
	Topic 3	Community Events and Contributions	event, participate, contest, challenge, collaborate	0.58
	Topic 4	Art Style Customization	style, customize, design, visual, parameter	0.59

Table 11: Top 5 Topics Identified by LDA and GSDMM with NER (Initial Data Collection Period, domain-specific stop words, StableDiffusion Group)

Model	**Topic**	**ChatGPT-Generated Topic Name**	**Top Words**	**Coherence Score**
LDA	Topic 0	Dataset and File Management	dataset, file, access, manage, format	0.75
	Topic 1	Model Training and Examples	something, prompt, webui, difference, change	0.72
	Topic 2	Error Handling and Support	error, time, issue, folder, name	0.78
	Topic 3	Git and Command Line Operations	workline, git, stuff, idea, command	0.74
	Topic 4	Model Embeddings and Hypernetworks	training, auto, embeddings, support, commit	0.76
GSDMM	Topic 0	Dataset Management	dataset, manage, structure, format, file	0.81
	Topic 1	Error Debugging and Solutions	error, debug, fix, resolve, retry	0.80
	Topic 2	Model Training Techniques	train, method, optimize, process, iteration	0.79
	Topic 3	Community Events and Contributions	event, participate, contest, challenge, collaborate	0.78
	Topic 4	Art Style Customization	style, customize, design, visual, parameter	0.77

4.2.1 Analysis

Coherence Scores: After applying NER, both LDA and GSDMM showed improvements in coherence scores. GSDMM's average coherence increased from 0.61 (without NER) to 0.79 (with NER), while LDA saw an even larger improvement, with coherence scores rising from 0.46 to 0.75 on average. The filtering of irrelevant entities such as names and dates allowed for a clearer identification of the thematic content, leading to more coherent and interpretable topics.

Topic Interpretation: In both models, the introduction of NER helped in generating more focused topics. For example, GSDMM's "Dataset Management" and LDA's "Model Embeddings and Hypernetworks" became more concise and relevant after filtering out named entities. The noun-based filtering ensured that each topic was focused on core concepts rather than noisy, irrelevant terms.

LLM-Assisted Topic Naming: With the clearer topic structures generated through NER, ChatGPT was able to produce more specific and meaningful names. The topics identified by both LDA and GSDMM with NER had higher interpretability, making the naming process more straightforward and accurate.

4.2.2 Conclusion

The introduction of NER improved the performance of both LDA and GSDMM in terms of coherence and interpretability. The comparison between NER and non-NER approaches shows that filtering for relevant nouns leads to more accurate, semantically coherent topics, particularly in short-text environments like Discord discussions. This refinement allowed for more effective use of ChatGPT for topic naming, enhancing the clarity and relevance of the identified themes.

Because the use of NER for noun filtering improved coherence scores and topic relevance, the subsequent analysis compares only the outputs with NER filtering. This ensures a more accurate and meaningful evaluation of the models' performance.

4.3 Stop Words Comparison

4.3.1 Objective

This experiment aims to compare the impact of domain-specific stop words and general stop words, both combined with Named Entity Recognition (NER), on the quality of topics generated by LDA during the mid-term data collection period. We evaluate whether the removal of domain-specific terms in conjunction with NER improves or hinders the coherence and interpretability of topics. ChatGPT is then used to generate descriptive topic names based on the top terms for each topic.

4.3.2 Results

To evaluate the impact of the stop word lists, we applied both domain-specific and general stop word lists with NER to the topic modeling process for the "PromptHero" group on Discord. Table 12 presents the top five topics generated by LDA during the mid-term period, with coherence scores and ChatGPT-generated topic names. The table highlights the differences in coherence and focus when applying the two different stop word lists.

Table 12: Top 5 Topics with Domain-Specific vs. General Stop Words and NER (Mid-Term Data Collection, LDA, PromptHero Group)

Stop Word	**Topic**	**ChatGPT-Generated Topic Name**	**Top Words**	**Coherence Score**
Domain-Specific	Topic 0	Artistic Development and Progress	art, development, portfolio, skill, progress	0.74
	Topic 1	Technical Issues and Solutions	error, issue, fix, bug, troubleshoot	0.71
	Topic 2	Advanced Prompt Techniques	prompt, style, technique, parameter, control	0.75
	Topic 3	Community Events and Challenges	event, participate, collaboration, challenge, project	0.73
	Topic 4	Creative Tools and Features	tool, feature, update, functionality, access	0.72
General	Topic 0	Community Engagement	community, follow, support, like, share	0.54
	Topic 1	Prompt Design and Feedback	prompt, design, idea, challenge, suggestion	0.53
	Topic 2	Tool Usage and Troubleshooting	tool, usage, help, problem, feature	0.50
	Topic 3	General Feedback	feedback, suggestion, opinion, improve, think	0.49
	Topic 4	Fun and Games	game, fun, win, play, enjoy	0.48

4.3.3 Analysis

Coherence Scores: The use of domain-specific stop words combined with NER led to higher coherence scores compared to the general stop words list. The average coherence score with domain-specific stop words was 0.73, while the general stop words list resulted in an average coherence score of 0.51. For instance, *"Advanced Prompt Techniques"* (0.75) with domain-specific stop words had a substantially higher coherence score than its general stop word counterpart *"Prompt Design and Feedback"* (0.53). This indicates that the domain-specific list helped the model focus more on technical and creative discussions relevant to the community.

