National University of "Kyiv-Mohyla Academy"

Kyiv — 2025

Report Team # 17

Homework #5. Professional Behavioural Research Report

Done by Sofiia Zakharuk

Kyiv — 2025

Executive Summary

Project Overview

This report presents a comprehensive behavioral analysis of the impact of Generative AI on the creative sector. By examining digital footprint data over a six-year period, we track how creative professionals – artists, writers, and designers – have reacted to, adopted and resisted the introduction of AI technologies.

Dataset

The analysis is based on a large-scale dataset of 2.5 million interactions extracted from 10 professional creative communities on Reddit (Art, illustration, DigitalArt, painting, graphic_design, photography, writing, copywriting, blender, gamedev).

To ensure objective comparability and eliminate seasonal bias, the study utilizes five longitudinal snapshots: September 2018, September 2021, May 2023, October 2023, and September 2024.

Data source: Reddit Public Corpus ([Academic Torrents](#))

Main Research Questions

Catalysts: did the disruption correlate with general AI awareness or specific visual capabilities?

Evolution: how has the professional discourse shifted from the initial shock of 2023 to the new normal of 2024?

Polarization: is the creative workforce reacting as a monolith, or is there a structural divide between artistic and technical disciplines?

Top-Level Findings

1. Visual disruption (activity timeline). Contrary to general tech industry trends, the release of text-based models (GPT-3 in 2020) had a minimal impact on creative community activity. The true disruption occurred in late 2023, driven directly by the mass accessibility of text-to-image tools like Midjourney and Stable Diffusion. Activity peaked in October 2023, confirming that visual generative AI poses a more immediate perceived threat to the creative workforce than textual AI.

2. From panic to integration. The discourse has evolved through two distinct phases:

The Crisis Phase (2023). Dominated by existential anxiety. Semantic analysis confirms that "copyright infringement" was a top recurring phrase in professional debates, while the broader "Ethics and Jobs" category (encompassing concerns about displacement and professional rights) reached its absolute peak frequency in October 2023.

The Normalization Phase (2024). Data reveals a qualitative shift in professional discourse. While ethical concerns remain present, the conversation has largely pivoted from abstract fear to practical implementation. This is evidenced by a significant rise in technical terminology: our

analysis shows that terms like "workflow," "Adobe firefly," and "Generative fill" have become dominant. The high frequency of N-grams such as "using adobe" and "industry standard" indicates that professionals are no longer rejecting AI entirely, but are selectively adopting legally compliant tools that fit into existing production pipelines without compromising copyright safety. This signals a transition to a new normal where AI is treated as a software feature rather than an existential threat.

3. Hidden segmentation and discourse depth. While automated sentiment analysis scores the discussions as neutral, deeper investigation reveals a sharp professional divide. Visual art communities remain the epicenter of ethical resistance, whereas technical communities demonstrate a higher readiness to adopt AI as a pipeline tool. Furthermore, engagement quality is high: comments on AI topics are 3x longer than general comments, indicating deep cognitive involvement rather than low-effort noise.

Why It Matters

The shock phase of AI adoption has passed. The creative industry is not rejecting AI entirely but is demanding legal safety and workflow integration. The data suggests that professionals are moving away from open generative models towards controlled, industry-standard tools that mitigate copyright risks. Future strategies must address this bifurcation: protecting the rights of "resisters" while empowering the workflows of "adopters."

Dataset Description & Methodology

Data Source & Scope

To capture the authentic voice of the creative community, we constructed a specialized dataset derived from the Reddit Public Corpus. The data was filtered to include only interactions from 10 distinct subreddits representing key creative industries:

Visual Arts	r/Art, r/DigitalArt, r/illustration, r/painting, r/photography
Design & 3D	r/graphic_design, r/blender, r/gamedev
Writing	r/writing, r/copywriting

Temporal Sampling Strategy

Instead of a continuous timeline, we utilized a longitudinal snapshot approach to ensure data comparability and isolate the impact of specific AI milestones. The dataset consists of five specific one-month periods:

Baseline (Pre-AI): September 2018, September 2021

The "Shock" Period: May 2023, October 2023

The "New Normal": September 2024

Note: September was chosen as a recurrent reference month to minimize seasonal variance (summer holidays or holiday season spikes).

Data Volume & Composition

After cleaning and filtering, the final dataset comprises 2,379,400 interactions:

Comments: 2,073,391

Submissions (Posts): 306,009

This volume ensures statistical significance for high-level trend analysis and sufficient granularity for semantic pattern recognition.

Data Processing Pipeline

The raw data (zst-compressed) underwent a multi-stage cleaning process using Python:

1. Extraction. Parsing raw text lines to separate metadata (timestamp, author, score) from content.

2. Normalization. Converting text to lowercase for NLP analysis.

3. Categorization. Comments were tagged based on custom keyword dictionaries.

Analytical Methods

We employed a mixed-methods approach to interpret the data:

1. Exploratory data analysis. To track volume dynamics and engagement metrics (comment length, scores).

2. N-gram analysis. Extracting the most frequent bigrams and trigrams to identify dominant semantic themes.

3. Comparative Topic Modeling. Analyzing the distribution of keyword groups across different subreddits to map community polarization.

Limitations

1. Sentiment analysis bias. Automated tools (VADER) struggled to detect professional anxiety, often classifying legal discussions as neutral. Therefore, we relied on semantic N-gram analysis as a more accurate proxy for community sentiment.

2. Platform specificity. The data reflects the demographics of Reddit users and may not fully represent the global offline creative workforce.

