

Never Care For What They Say? Platform vs Genre Rules in Online Horror Narratives (2007–2024)

Alexandre Lionnet-Rollin

École Pratique des Hautes Études, Université PSL
Paris, France

Florian Caffero

LRE, EPITA
CJM, École nationale des chartes, Université PSL
Paris, France

Abstract

Research on online cultural production shows that platforms are acting as mediators that can heavily shape textual form. Yet, empirical work is often platform-bounded, making it difficult to assess whether stylistic regularities that we observe are indeed genre signals or if some of them are platform artefacts. We address this question through a cross-platform design focused on *creepypasta*, a digital-born horror genre circulating across heterogeneous infrastructures. Using a corpus of ~23,500 English-language stories published from 2007 to 2024 on Reddit's /r/nosleep and the Creepypasta Fandom wiki, we compare stylistic profiles across platforms and relate them to differences in rule regimes and moderation practices, established through qualitative extraction and close reading of platform guidelines. Across readability indices, lexical diversity measures, syntactic proxies, and a cross-fit feature-based model, we find that platform membership leaves only a narrow stylistic imprint, largely reducible to a single architectural rule: r/NoSleep's mandatory first-person narration. Beyond this constraint, differences are modest and fail to form coherent platform-specific stylistic signatures. This helps us define what is stylistically common in creepypastas, and understand what the genre is to its writers beyond the topics it deals with or the platform it is written on.

Keywords: creepypasta, platform studies, computational stylistics, online literature, digital humanities

1 Introduction

While literary sociology has long emphasised that genres and canons emerge through the action of mediating institutions (Guillory, 1993), and distant-reading approaches have shown how large-scale stylistic patterns map onto such mediation (Moretti, 2000; Piper, 2018), online vernacular writing is

still often treated as if it circulated in a homogeneous digital space. Yet cultural production is increasingly organised by platforms whose technical architectures, rule regimes, and visibility metrics structure how texts are written, read, and circulated (van Dijck et al., 2018; Poell et al., 2021; Nieborg et al., 2022). In this sense, platforms could operate as new mediators: they reward certain formal choices, penalise others, and thereby shape stylistic equilibria in ways comparable to publishers, editors, or periodicals in earlier media regimes (Poell et al., 2021).

At the same time, a large share of empirical work remains platform-bounded: researchers typically analyse cultural production within a single socio-technical environment—for instance social reading and writing on Wattpad (Pianzola et al., 2020; Kraxenberger and Lauer, 2022; Rebora et al., 2021), evaluation practices on online book-reviewing platforms (Boot, 2023), or the poetics of Twitter/X bot poetry (Waliya and Tijani, 2024; McNulty, 2025). Even studies that synthesise how Reddit is used as a research site largely describe one platform at a time rather than comparing textual outcomes across infrastructures (Proferes et al., 2021). These single-platform designs are indispensable for internal validity, but they make it harder to test whether a genre's stylistic regularities generalise across hosting environments, or whether they are artefacts of platform-specific affordances.

Research on online literature communities nonetheless strongly suggests that textual form diverges across sites. Interface constraints, tagging systems, moderation cultures, and feedback mechanisms have been shown to condition not only circulation and reception but also the textual properties of user-generated writing (Pianzola et al., 2020; Kraxenberger and Lauer, 2022; Silberstein-Bamford, 2024). This implies that even when a genre label is shared, its realisation can be platform-specific: each site constitutes a distinct ecology of