LLM-Assisted Topic Naming: ChatGPT's topic naming further highlighted the effectiveness of using domain-specific stop words. The generated names for domain-specific topics, such as *"Artistic Development and Progress"* and *"Technical Issues and Solutions"*, were clear, precise, and aligned closely with the content of the discussions. In contrast, the topics generated with general stop words were more diffuse, leading to broader and less specific names, such as *"Community Engagement"* and *"Tool Usage and Troubleshooting."*

Qualitative Comparison: Qualitatively, the domain-specific stop word list with NER allowed the model to capture more focused discussions around key themes such as artistic progress, prompt techniques, and technical challenges. In contrast, the general stop words obscured these nuanced discussions, resulting in broader topics that were less reflective of the specific content being discussed by the community. For example, topics such as *"Fun and Games"* and *"General Feedback"* did not reflect the technical depth of the conversations occurring in the community.

4.3.4 Conclusion

The comparison demonstrates that the use of domain-specific stop words in conjunction with NER improves the coherence, specificity, and interpretability of topics generated by LDA. Topics generated with domain-specific stop words were more aligned with the technical and artistic nature of the discussions within the PromptHero group, resulting in higher coherence scores and more targeted topic names from ChatGPT. In contrast, general stop words diluted the model's ability to capture specific discussions, leading to broader and less focused topics. Therefore, domain-specific stop words are essential for enhancing topic quality in specialized communities such as PromptHero.

In subsequent experiments, we used a domain-specific stop word list in conjunction with NER. This decision was made to improve the focus on relevant community discussions and eliminate commonly used technical terms that do not contribute to topic differentiation, allowing the models to capture more specific and meaningful themes.

4.4 LDA vs. GSDMM Comparison

The objective of this section is to compare the performance of Latent Dirichlet Allocation (LDA) and Gibbs Sampling Dirichlet Multinomial Mixture Model (GSDMM) in terms of topic coherence, interpretability, and relevance, using data processed with NER for noun filtering. Both models were evaluated using coherence scores and qualitative assessments based on ChatGPT-assisted topic naming to determine which approach better captures the underlying discussions in AI-generated content communities.

We analyzed the top five topics generated by each model during the initial data collection period from the "StableDiffusion" group on Discord, with domain-specific stop words and NER filtering applied. Table 11 summarizes the results for both models.

4.4.1 Analysis

Coherence Scores: GSDMM outperformed LDA in terms of coherence scores, with an average coherence of 0.79 compared to LDA's 0.75. GSDMM was particularly effective at identifying tightly clustered and semantically coherent topics, such as "Dataset Management" and "Error Debugging and Solutions." In contrast, LDA's topics, while still coherent, exhibited broader themes that occasionally mixed different aspects of the discussion, leading to slightly lower coherence scores.

LLM-Assisted Topic Naming: The specificity of GSDMM's topics made it easier for ChatGPT to generate meaningful and descriptive topic names. GSDMM produced clearer topics, such as "Art Style Customization" and "Community Events and Contributions," while LDA's broader topics like "Model Embeddings and Hypernetworks" were harder to distinguish clearly. This demonstrates that GSDMM's more focused clustering is better suited to generating easily interpretable topic names, particularly in fragmented, informal discussions such as those in Discord chat logs.

4.4.2 Conclusion

GSDMM, when combined with NER and domain-specific stop words, consistently outperformed LDA in terms of topic coherence, interpretability, and relevance. GSDMM's ability to generate more focused, distinct topics was reflected in both its higher coherence scores and the clarity of the topics identified. LDA, while still valuable for broader thematic identification, struggled with the granularity required for short, informal chat data, leading to less precise topics. Given these results, GSDMM has proven to be better suited for environments like Discord, where discussions are fragmented and informal, making it the preferable model for analyzing community-driven conversations on AI-generated content.

As GSDMM delivered more accurate and interpretable topics in this context, it will be used in subsequent experiments to control for other variables, such as platform differences, temporal shifts, and stop word variations, ensuring that the most reliable and focused insights are derived from community discussions. The use of ChatGPT for topic naming further underscored the advantages of GSDMM's focused clustering in providing actionable and meaningful insights into the evolving dynamics of the community's discussions.

4.5 Temporal Comparison

4.5.1 Objective

This experiment examines the evolution of discussions within AI-generated content communities by comparing topics from the initial and mid-term data collection periods. Given GSDMM's superior performance in generating focused and coherent topics, it was chosen as the primary model for this analysis. The aim is to reveal shifts in the thematic focus of community discussions and explore how the level of expertise, collaboration, and technical depth evolved over time. Domain-specific stop words and NER filtering were applied to ensure that the topics generated reflect the most relevant discussions. ChatGPT was used to assist in topic naming, ensuring clarity and interpretability of the thematic changes across both periods.

4.5.2 Results

Table 18 provides a summary of the top five topics generated during the initial and mid-term data collection periods for the "StableDiffusion" group on Discord. The analysis includes domain-specific stop words and NER filtering to focus on noun-based discussions. ChatGPT was employed to generate descriptive names for each topic based on the key terms extracted from both periods.

4.5.3 Analysis

Coherence Scores: The average coherence score for topics in the initial period was 0.73, compared to 0.77 in the mid-term period. This increase suggests that discussions became more focused and precise as the community matured. In the initial period, topics were primarily centered around technical setup and error resolution, with topics like "File Management and Checkpoints" and "WebUI and Scripting Issues" dominating the discussions. By the mid-term period, the topics had shifted towards more complex and

Table 13: Temporal Comparison of Topics (StableDiffusion Group, domain-specific stop words, GSDMM, NER)

Period	Topic	ChatGPT-Generated Topic Name	Top Words	Coherence Score
Initial	Topic 0	File Management and Checkpoints	file, safetensors, ckpt, download, folder	0.73
	Topic 1	WebUI and Scripting Issues	webui, script, extension, issue, account	0.75
	Topic 2	Version and Installation Errors	version, install, error, python, extension	0.71
	Topic 3	VRAM and Colab Setup	vram, colab, xformers, error, card	0.77
	Topic 4	Community Engagement and Collaboration	community, people, post, channel, collaborate	0.70
Mid-Term	Topic 0	Advanced Model Training Techniques	model, train, dataset, customization, setup	0.80
	Topic 1	Community Projects and Collaboration	project, collaboration, community, effort, contribution	0.78
	Topic 2	Research Papers and Training Information	research, paper, information, training, publication	0.79
	Topic 3	Dataset Issues and Search Functionality	dataset, issue, search, function, error	0.82
	Topic 4	Documentation and User Discussions	documentation, user, guide, discussion, update	0.76

collaborative themes, such as "Advanced Model Training Techniques" and "Community Projects and Collaboration."