Research Storytelling

The Timeline of Disruption: It Was Never About Text

Our initial hypothesis was that the release of GPT-3 in 2020 would trigger the first wave of AI discussions in creative communities. The data refutes this. As shown in Figure 1, activity remained relatively stable throughout the "Text AI" era (2020–2021). The true inflection point occurred in mid-2022, correlating perfectly with the open beta releases of Midjourney and Stable Diffusion. The discussion volume surged, reaching its absolute peak in October 2023. Creative professionals did not perceive text generation as an existential threat. The disruption was driven almost exclusively by visual generative AI. The sustained high activity levels in 2024 (dropping only slightly from the peak) indicate that this is not a temporary trend but a permanent shift in the industry's attention economy.
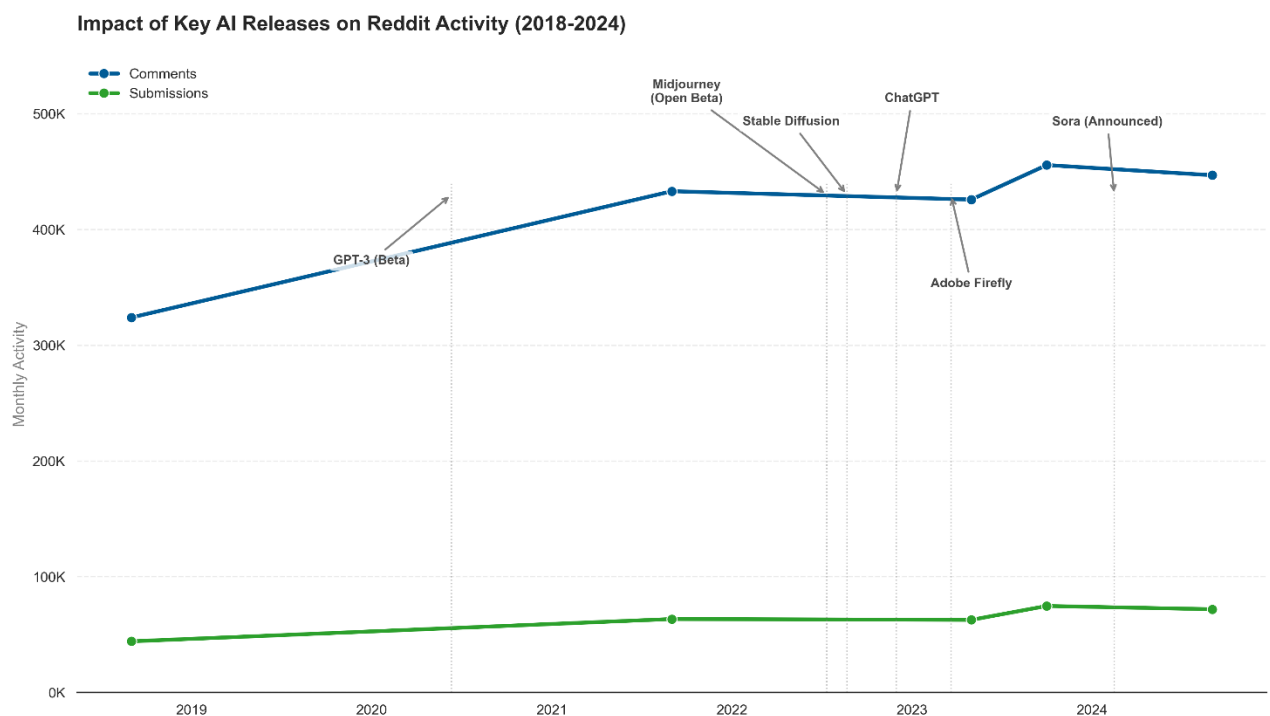

Figure 1. Community activity surged in direct correlation with the public release of text-to-image tools, while early text models (GPT-3) had minimal impact.

The Evolution of Discourse: From Panic to Integration

What are professionals actually discussing? By tracking keyword groups over time, we identified a clear transition in community sentiment and priorities.

The Crisis Phase (2023) As illustrated in Figure 2, the red line representing the "Ethics & Jobs Debate" dominates the conversation in 2023. This topic grew faster than any other category. To understand the specific nature of this anxiety, we analyzed the most frequent N-grams. Figure 3 reveals that the discourse was centered on legality and rights. Terms like "copyright infringement" and "copyright law" appear with high frequency.
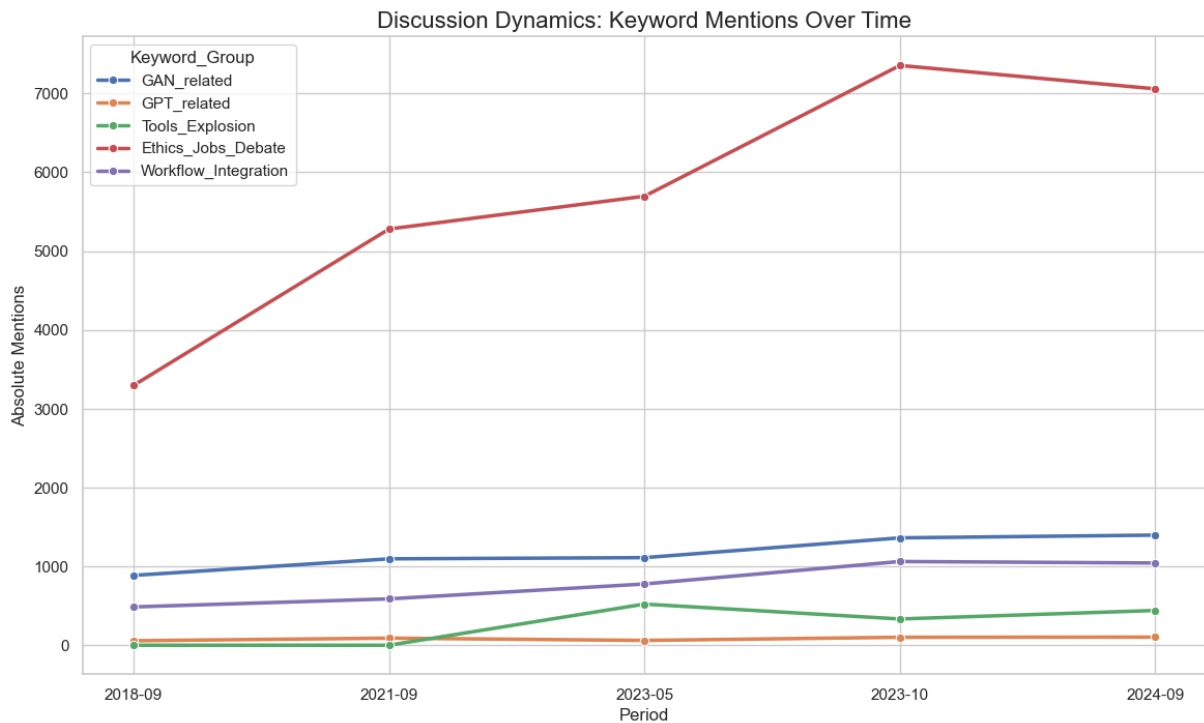
Figure 2. Discussion Dynamics: The "Ethics & Jobs" debate (red line) dominated the conversation in 2023, overshadowing technical discussions.