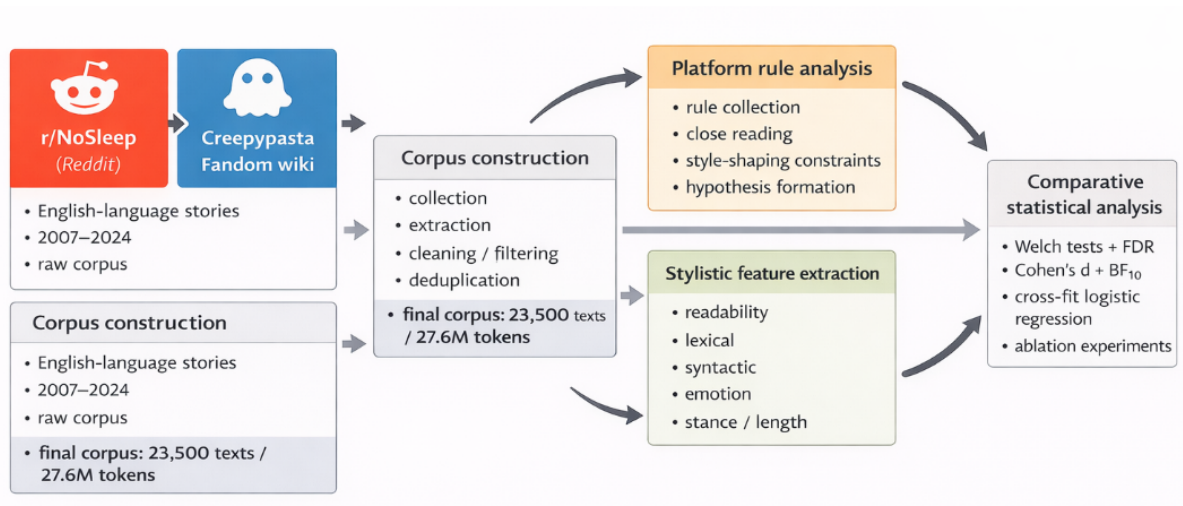


Figure 1: Analytical workflow of the study.

writing and reading, with its own norms of style and its own incentives for particular narrative postures.

Digital-born genres such as creepypasta provide an unusually clear window onto these dynamics. Creepypasta is a vernacular horror genre spanning subreddits and wikis, with relatively low editorial gatekeeping and vast quantities of comparable short-form narratives. We have previously approached creepypasta through the lens of bottom-up canon formation (Lionnet-Rollin and Cafiero, 2025); here, we shift the focus from canonisation outcomes to the cross-platform stylistic regimes through which the genre is produced and stabilised.

In this paper, we conceptualise creepypasta as a platformed genre: a set of shared narrative motifs realised through divergent stylistic norms depending on the hosting infrastructure. Using a corpus of approximately 23,500 English-language stories published between 2007 and 2024 on Reddit’s /r/nosleep and the Creepypasta Fandom wiki, we systematically compare their stylistic profiles. Building on computational stylistics and platform studies (Piper, 2018; van Dijck et al., 2018; Poell et al., 2021), our analysis addresses two questions: do creepypastas exhibit consistent stylistic differences across platforms? can eventual differences be linked to platform-specific rule regimes, interfaces, and visibility logics?

2 Material and methods

2.1 Research design and platform selection

This study adopts a cross-platform comparative design to disentangle genre-level stylistic pat-

terns from platform-specific effects. We focus on two major hosting environments for creepypasta that represent distinct infrastructural and editorial regimes: (1) Reddit’s /r/nosleep, a subreddit organised around rule-bound immersive storytelling and algorithmic (upvote-driven) visibility, and (2) the Creepypasta Fandom wiki, which operates through community editorial gatekeeping and explicit quality guidelines. Both platforms host substantial quantities of comparable horror narratives, making them ideal for comparative analysis.

2.2 Platform rules and governance structures

To ground our quantitative analysis, we conducted qualitative close reading of the official rules and guidelines for each platform, extracting key differences in:

- Narratorial constraints (e.g., first-person requirements, stylistic prohibitions)
- Content restrictions and moderation priorities
- Visibility and recommendation mechanisms
- Quality standards and editorial framing

This qualitative examination informs our hypothesis development (Section 3) and situates our quantitative findings within each platform’s specific affordance structure. The differential emphasis on narratological rules (r/NoSleep) versus community order and content originality (Fandom) provides the conceptual groundwork for our stylistic expectations.

2.3 Dataset construction and preprocessing

We collected texts from two main sources: the Creepypasta Fandom wiki, using *Selenium* and *BeautifulSoup* to navigate and extract story content, and the */r/nosleep* subreddit via the Pushshift database, from which we selected a balanced subset of posts to ensure temporal and qualitative representativeness.