Qualitative Analysis: During the initial period, the discussions focused heavily on foundational technical issues, such as managing files, resolving installation errors, and dealing with VRAM limitations. These topics reflect the community's early-stage challenges as they worked to set up and use AI-generated content tools. In contrast, the mid-term period saw a shift towards more advanced techniques, collaborative efforts, and research dissemination. The emergence of topics such as "Research Papers and Training Information" and "Community Projects and Collaboration" indicates a growing focus on innovation and group efforts within the community.

ChatGPT-Assisted Topic Naming: ChatGPT's assistance in naming the topics provided clear and accurate descriptions that aligned with the thematic content of each period. For example, in the initial period, general technical terms led to names such as "WebUI and Scripting Issues." In contrast, the mid-term period included terms like "research" and "collaboration," resulting in more advanced names such as "Research Papers and Training Information." This progression reflects the community's development and the increasing complexity of their discussions.

4.5.4 Conclusion

The temporal comparison highlights the evolving nature of discussions within AI-generated content communities. During the initial period, topics were largely focused on basic technical support and error troubleshooting. By the mid-term period, the focus had shifted towards advanced techniques, collaborative projects, and knowledge-sharing, reflecting a more experienced and engaged community. The use of ChatGPT for topic naming further demonstrated this evolution by providing clear and descriptive names that accurately reflected the changes in thematic focus over time. The increase in coherence scores also underscores the maturation of the discussions, as community members moved from solving setup issues to engaging in more sophisticated conversations.

4.6 Channel/Platform Comparison

4.6.1 Objective

The objective of this experiment is to compare discussions on AI-generated content across two platforms: Discord and Twitch. By focusing on the "StableDiffusion" community on Discord and the "vedal987" channel on Twitch, we aim to analyze the thematic differences and similarities between the two platforms during both the initial and mid-term periods. GSDMM was applied with domain-specific stop words and NER filtering to generate topics. ChatGPT was used for naming these topics, ensuring they were interpretable and accurately reflected the discussions on each platform.

4.6.2 Results

Tables 14 and 15 compare the top five topics from both Discord's "StableDiffusion" and Twitch's "vedal987" during the initial and mid-term data collection periods. These tables highlight key thematic differences between the two platforms, showcasing how Discord and Twitch support different styles of discussions, both initially and as the discussions evolve.

Table 14: Initial Period Comparison of Topics (Discord-StableDiffusion vs. Twitch-vedal987, domain-specific stop words, GSDMM, NER)

Platform	**Topic**	**ChatGPT-Generated Topic Name**	**Top Words**	**Coherence Score**
Discord	Topic 0	File Management and Checkpoints	file, safetensors, ckpt, download, folder	0.73
	Topic 1	WebUI and Scripting Issues	webui, script, extension, issue, account	0.75
	Topic 2	Version and Installation Errors	version, install, error, python, extension	0.71
	Topic 3	VRAM and Colab Setup	vram, colab, xformers, error, card	0.77
	Topic 4	Community Engagement and Collaboration	community, people, post, channel, collaborate	0.70
Twitch	Topic 0	AI Model Training and Debugging	model, train, debug, error, fix	0.72
	Topic 1	Game Strategy and AI Integration	game, ai, strategy, play, integration	0.75
	Topic 2	Community Interaction and Feedback	community, feedback, interaction, response, comment	0.70
	Topic 3	AI Tools and Software Discussion	tool, software, code, development, feature	0.73
	Topic 4	Streamer Collaboration and Events	streamer, collaborate, event, participate, organize	0.78

Table 15: Mid-Term Period Comparison of Topics (Discord-StableDiffusion vs. Twitch-vedal987, domain-specific stop words, GSDMM, NER)

Platform	**Topic**	**ChatGPT-Generated Topic Name**	**Top Words**	**Coherence Score**
Discord	Topic 0	Advanced Model Training Techniques	model, train, dataset, customization, setup	0.80
	Topic 1	Community Projects and Collaboration	project, collaboration, community, effort, contribution	0.78
	Topic 2	Research Papers and Training Information	research, paper, information, training, publication	0.79
	Topic 3	Dataset Issues and Search Functionality	dataset, issue, search, function, error	0.82
	Topic 4	Documentation and User Discussions	documentation, user, guide, discussion, update	0.76
Twitch	Topic 0	Community Interaction and Support	vedal, swarm, update, support, bit	0.78
	Topic 1	Emotional Responses and Celebrations	birthday, sister, party, hug, evil	0.80
	Topic 2	Gaming and Twitch Events	counterstrike, update, game, event, watchalong	0.76
	Topic 3	User Reactions and Moderation	chatter, timeout, watchalong, reaction, coldfish	0.77
	Topic 4	Twitch Viewership and Content Feedback	feedback, viewer, neuro, time, heart	0.79

4.6.3 Analysis

Discord vs. Twitch: Initial Period: In the initial period, Discord's discussions were heavily focused on technical topics, such as handling file formats, VRAM errors, and installation issues. This reflects the early-stage struggles of the community as users focused on troubleshooting and model setup. By contrast, Twitch's initial period centered more on community interaction and entertainment, with discussions on game strategy, AI integration, and collaboration with streamers. The difference in content reflects how each platform's nature shapes its discussions: Discord attracts more technical, problem-solving dialogue, while Twitch focuses on interaction, engagement, and entertainment.