While ethical concerns remain high, 2024 data reveals a qualitative shift. The "Workflow Integration" topic (purple line in Figure 2) shows steady growth. Crucially, N-gram analysis (Figures 3, 4) highlights a rise in pragmatic terminology: "using Adobe", "Generative Fill". The industry is moving from a state of panic (focusing on unregulated models and theft) to integration (focusing on legally compliant tools like Adobe Firefly). Professionals are no longer just debating the existence of AI – they are figuring out how to use it safely.
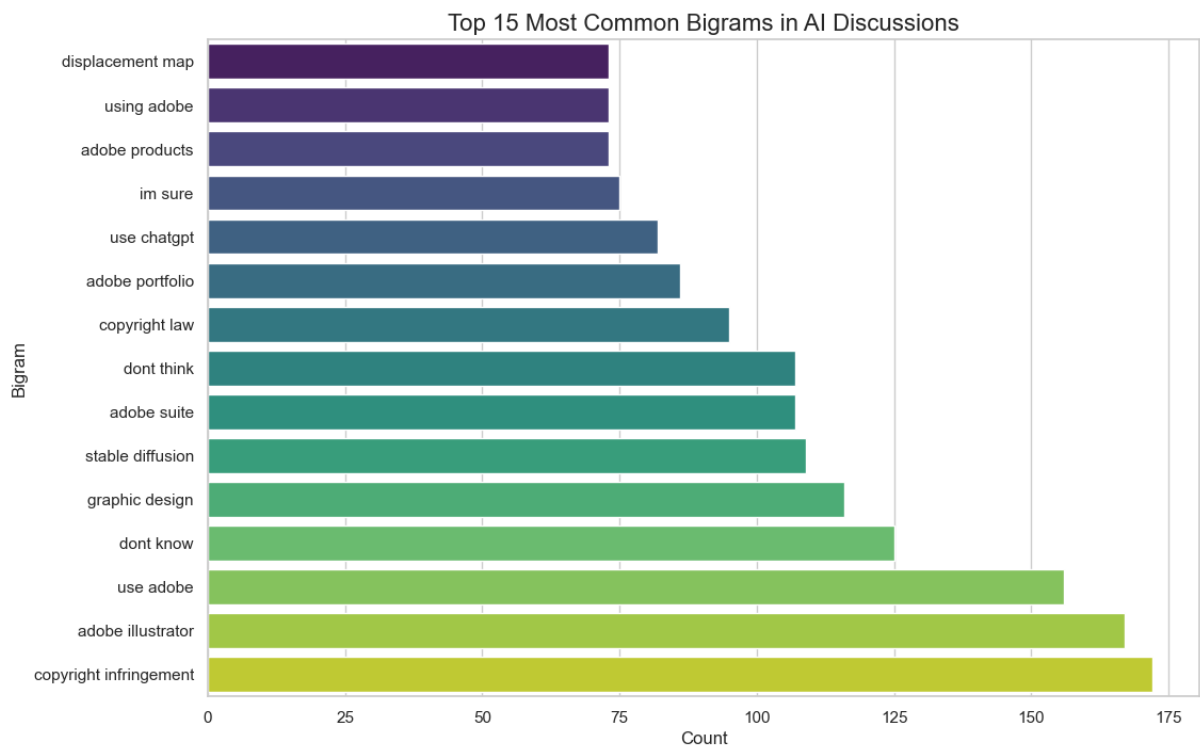

Figure 3. Top Bigrams reveal that "Copyright" and "Adobe" are the two central pillars of the conversation.