The texts were preprocessed through a multi-step pipeline including HTML stripping and normalisation, language detection and duplicate removal, sentence segmentation and tokenisation, and stop-word masking for lexical analyses. From the cleaned corpus, we extracted 23,500 texts (27.6 million tokens) with associated metadata capturing publication date, author, subreddit score, and comment count.

Table 1: Corpus distribution by platform.

Platform	Total
<i>r/nosleep</i>	13,601
<i>Fandom</i>	11,379

2.4 Feature extraction and computational stylistic analysis

For each text, we extracted features along four formal dimensions and two semantic layers, organised as follows:

2.4.1 Readability indices

Readability was assessed through eight standard indices—Flesch Reading Ease, Flesch–Kincaid Grade Level, Gunning Fog, SMOG (Simple Measure of Gobbledygook), Automated Readability Index (ARI), Coleman–Liau, Dale–Chall, and LIX—chosen for their complementary sensitivity to syllabic, word-length, and sentence-level complexity. These indices estimate the educational level required to comprehend a text and probe whether creepypasta demonstrates consistent accessibility across platforms or whether one site favours more complex prose.

2.4.2 Lexical richness metrics

Lexical richness was captured via four lemma-based metrics: MATTR (moving-average type-token ratio, window = 500 tokens), Honoré’s R, lexical density (ratio of content words to total words), and hapax ratio (proportion of words appearing exactly once). These metrics were *z*-score

normalised per corpus to allow cross-platform comparison. They measure vocabulary diversity and the balance between function words (often associated with narrative stance) and content words (typically denser in descriptive or report-like registers).

2.4.3 Syntactic complexity proxies

Syntactic complexity was derived with spaCy v3.7 (en_core_web_sm), yielding: (1) mean sentence length (word count per sentence), (2) passive voice ratio (proportion of passive constructions), and (3) a nominal style index (ratio of nouns and adjectives to total tokens). Together, these probe whether creepypasta prose tends toward agent-deemphasising or descriptive constructions, which would reflect differences in how each platform’s constraints shape narrative positioning.

2.4.4 Language model perplexity

Language-model perplexity was estimated using GPT-2 (117M parameters) via the HuggingFace transformers library and averaged per document. Lower perplexity values indicate more predictable, formulaic prose patterns, potentially reflecting how tightly platform rules constrain stylistic variation.

2.4.5 Emotion analysis

Emotion scores (fear, disgust, anger, sadness, surprise, joy, neutral) were obtained by applying a DistilRoBERTa (Sanh et al., 2019) model fine-tuned on the GoEmotions dataset (Demszky et al., 2020) to each story, retaining the per-label probability as a continuous feature. This semantic layer captures which affective dimensions dominate each corpus, probing whether first-person experiential narration on *r/NoSleep* amplifies fear-related language, while *Fandom*’s looser constraints allow for greater affective diversity.

2.5 Statistical testing strategies

We employ three complementary analytical strategies, each addressing a different aspect of the platform-genre relationship:

- **Univariate hypothesis tests:** We conduct Welch *t*-tests—chosen for their robustness to unequal variances across groups—with Benjamini–Hochberg false discovery rate (FDR) correction across all stylistic features, reporting Cohen’s *d* (standardised effect size) and Bayes factors (BF₁₀, alternative vs. null) as measures of effect magnitude. This ap-

proach tests whether each feature shows meaningful differences between platforms.

- **Cross-fit logistic regression:** We implement a cross-fit procedure to obtain more robust inferential statistics than standard p -values from a single fitted model. An L1-regularised logistic regression selects features on each training fold, and held-out generalised linear model (GLM) inference is conducted on the corresponding test fold alone. P -values are aggregated conservatively (Bonferroni across folds) and then corrected for multiple testing via Benjamini–Hochberg. This approach identifies which features most robustly predict platform membership when considered jointly.
- **Ablation experiments:** We run a series of feature-ablation experiments on the logistic classifier to disentangle the contribution of first-person narration from other stylistic dimensions. A key experiment residualises all remaining features on the first-person ratio, thus removing the variance of every feature linearly attributable to narrator stance.