Discord vs. Twitch: Mid-Term Period: By the mid-term period, Discord's discussions had evolved into more advanced topics like model training techniques, research paper sharing, and collaborative projects. This suggests a growing expertise within the community, as users moved beyond basic troubleshooting to focus on innovation and in-depth technical exploration. In contrast, Twitch's "vedal987" channel saw a shift toward more community-driven and emotionally engaging content, with topics such as "Emotional Responses and Celebrations" and "User Reactions and Moderation." This indicates that Twitch's discussions remained more focused on real-time interaction, social engagement, and entertainment, even as the community matured.

Cross-Platform Comparison: The comparison between Discord and Twitch reveals distinct patterns in how discussions about AI-generated content evolve on each platform. Discord, with its structured and community-driven nature, promotes more technical and in-depth discussions, especially as users become more knowledgeable. Twitch, on the other hand, emphasizes real-time interaction, emotional engagement, and community participation, with discussions often centered around events, games, and streamer collaborations. The evolution of discussions on both platforms shows a maturation of topics over time, but the focus on each platform remains true to its core strengths: technical depth for Discord and social engagement for Twitch.

4.6.4 Conclusion

The comparison between Discord's "StableDiffusion" group and Twitch's "vedal987" channel highlights the differing natures of these platforms. Discord fosters more technical and collaborative discussions, while Twitch emphasizes community interaction, emotional responses, and entertainment-focused content. Both platforms saw an increase in coherence scores over time, reflecting more sophisticated discussions, but the thematic focus remained distinct between the platforms. GSDMM, combined with domain-specific stop words and NER, was instrumental in capturing these platform-specific themes, providing high coherence and clear topic identification across both platforms.

4.7 Overall Performance Evaluation

4.7.1 Objective

The objective of this section is to evaluate the overall performance of the topic modeling methods (GSDMM), the use of domain-specific stop words, and the evolution of discussions in AI-generated content communities across multiple channels and platforms. By combining insights from previous experiments, this section offers a comprehensive assessment of how these models and preprocessing methods perform across different types of social media platforms, focusing on AI-driven discussions.

4.7.2 Results Overview

Tables 16, 17, and 18 summarize the performance across platforms (Discord and Twitch), stop word approaches (general vs. domain-specific), and the evolution of topics over time. These tables highlight coherence scores and qualitative findings from the top topics generated using GSDMM with NER across all analyzed channels.

Table 16: Platform Comparison (Discord vs. Twitch, GSDMM, domain-specific stop words, NER)

Platform	Coherence Score Range	Key Findings
Discord	0.70 - 0.82	Technical and collaboration-focused discussions, such as model training and troubleshooting
Twitch	0.72 - 0.80	Emotionally driven interactions and community-centric topics, focusing on gaming, feedback, and events

Table 17: Stop Words Comparison (General vs. Domain-Specific, GSDMM, NER)

Platform	Stop Word List	Coherence Score Range	Key Findings
Discord	General	0.45 - 0.55	Broader, less focused discussions due to general stop word dilution
Discord	Domain-Specific	0.70 - 0.82	Clearer, more focused discussions on technical topics and collaborative efforts
Twitch	General	0.52 - 0.60	Adequate for capturing emotionally charged, community-driven interactions
Twitch	Domain-Specific	0.72 - 0.80	Better differentiation of discussions on community feedback, gaming events, and viewer reactions

Table 18: Temporal Comparison Across Platforms

Platform	Period	Coherence Score Range	Key Findings
Discord	Initial	0.70 - 0.77	General discussions on setup, troubleshooting, and basic usage
Discord	Mid-Term	0.78 - 0.82	Shift to advanced model training techniques, community projects, and research sharing
Twitch	Initial	0.70 - 0.75	Community-focused interactions around gaming, AI debugging, and emotional responses
Twitch	Mid-Term	0.76 - 0.80	Increased focus on community events, emotional support, and collaborative projects

4.7.3 Analysis Across Variables

Platform Comparison: Discord vs. Twitch The platform comparison reveals distinct patterns in how AI-generated content discussions take place on Discord and Twitch. Discord conversations were more technical, focusing on problem-solving, collaboration, and AI model customization. GSDMM performed exceptionally well in these environments, producing higher coherence scores (0.70 - 0.82) when combined with domain-specific stop words. In contrast, Twitch featured more emotionally driven, community-centric interactions, with topics like gaming, viewer reactions, and live event discussions. The coherence scores on Twitch ranged from 0.72 to 0.80, reflecting the platform's emphasis on real-time and emotionally charged interactions.

General vs. Domain-Specific Stop Words The comparison of stop word lists showed that domain-specific stop words consistently improved topic coherence across both platforms. On Discord, domain-specific stop words enhanced the clarity of technical discussions (coherence scores of 0.70 - 0.82), while general stop words introduced noise and diluted the focus. On Twitch, the impact of domain-specific stop words was still notable, helping differentiate between community feedback, gaming events, and emotional reactions, resulting in coherence scores between 0.72 and 0.80. General stop words were sufficient for broader emotional interactions but lacked precision for more detailed discussions.

Temporal Comparison: Initial vs. Mid-Term Both platforms demonstrated an evolution in the complexity and depth of discussions over time. Initially, discussions on Discord centered around setup, error fixing, and basic AI model usage (coherence scores of 0.70 - 0.77). By the mid-term period, these discussions shifted toward more advanced topics, including model training, community collaboration, and research sharing (coherence scores of 0.78 - 0.82). A similar trend was observed on Twitch, where initial conversations focused on general community interaction and AI debugging (coherence scores of 0.70 - 0.75). By the mid-term period, topics evolved to include emotionally driven discussions around community support, gaming events, and collaborative projects (coherence scores of 0.76 - 0.80).