Community Segmentation: Two Different Worlds

A key finding of this research is that distinct creative sub-disciplines have unique "behavioral fingerprints." Figure 4 visualizes the normalized distribution of topics, revealing three specific community clusters.
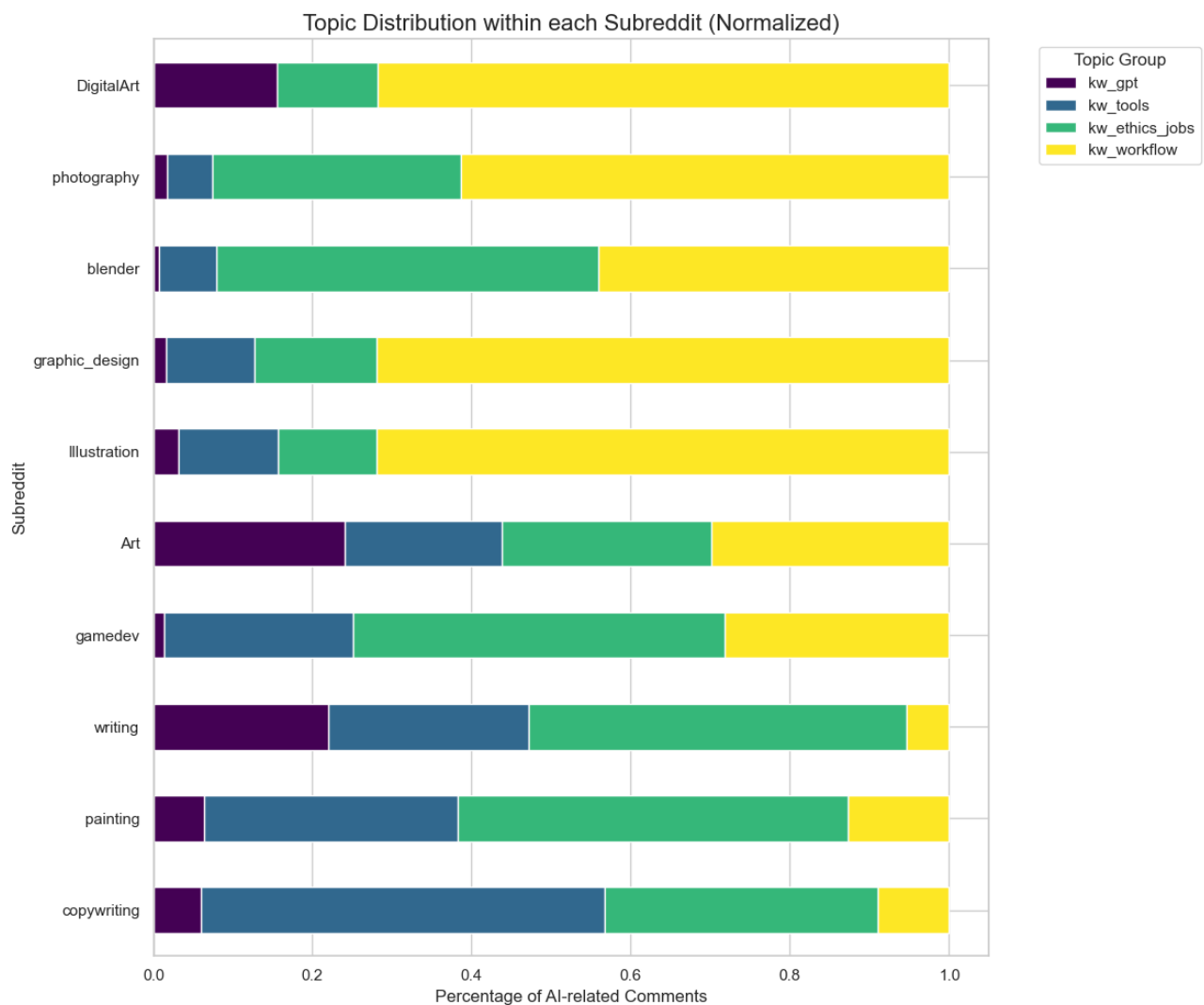

Figure 4. Topic Distribution: r/DigitalArt is dominated by Workflow (Yellow), while r/Art and r/writing show heavy engagement with Plagiarism/GPT (Purple) and Ethics (Green).

1. The "Workflow" Cluster (Adopters). Communities like r/DigitalArt, r/photography are dominated by the yellow "Workflow Integration" segment. For example, in r/DigitalArt, workflow topics account for over 50% of the discussion. These communities appear to be the most pragmatic, focusing on how to apply tools (like Adobe Firefly) rather than debating their existence.

2. The "Displacement" Cluster (Anxious Professionals). Surprisingly, r/gamedev and r/writing show the largest proportional shares of the "Ethics & Jobs" debate (green segment). This indicates that writers and game developers – who rely heavily on intellectual property and assets – feel the threat of professional displacement most acutely.

3. The Text vs. Visual Split. Why is the "GPT" (Dark Purple) segment so high in r/Art? This is driven by the keyword "plagiarism," which falls into this category. While writers in r/writing discuss GPT-3 as a tool, artists in r/Art are using this semantic space to debate style mimicry and intellectual

property theft. This confirms that for "traditional" creatives (writers and fine artists), the technology is primarily framed through the lens of integrity and rights, rather than utility.