3 Results

3.1 Qualitative analysis of platforms’ rules and hypothesis formation

Reddit’s *r/NoSleep* and the Creepypasta Fandom wiki regulate the genre in fundamentally different ways. *r/NoSleep* is primarily an editorial and narratological regime: its rules tightly specify how stories must be written and received (first-person “as if real” immersion, a protagonist able to write the account, strict formatting and length constraints, and strong restrictions on out-of-character interaction), and they also delimit admissible horror content through detailed thematic prohibitions. By contrast, Fandom’s governance is largely about platform use and community order (behavioural norms, anti-abuse provisions, account/security rules), with editorial constraints that are comparatively general and oriented toward originality and compliance (no plagiarism/unauthorised reuse, no AI-generated content, no trollpastas), rather than a single mandated storytelling device. Finally, the two sites operationalise “quality” differently: Fandom explicitly frames deletion around low-quality or overly generic writing, whereas *r/NoSleep*’s removals are mainly driven by conformity to its immersive storytelling apparatus and content boundaries, with

quality enforced indirectly through language and format requirements.

This governance structure leads us to formulate and test five hypotheses, grounded in each platform’s specific affordances:

- **H1** (Variance compression on *r/NoSleep*). Because *r/NoSleep* imposes strict constraints on narration (first-person, “as if real”), format, and language correctness, texts published there should display lower dispersion than Fandom on core stylistic metrics (readability indices, sentence-length proxies, lexical-density measures), even if mean differences are small.
- **H2** (Readability standardisation on *r/NoSleep*). Due to minimum-length requirements and explicit expectations of “correct English,” *r/NoSleep* stories should show fewer extreme readability outliers and more central clustering on standard readability indices than Fandom (i.e., shorter tails and fewer highly noisy/fragmentary texts).
- **H3** (Lower lexical density on *r/NoSleep*). The enforced first-person experiential stance on *r/NoSleep* should increase the relative frequency of function words and stance markers (e.g., first-person pronouns, perception verbs), yielding lower lexical density (content-word share) than on Fandom, where narratorial stance is less constrained.
- **H4** (Lower nominal/passive style on *r/NoSleep*). *r/NoSleep*’s immersion rules and bans on “found media” formats should reduce the prevalence of report-like registers; consequently, *r/NoSleep* texts should exhibit lower nominal style indices and lower passive-voice ratios than Fandom on average.
- **H5** (Greater register heterogeneity on Fandom). Because Fandom does not enforce a single storytelling mode, its creepypasta corpus should exhibit greater between-text heterogeneity in syntactic proxies (sentence length, passive ratio, nominal style) and lexical measures (lexical density, diversity), reflected in higher variance and potentially multimodal feature distributions.

Table 2: Cross-platform comparison of stylistic features by domain. Values are reported as mean \pm SD. Cohen’s d gives the standardised effect magnitude; BF_{10} reports Bayesian evidence for the alternative over the null. The “Eff.” column provides a compact interpretation of effect size: 0 = negligible ($|d| < 0.20$), + = small-to-medium ($0.20 \leq |d| < 0.50$), ++ = moderate ($0.50 \leq |d| < 0.80$), +++ = large ($|d| \geq 0.80$). † indicates $BF_{10} < 1$, i.e. evidence favouring the null. Each row is labelled with its corresponding hypothesis (H1–H5) where applicable.