4.7.4 Conclusion: Holistic Performance Evaluation

In conclusion, GSDMM with domain-specific stop words and NER consistently provided superior performance across both Discord and Twitch. The platform comparison revealed distinct patterns of technical discussions on Discord and emotionally driven, community-focused interactions on Twitch. Domain-specific stop words were critical for improving the clarity and coherence of technical topics, especially on Discord. Temporally, both platforms exhibited an increase in topic coherence as discussions matured and became more specialized, reflecting the growing expertise and engagement of their respective communities.

This holistic evaluation highlights GSDMM's effectiveness in capturing diverse discussions across platforms and provides insights into the dynamics of AI-generated content discussions in different social media environments.

5 Discussion and Conclusion

This chapter provides a comprehensive summary of the findings from the experiments conducted using both LDA (Latent Dirichlet Allocation) and GSDMM (Gibbs Sampling Dirichlet Mixture Model) for topic modeling. Additionally, it evaluates the impact of stop word selection (general vs. domain-specific) and temporal shifts in the data collection periods. The results offer valuable insights into the dynamics of AI-generated content discussions and the performance of each model in detecting distinct, interpretable topics across different platforms and community environments.

5.1 Model Performance and Comparison

- **GSDMM vs. LDA**:
  - **LDA**: LDA was effective at capturing broad, overarching themes but struggled to identify granular, context-specific discussions, particularly in platforms with short-form content like Discord and Twitch. Its performance was better suited for long-form text or structured discussions but failed to capture the dynamism of real-time user interactions.
  - **GSDMM**: GSDMM excelled in environments with fragmented, short-form conversations, such as those typical on Discord and Twitch. The model's ability to generate highly focused, distinct topics was evident in its higher coherence scores, particularly in the mid-term period when discussions became more specialized. GSDMM was better equipped to model these nuanced, real-time interactions, offering clearer, more interpretable topics compared to LDA.

5.2 Discussion Topic Evolution

- **Discord (GSDMM)**: The discussions on Discord exhibited a clear evolution from general technical support and troubleshooting in the initial period to more specialized and collaborative topics by the mid-term period. In the early stages, GSDMM captured broad topics such as *file management* and *error fixing*. As the community matured, topics shifted to more complex themes like *advanced model training* and *community collaboration*. The use of domain-specific stop words, combined with GSDMM's capacity to handle fragmented conversations, allowed for better capture of these evolving discussions, highlighting the platform's technical focus.

- **Twitch (GSDMM)**: On Twitch, discussions followed a similar trend but with a focus on emotionally driven, community-centric interactions. Early topics included *game updates* and *community engagement*, but by the mid-term period, the discussions evolved into more nuanced and personal themes, such as *emotional responses to content* and *community events*. GSDMM was particularly effective at identifying these smaller, niche topics, reflecting the platform's emphasis on real-time, socially driven conversations.

5.3 Changes in Discussion Frequency and Engagement

- **Discord (GSDMM)**: As the discussions became more frequent and sophisticated, GSDMM captured an increase in topic diversity and engagement. Early conversations focused on *setup issues* and *tool usage*, while later discussions involved more specific topics, such as *AI model optimization* and *collaborative projects*. This increase in technical complexity coincided with higher user engagement and collaboration.

- **Twitch (GSDMM)**: On Twitch, engagement deepened as users formed stronger emotional connections through the platform. GSDMM identified a transition from casual, content-driven interactions to more intimate discussions, focusing on *personal anecdotes*, *emotional support*, and *community events*. The increasing personal nature of the conversations suggested a growing emotional investment from the user base over time.

5.4 Impact of Stop Words on Topic Clarity

The use of stop words influenced the clarity and coherence of the topics generated by both models:

- **General Stop Words**: General stop words were useful in filtering out common terms but sometimes resulted in the loss of specific, technical vocabulary, especially in discussions on Discord. This led to broader, less focused topics, where LDA, in particular, struggled with overlapping themes.

- **Domain-Specific Stop Words**: Domain-specific stop words enhanced topic coherence, especially in technical discussions on platforms like Discord. GSDMM particularly benefited from this approach, producing clearer and more focused topics such as *error debugging* and *AI training optimization*. On Twitch, domain-specific stop words helped distinguish between emotionally driven conversations, making topics more interpretable.

5.5 Temporal Evolution of Discussions

The comparison of initial and mid-term periods revealed clear temporal shifts in the nature of discussions:

- **Discord**: Initial conversations on Discord focused on basic topics such as *tool setup* and *error troubleshooting*. By the mid-term period, the discussions had evolved into more specialized subjects, such as *community-driven AI tool development* and *advanced model training techniques*. This temporal evolution reflects the growing expertise and engagement of the community.

- **Twitch**: On Twitch, initial discussions revolved around broad, content-focused topics like *game updates* and *community reactions*. By the mid-term period, discussions had become more emotionally charged and personal, with themes such as *community conflicts*, *emotional reactions to content*, and *viewer engagement*. This shift underscores the platform's role in fostering social and emotional connections among users.

5.6 Cross-Platform Comparison: Discord vs. Twitch

Differences were observed between how discussions evolved on Discord and Twitch:

- **Discord**: Conversations on Discord were primarily technical, focusing on AI tool development, troubleshooting, and collaboration. GSDMM, combined with domain-specific stop words, effectively captured the evolution of these discussions from general support to more advanced, collaborative projects.

- **Twitch**: On Twitch, discussions were more socially driven and emotionally focused. GSDMM excelled at capturing real-time, user-specific conversations, reflecting the platform's emphasis on live interaction and community engagement. While domain-specific stop words improved clarity, general stop words performed adequately for capturing broader, emotionally charged topics.