Engagement Quality: Deep Debate, Not Spam

Figure 5 presents a striking finding: comments related to AI are, on average, 3x longer (approx. 225 characters) than the historical average for these subreddits (approx. 75 characters). This high "cognitive load" suggests that users are engaging in complex, nuanced arguments rather than posting short reactions.
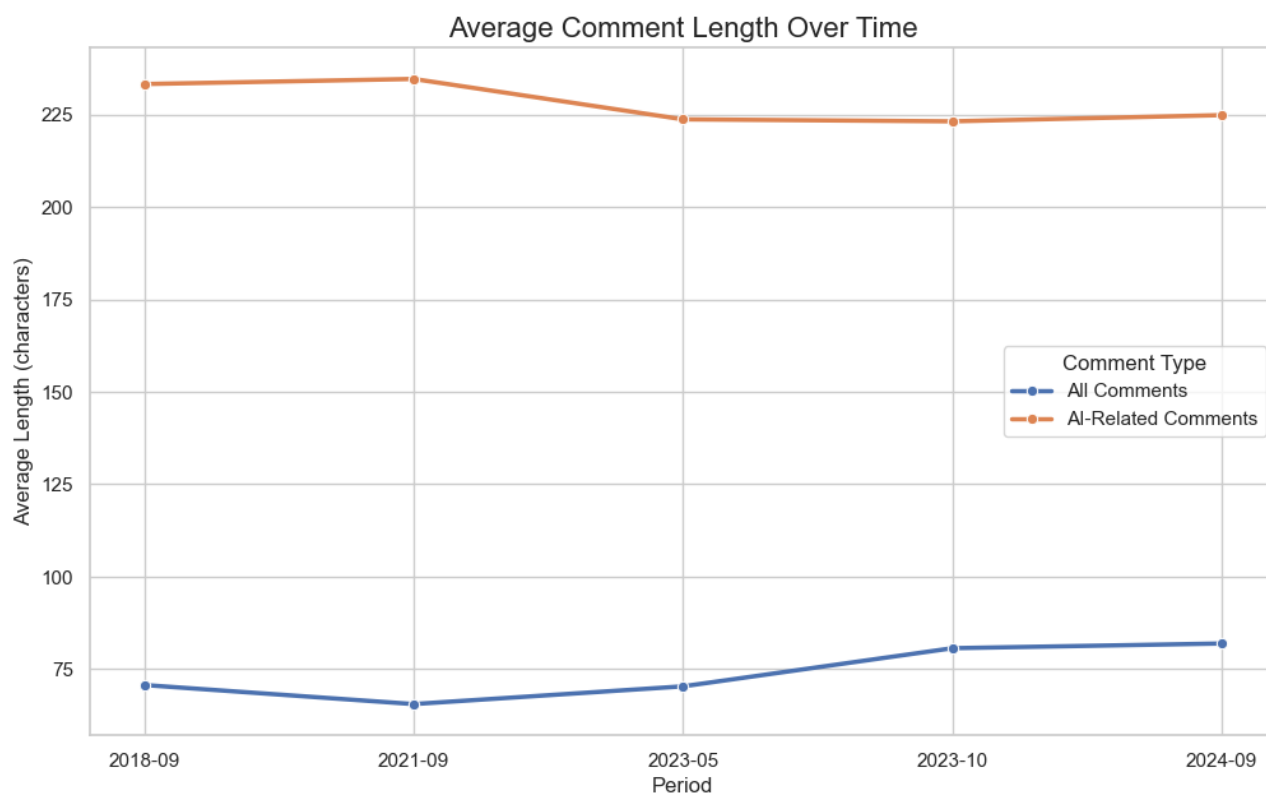

Figure 5. AI-related comments (orange line) are consistently longer and more detailed than the average community contribution.

Rejection of "Showcase" Content However, the community clearly distinguishes between talking about AI and seeing AI art. Figure 6 shows that posts showcasing "Tools" (AI-generated images) receive significantly lower engagement scores (~5) compared to posts discussing "Ethics" (~48). The community wants to debate the technology, but they largely reject AI-generated content itself.
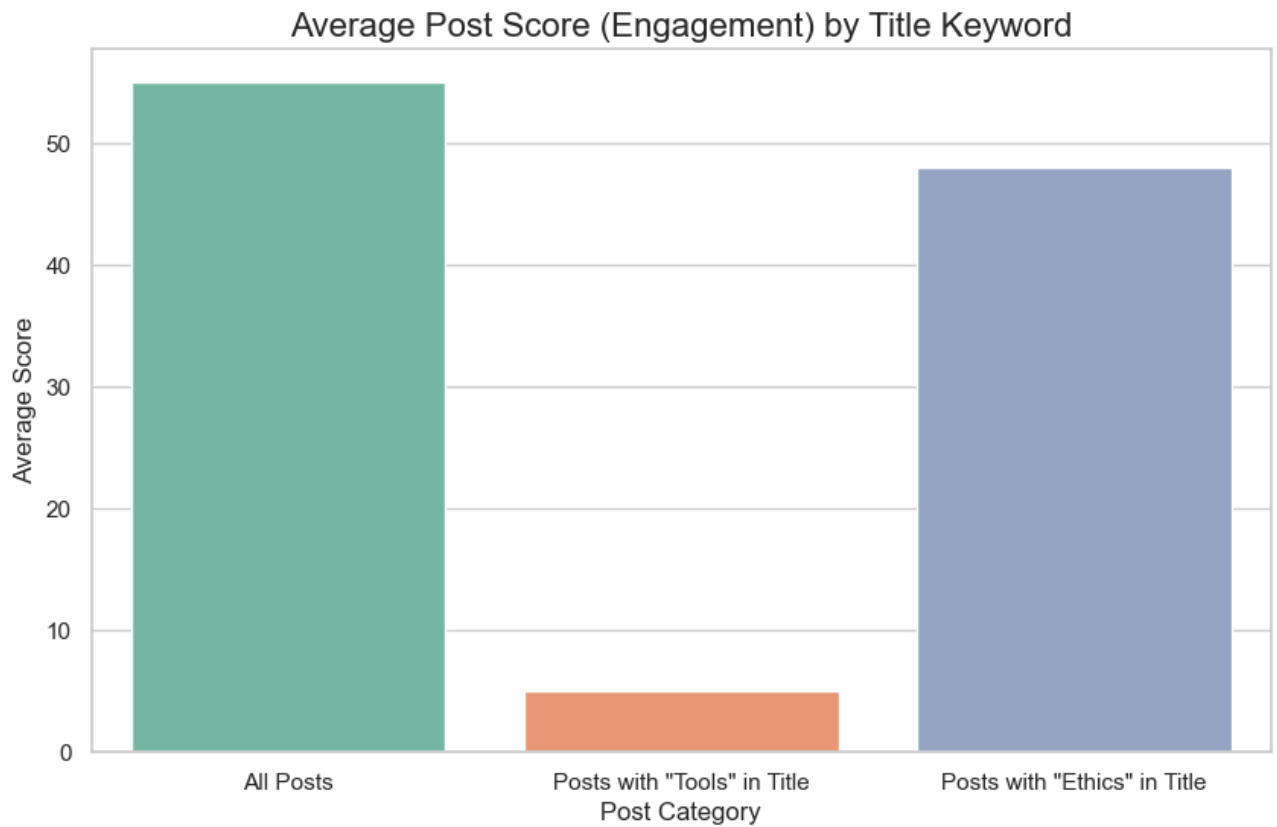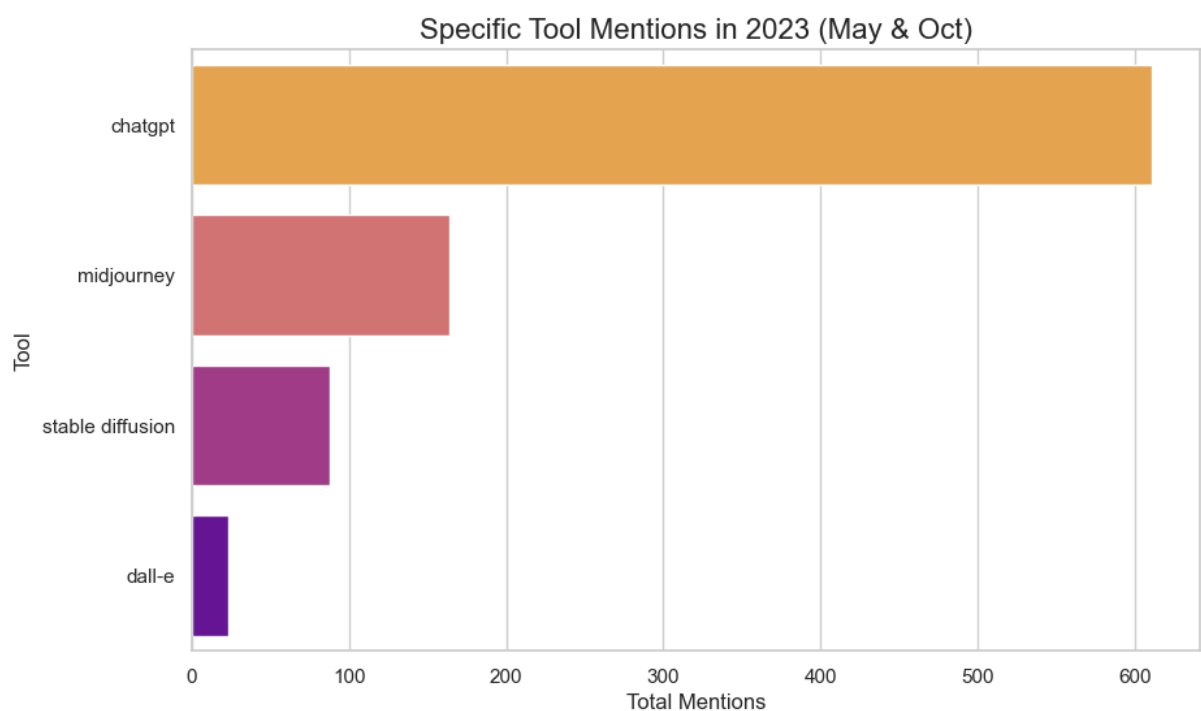
Average Post Score (Engagement) by Title Keyword

Figure 6. Engagement Paradox: Users actively debate Ethics (high scores) but ignore or downvote posts showcasing AI Tools (low scores).

Professional Visualizations

1. The "Elephant in the Room" (Tool Dominance)

Dominance of Specific Tools (2023). While the disruption is visual, the terminology is dominated by ChatGPT. It appears nearly 3x more frequently than image generators (Midjourney, Stable Diffusion), suggesting it has become the generalized proxy term for "Generative AI" across all creative disciplines.