5.7 Limitations and Future Work

While this research has yielded important insights into AI-generated content discussions on Discord and Twitch, several limitations and opportunities for future exploration have been identified. These considerations are critical for advancing the performance of topic modeling techniques, as well as for broadening the scope and depth of future analyses.

5.7.1 Model Limitations and Future Directions

- **Limitations of GSDMM and LDA:** Although GSDMM outperformed LDA in capturing focused, real-time discussions, both models faced challenges when handling short, fragmented social media content. GSDMM's assumption that each message belongs to one topic works well in some cases but falls short when messages are highly ambiguous or multitopic. LDA, on the other hand, struggled with shorter text, often merging distinct discussions into broader, less interpretable themes. Future research could explore more advanced neural topic models such as BERTopic, which uses transformer embeddings to capture semantic meaning. This would allow for a deeper understanding of the nuanced, context-specific discussions prevalent in short-form, dynamic platforms like Discord and Twitch.

- **Keyword-Based Filtering:** One limitation in the data-cleaning process was the use of keyword-based filtering, which, while effective in narrowing the scope of analysis, may have inadvertently excluded valuable discussions on peripheral topics. The current threshold-based filtering might have removed important conversations related to the ethical and societal dimensions of AI-generated content. Future studies could refine this approach by employing semantic-based filtering methods, potentially using contextual embeddings to better capture conversations that are loosely but meaningfully connected to AI-generated content. This would allow for a more holistic analysis of the discussions, including ethical debates and societal concerns.

- **LLM Integration for Topic Generation:** While ChatGPT was instrumental in generating interpretable topic names, future work could integrate large language models directly into the topic generation process. Models like GPT-4 or BERTopic could provide deeper, contextually aware clustering of topics, offering a more nuanced understanding of complex discussions. While resource constraints prevented the full application of GPT-based models in this research, further exploration of these techniques could enhance topic interpretability and relevance. Integrating sentiment analysis into this process could also help capture the emotional and ethical subtexts in discussions, particularly on platforms like Twitch, where user sentiment plays a important role.

5.7.2 Data Scope and Temporal Considerations

- **Expanding the Dataset:** This study focused on a limited number of communities and time periods, which could limit the generalizability of the results. Future work should aim to expand the dataset by including a broader range of AI-generated content communities across different platforms, as well as non-English discussions. Including diverse communities from platforms like Reddit, WeChat, or Douyin could offer additional insights into global perspectives on AI-generated content. Moreover, analyzing longer time spans or performing a week-by-week analysis could reveal finer shifts in community engagement and thematic focus, helping to track the evolution of discussions over time with greater precision.

- **Temporal Granularity:** Although this research included a comparison of initial and mid-term periods, a more granular temporal analysis would provide further insights into how discussions evolve over shorter intervals. Future studies could analyze shifts in topics on a day-to-day or week-by-week basis, allowing for a more dynamic understanding of how community discussions change in response to major AI developments or cultural events. This could be particularly useful in fast-paced environments like Twitch, where user engagement fluctuates significantly over short periods.

5.7.3 Improving Stop Words Selection and Replicability

- **Domain-Specific Stop Words and Replicability:** While domain-specific stop words improved topic coherence and clarity, particularly for technical discussions on Discord, the stop word selection process was partly qualitative, leading to potential biases and replication challenges. To ensure replicability and reduce bias, future research should employ more systematic approaches for selecting stop words, combining automated frequency-based selection with expert input to curate more objective lists. This will allow for a more transparent and replicable methodology, making it easier for future researchers to refine or build upon these findings. Additionally, conducting experiments with varying sets of stop words, including technical and non-technical terms, could help identify the optimal stop word list for specific domains or platforms.

5.7.4 Enhanced Visualization and Interaction

- **Interactive and Real-Time Visualization:** While static visualizations, such as word clouds and co-occurrence networks, provided useful insights, future work could benefit from more interactive and real-time visualizations. These dynamic tools would allow researchers and users to explore evolving trends and shifts in discussions more intuitively. Developing real-time topic modeling dashboards could provide immediate feedback on how community conversations evolve, offering valuable insights into emerging trends, shifts in sentiment, or sudden spikes in engagement.
- **Cross-Platform Visualization:** Another future direction involves building tools that compare and visualize discussions across multiple platforms simultaneously. This could be particularly valuable for understanding how discussions migrate from one platform to another or how the same AI-generated content is perceived differently in various online communities.

This research provided a detailed exploration of AI-generated content discussions using GSDMM and LDA for topic modeling on Discord and Twitch, revealing critical insights into the dynamics of online conversations. However, several limitations, such as the need for more advanced models, more nuanced data filtering techniques, and improved stop word selection, offer areas for future improvement. By integrating transformer-based models, refining semantic filtering techniques, and expanding the dataset across platforms and languages, future research can provide a more comprehensive and precise understanding of AI discourse across online communities.

5.8 Discussion and Future Directions

The rapid growth of AI-generated content (AIGC) has introduced both opportunities and ethical challenges, sparking global debates within creative and social domains. While AIGC technologies like text-to-image models and large language models (LLMs) have the potential to transform content creation, their widespread adoption has raised critical concerns regarding authorship, ownership, misinformation, and job displacement. As these technologies evolve, it is essential to address these challenges through comprehensive ethical frameworks, informed research, and effective policy implementation.