## 2. The Shift in Influence (Share of Voice)
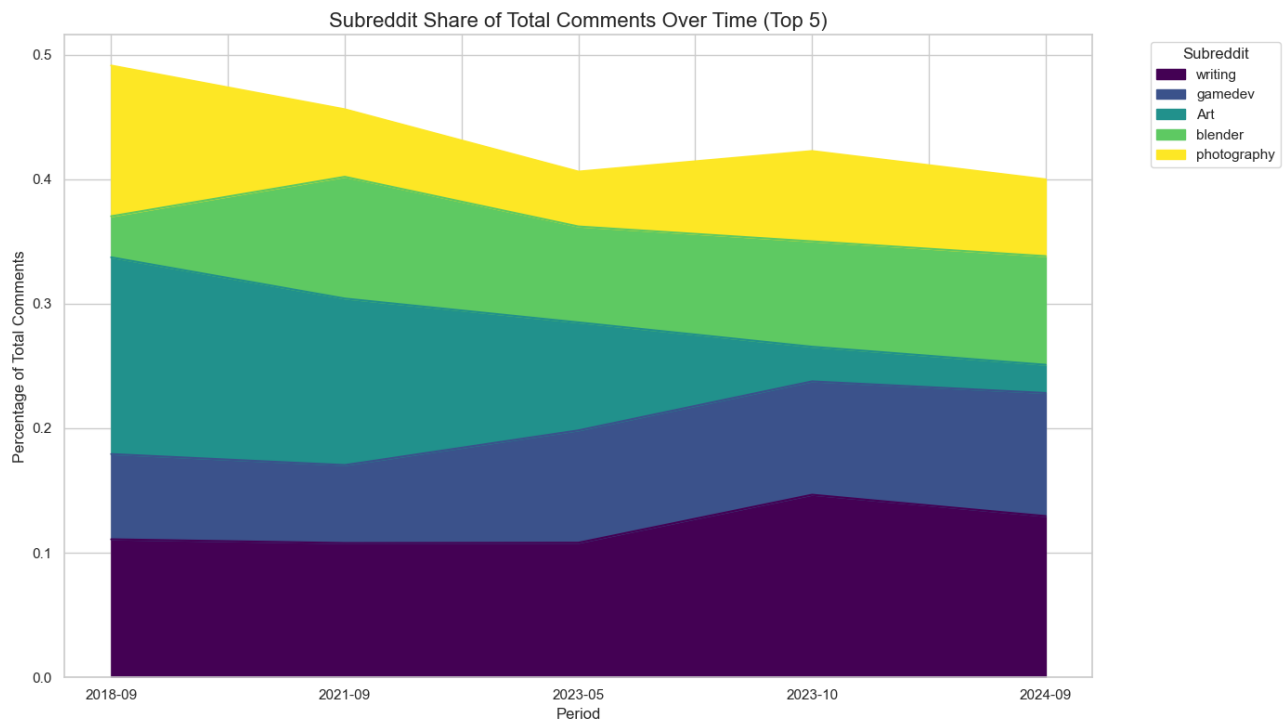
Evolution of Community Influence (Share of Voice). Historically, textual communities (r/writing, purple area) dominated the discourse volume. However, the 2023 AI boom caused a structural shift: visual communities (r/Art, r/blender) significantly expanded their share of the total conversation, proving that the visual AI crisis mobilized users more aggressively than text AI.

3. The "Pre-History" (Early Adopters)

Early Adopters (2018–2021). Before the mainstream boom, AI discussion was highly niche. r/writing was the primary hub for early GPT discussions, while r/Art showed minimal interest in early visual tech (GANs). This explains why the sudden arrival of high-quality generators in 2022 caused such a shock – the visual community was largely unprepared compared to writers.

Early Adopters: GAN vs. GPT Mentions (2018-2021)
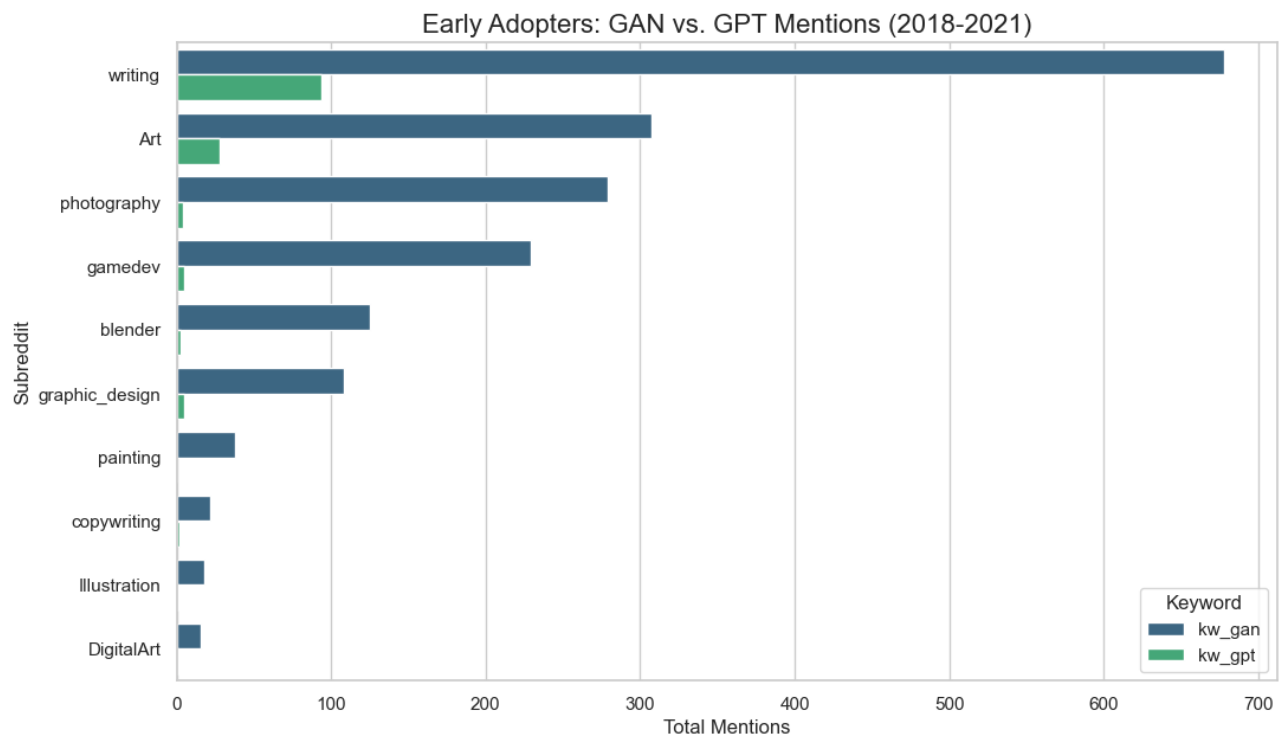
4. The "Mod War" (Specific Trigrams)

Top 15 Most Common Trigrams in AI Discussions
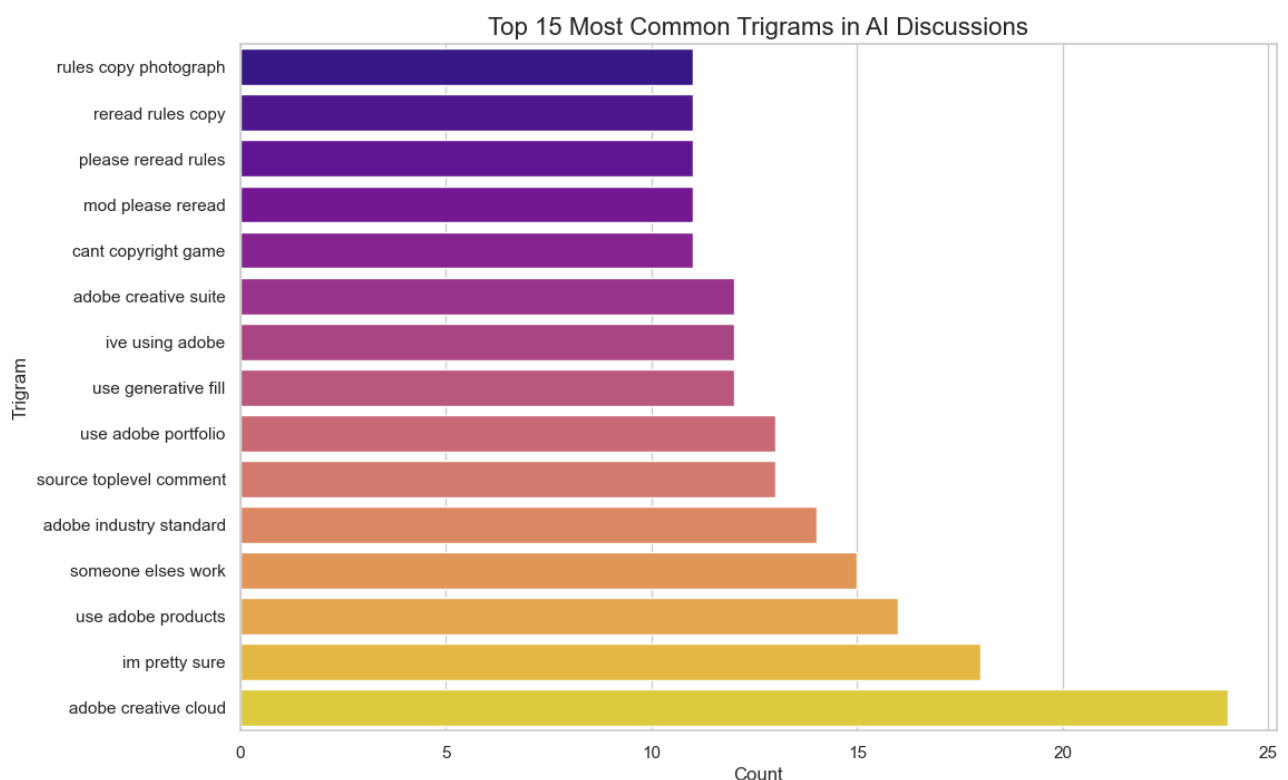
Operationalizing the Debate (Top Trigrams). Deeper semantic analysis reveals a struggle over community governance. Phrases like "rules copy photograph" and "mod please reread" indicate intense conflicts regarding subreddit rules and moderation policies concerning AI content.

Conclusions

Summary of Findings

Our analysis of 2.5 million interactions confirms that the creative industry has undergone a fundamental structural transformation. The "AI Revolution" on Reddit was not a gradual adoption curve but a violent disruption triggered specifically by visual generative tools. The data reveals a clear psychological progression: from the initial shock and existential panic of 2023 (dominated by "copyright" and "displacement" rhetoric) to a tentative stabilization in 2024, characterized by the integration of AI into established workflows (Adobe).

Patterns & Anomalies

The visual bias anomaly. A striking finding is the disproportionate reaction to visual AI compared to text AI. The release of GPT-3 in 2020—arguably as revolutionary as Midjourney, barely registered in creative communities. This suggests that visual professionals feel significantly more vulnerable to automation than writers, or that visual mimicry triggers a stronger emotional response than text generation.

The engagement paradox. While community sentiment is defensive, engagement metrics are paradoxically high. AI-related discussions generate comments that are 3x longer than the historical average. The community rejects the product (AI art posts often have low engagement scores) but is deeply invested in the process of debating its implications.

The Most Important Insight

The single most critical finding is the structural bifurcation of the creative workforce. We are not seeing a unified "Creative Industry" reaction, but rather two distinct realities:

Ethical resisters. Focused on copyright protection and maintaining human-centric definitions of art.

Pragmatic adopters. Focused on efficiency, pipeline optimization and using AI as a "gluing" tool (text generation, coding assistance). Strategies that treat these two groups as a monolith are destined to fail.

Unanswered Questions

Demographic nuance. Due to Reddit's anonymity, we cannot correlate sentiment with age or professional seniority. It remains unclear if the resistance is generational (older professionals vs. younger students).

Future Research Directions

With more time and resources, the next logical step would be User Clustering. By tracking individual user histories, we could segment the population into "Pro-AI," "Anti-AI," and "Convert" cohorts to quantify exactly how many professionals have changed their stance from resistance to adoption.