5.8.1 Ethical Considerations in AIGC

One of the most pressing ethical issues surrounding AIGC is the question of authorship and ownership. Traditionally, creativity has been viewed as a human endeavor, grounded in intention, originality, and emotional expression. However, AI systems such as GPT-3 and text-to-image models like DALL-E challenge these notions by autonomously generating content that rivals human creativity [28, 6]. This raises the question of who, if anyone, owns the rights to AI-generated works. Legal frameworks, such as the European Union's Copyright Directive, struggle to keep pace with these developments, as it remains unclear whether AI-generated works can be protected under traditional copyright law or if the creators of the AI systems should be granted ownership rights [31].

The proliferation of AIGC on digital platforms also heightens the risk of misinformation. Deepfakes, powered by Generative Adversarial Networks (GANs), are particularly concerning, as they enable the creation of realistic but entirely fabricated media content. This technology has been weaponized in political disinformation campaigns, contributing to the erosion of public trust in media and institutions [8]. The real-time and rapid dissemination of AI-generated content across social media platforms like Twitch and Discord present additional challenges, as traditional content moderation systems struggle to keep up. Researchers and technologists must develop more advanced detection mechanisms to identify and flag AI-generated misinformation before it can cause widespread harm [34].

Another critical issue is the potential displacement of human labor in creative industries. AIGC technologies can automate tasks that were once the domain of human workers, such as graphic design, writing, and animation [12]. While this automation democratizes access to high-quality creative tools and lowers the barrier for entry into creative fields, it also threatens traditional job roles. In particular, industries that rely on skilled labor, such as digital marketing, game development, and film production, may see job losses as AI systems become more capable of producing professional-quality outputs with minimal human intervention. Balancing the benefits of AIGC with its potential impact on employment requires careful consideration from both researchers and policymakers [16, 10].

5.8.2 Future Research Directions

As AIGC continues to expand its influence, future research must focus on developing ethical frameworks and practical applications that align with societal values. Scholars such as Bright et al. and Hollywood et al. emphasize the importance of establishing robust frameworks for utilizing social media analysis tools, particularly in domains like law enforcement, content moderation, and digital rights protection [5, 15]. Future studies should not only investigate the ethical implications of AIGC but also explore how these technologies can enhance user experiences on platforms like Twitch and Discord.

A key area for future research is improving the accuracy and transparency of AI models to ensure their responsible use in digital spaces. This includes refining moderation tools capable of detecting deepfakes, misinformation, and other harmful content. Collaboration between researchers in AI ethics, social media analysis, and law enforcement is essential to developing models that effectively safeguard against misinformation while preserving freedom of expression [34, 24].

Furthermore, understanding how AIGC influences user behavior and community dynamics remains an important research focus. As platforms like Twitch and Discord continue to host vibrant communities around AI-generated content, studies must investigate how these technologies affect user engagement, community formation, and content generation. By analyzing the real-time and asynchronous discussions on these platforms, researchers can gain deeper insights into how AIGC is reshaping the cultural landscape and influencing public discourse [22, 3]. Interdisciplinary collaboration will be critical in this effort, combining insights from AI development, social science, and creative fields to ensure that AIGC technologies are used ethically and responsibly, fostering human creativity and community engagement rather than replacing it.

In conclusion, the future of AIGC lies at the intersection of innovation, ethics, and policy. By addressing the societal and legal challenges posed by these technologies, researchers and policymakers can help shape an AI-driven future that enhances creativity while safeguarding against potential risks.

5.9 Conclusion

This thesis has provided a comprehensive analysis of discussions surrounding AI-generated content on two distinct platforms—Discord and Twitch—by applying and evaluating topic modeling techniques, specifically LDA (Latent Dirichlet Allocation) and GSDMM (Gibbs Sampling Dirichlet Mixture Model). The findings consistently demonstrated that GSDMM outperformed LDA in capturing more granular, real-time discussions, particularly in fragmented and fast-paced environments like Twitch, where emotionally charged, user-specific interactions dominated. In contrast, LDA proved more effective at identifying broader thematic trends, particularly in longer, structured discussions common on Discord.

A critical insight from this research is the evolving nature of discussions over time. Initial conversations on both platforms largely focused on general topics such as tool setup and community interaction. However, as the communities matured, discussions became more specialized, with Discord communities emphasizing technical aspects like model training and error troubleshooting, while Twitch conversations transitioned toward more personal, emotionally driven topics. This temporal evolution highlights the growing expertise and engagement of users as they become more comfortable with AI technologies, leading to more sophisticated and collaborative discussions.

The comparison between general and domain-specific stop words further underscored the importance of tailored preprocessing techniques, particularly in technical communities like Discord. By refining stop word lists to focus on domain-specific terms, the topic models were able to generate clearer, more coherent topics, especially in contexts requiring precise technical language. This approach allowed GSDMM to shine in capturing highly focused discussions on both platforms, offering valuable insights into user behavior and platform dynamics.

Despite its strengths, the research also revealed several limitations in the models applied. Both GSDMM and LDA faced challenges when handling short, fragmented text, and the need for more advanced models such as BERTopic or large language models became evident. These models could potentially capture the nuanced, context-specific discussions that were sometimes lost with the current approaches. Furthermore, the keyword-based filtering method used in data preprocessing may have excluded valuable discussions on peripheral topics, particularly ethical debates surrounding AI, which could offer new dimensions to the analysis.

In conclusion, this thesis contributes to the growing understanding of AI-generated content discussions by combining quantitative topic modeling techniques with qualitative insights. GSDMM's clustering approach, particularly when paired with domain-specific stop word filtering, has proven highly effective for analyzing fragmented, context-driven conversations on platforms like Discord and Twitch. Future research could focus on integrating more sophisticated models, expanding the dataset to include a wider range of communities, and refining the temporal granularity of the analysis to capture more dynamic shifts in discussions. Ultimately, this work underscores the importance of adapting topic modeling methods to the characteristics of different platforms and communities in order to fully understand the complexities of digital discourse in the age of AI.

References

[1] Rasheda Alghamdi and Khalid Alfalqi. A survey of topic modeling in text mining. *International Journal of Advanced Computer Science and Applications*, 6(1):147–153, 2015.

[2] Fayya Anyatasia. Investigating motivation and usage of text-to-image generative ai for creative practitioners. Master's thesis, University of Helsinki, December 2023. Faculty of Science.

[3] Mila Anyatasia. The impact of ai on digital art: A quantitative analysis of creative communities. *International Journal of AI and Creativity*, 10:112–130, 2023.

[4] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[5] Jonathan Bright and Thomas Heidenreich. The role of social media in law enforcement research. *Police Quarterly*, 2014.

[6] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[7] Max Böhmer, Ralf Paul, and Volker Eisenlauer. The role of chat in shaping twitch streaming dynamics. *Proceedings of the ACM Conference on Human-Computer Interaction*, 2019.

[8] Robert Chesney and Danielle Keats Citron. Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Affairs*, 98:147–155, 2019.

[9] Patrick Ching and Wei Li. Deepfakes: The impact of ai-generated videos on social media trust. *Social Media Research Quarterly*, 7:33–48, 2021.

[10] John Danaher. The threat of job automation by artificial intelligence: An overview. *AI and Society*, 32(2):245–256, 2017.

[11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1:4171–4186, 2019.

[12] Ahmed Elgammal. Ai and art: The deep learning algorithms transforming creative production. *Communications of the ACM*, 64(2):92–99, 2021.

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27:2672–2680, 2014.

[14] William A. Hamilton, O. Garretson, and A. Kerne. Streaming on twitch: Fostering participatory communities of play. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1315–1324, 2014.

[15] John S Hollywood, Michael JD Vermeer, Dulani Woods, Sean E Goodison, and Brian A Jackson. Using social media and social network analysis in law enforcement: Creating a research agenda, including business cases, protections, and technology needs. *RAND Corporation*, 2018.

[16] Julia Howard. The role of ai in digital art and media. *Journal of New Media Research*, 16(1):112–127, 2022.

[17] Hyejin Kim and Jungsoo Kim. Sentiment analysis in social media: Challenges and future directions. *ACM Computing Surveys*, 55(3):1–30, 2023.

[18] H. Koo, J. Kim, and Y. Song. Community engagement on discord: How structured conversations enable deeper discussions. *Journal of Digital Communities and Communication*, 2021.

[19] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[20] Iman Naim and Leila Sadeghi. The rise of deepfakes: Implications for media and trust. *Journal of Media Ethics*, pages 15–30, 2018.

[21] Jonas Oppenlaender. Perceptions and challenges of ai-generated art. *International Journal of Human-Computer Studies*, 170:102–114, 2023.

[22] Jonas Oppenlaender, Panos Papadopoulos, and Drazen Adamovic. Taxonomy and perceptions of ai art: How generative models are reshaping creativity. *Computers in Human Behavior*, 150:106890, 2023.

[23] Long Ouyang, Jeffrey Wu, Xu Jiang, Carroll L Almeida, Christine Wainwright, Pamela Mishkin, Chong Zhang, et al. Training language models to follow instructions with human feedback. *arXiv preprint arXiv:2203.02155*, 2022.

[24] Frank Pasquale. Ai ethics and the role of transparency. *Harvard Law Review*, 31(2):40–45, 2018.

[25] Joe Pritchard and Alex Dong. Gsdmm: Short text topic modeling. *Proceedings of the ACM Conference on Information and Knowledge Management*, 2017.

[26] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. *International Conference on Machine Learning*, 139:8748–8763, 2021.

[27] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. 2018.

[28] Alec Radford, Jeffrey Wu, Dario Amodei, Jack Clark, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

[29] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21:1–67, 2020.

[30] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, et al. Zero-shot text-to-image generation. *arXiv preprint arXiv:2102.12092*, 2021.

[31] Lucy Samuelson and Charlie Brown. Eu regulation on artificial intelligence: The implications for deepfake technology. *European Journal of Law and Technology*, 11:1–21, 2020.

[32] Joshua Stein. Paradigms of social media interaction: Real-time communication and ai integration. *Digital Media Studies*, pages 102–122, 2023.

[33] Romal Thoppilan, Daniel De Freitas, Jason Hall, Noam Shazeer, Apoorv Kulshreshtha, Hongkun Cheng, Alicia Jin, Tuomas Bos, Simon Baker, Nan Du, et al. Lamda: Language models for dialog applications. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5835–5852. Association for Computational Linguistics, 2022.

[34] Cristian Vaccari and Andrew Chadwick. Deepfakes and disinformation: Exploring the impact on trust and democracy. *Journal of Internet Law*, 24(3):9–14, 2020.

[35] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[36] Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R Bowman. Glue: A multi-task benchmark and analysis platform for natural language understanding. *arXiv preprint arXiv:1804.07461*, 1:1–15, 2018.

[37] Chong Wang and David M. Blei. Topics over time: A non-markov continuous-time model of topical trends. *Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 424–432, 2012.

[38] John Xie, Lu Zhou, and Jiang Gao. Conversational ai: System design and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 34:987–1012, 2023.

[39] Jianhua Yin and Jianyong Wang. Dirichlet multinomial mixture model for topic modeling on short texts. *Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 2139–2148, 2014.

A Use of ChatGPT in Thesis Writing

In compliance with the official policy of Utrecht University, the use of generative AI tools in the writing of this thesis has been restricted to correction and editing purposes. ChatGPT was employed specifically to assist in naming topics identified through topic modeling and in conducting additional background research to clarify complex areas. No generative AI tools were used for the writing of original content, except for refining language and improving clarity. The insights and contributions provided by ChatGPT were carefully reviewed and validated to ensure they align with the academic integrity standards set by the university.