Feature	Fandom	Reddit	Cohen’s d	Eff.	BF_{10}
<i>Readability (H1, H2)</i>					
Flesch Reading Ease	81.87 \pm 12.73	83.55 \pm 13.84	0.126	0	7.45×10^{19}
Flesch–Kincaid Grade Level	5.61 \pm 4.00	5.43 \pm 4.99	0.040	0	2.81
Gunning Fog Index	7.55 \pm 4.05	7.39 \pm 5.09	0.035	0†	0.67
SMOG Index	8.09 \pm 1.65	7.85 \pm 1.15	0.169	0	2.27×10^{34}
Automated Readability Index	6.98 \pm 5.13	6.60 \pm 6.33	0.066	0	1.95×10^4
Coleman–Liau Index	6.58 \pm 1.90	5.98 \pm 1.35	0.374	+	9.34×10^{172}
Dale–Chall Score	6.98 \pm 1.13	6.79 \pm 0.80	0.206	+	6.42×10^{51}
LIX	31.72 \pm 8.87	30.31 \pm 13.58	0.121	0	1.57×10^{19}
<i>Lexical richness (H3, H5)</i>					
Lexical Density	0.48 \pm 0.04	0.46 \pm 0.03	0.398	+	2.02×10^{191}
Hapax Legomena	371.4 \pm 297.89	325.4 \pm 162.87	0.196	0	1.11×10^{45}
Honoré’s R	2084.31 \pm 388.14	1974.47 \pm 251.82	0.342	+	3.07×10^{142}
Type–Token Ratio	0.68 \pm 0.11	0.70 \pm 0.03	0.191	0	7.78×10^{39}
<i>Syntax (H4, H5)</i>					
Sentence length	14.41 \pm 9.69	14.41 \pm 9.67	0.001	0†	0.014
Perplexity (GPT-2)	2.89 \pm 0.39	2.80 \pm 0.30	—	—	—
Noun–verb ratio	1.15 \pm 0.36	1.06 \pm 0.24	0.291	+	2.40×10^{103}
Noun–adj.–verb ratio	1.56 \pm 0.47	1.43 \pm 0.31	0.321	+	1.12×10^{126}
Passive-voice ratio	0.04 \pm 0.03	0.03 \pm 0.02	0.277	+	9.72×10^{91}
<i>Emotion</i>					
Fear	0.36 \pm 0.25	0.46 \pm 0.24	0.409	+	1.19×10^{217}
Disgust	0.23 \pm 0.18	0.17 \pm 0.15	0.361	+	8.58×10^{164}
Anger	0.09 \pm 0.10	0.08 \pm 0.08	0.142	0	1.09×10^{24}
Sadness	0.07 \pm 0.13	0.06 \pm 0.08	0.165	0	9.59×10^{31}
Surprise	0.08 \pm 0.11	0.09 \pm 0.10	0.017	0†	0.036
Joy	0.02 \pm 0.05	0.01 \pm 0.04	0.043	0	3.19
Neutral	0.14 \pm 0.15	0.13 \pm 0.12	0.057	0	234
<i>Descriptive / stance</i>					
Text length (chars)	11 622 \pm 28 899	8 592 \pm 6 504	0.151	0	1.42×10^{24}
First-person pronoun ratio	0.398 \pm 0.256	0.549 \pm 0.145	0.746	++	∞

3.2 Univariate comparisons between platforms

All pairwise comparisons use Welch’s unequal-variance t -test. We report *Cohen’s d* as a standardised effect-size measure and the Bayes factor BF_{10} (alternative vs. null) to evaluate evidential strength independently of sample size. A $BF_{10} < 1$ signals that the data favour the null hypothesis of no difference; values between 1 and 3 are conventionally regarded as anecdotal evidence, and values above 10 as strong evidence for a difference. Cohen’s d thresholds follow conventional benchmarks: $|d| < 0.20$ negligible, 0.20–0.50 small-to-medium, 0.50–0.80 moderate, > 0.80 large (Cohen, 1988). Descriptive statistics (mean \pm SD) are retained in

Table 2 so that the reader can gauge absolute magnitudes alongside standardised effects.

3.2.1 H1: Variance compression on r/NoSleep

H1 receives partial support. The clearest difference concerns text length: Fandom displays much greater dispersion (SD = 28, 899 chars), whereas r/NoSleep appears more constrained at the document level (SD = 6, 504 chars). By contrast, most readability and sentence-level measures remain very close across platforms, and their variances do not strongly differ. In other words, r/NoSleep seems to compress coarse formal properties (document length) more than fine-grained stylistic ones (sentence complexity).

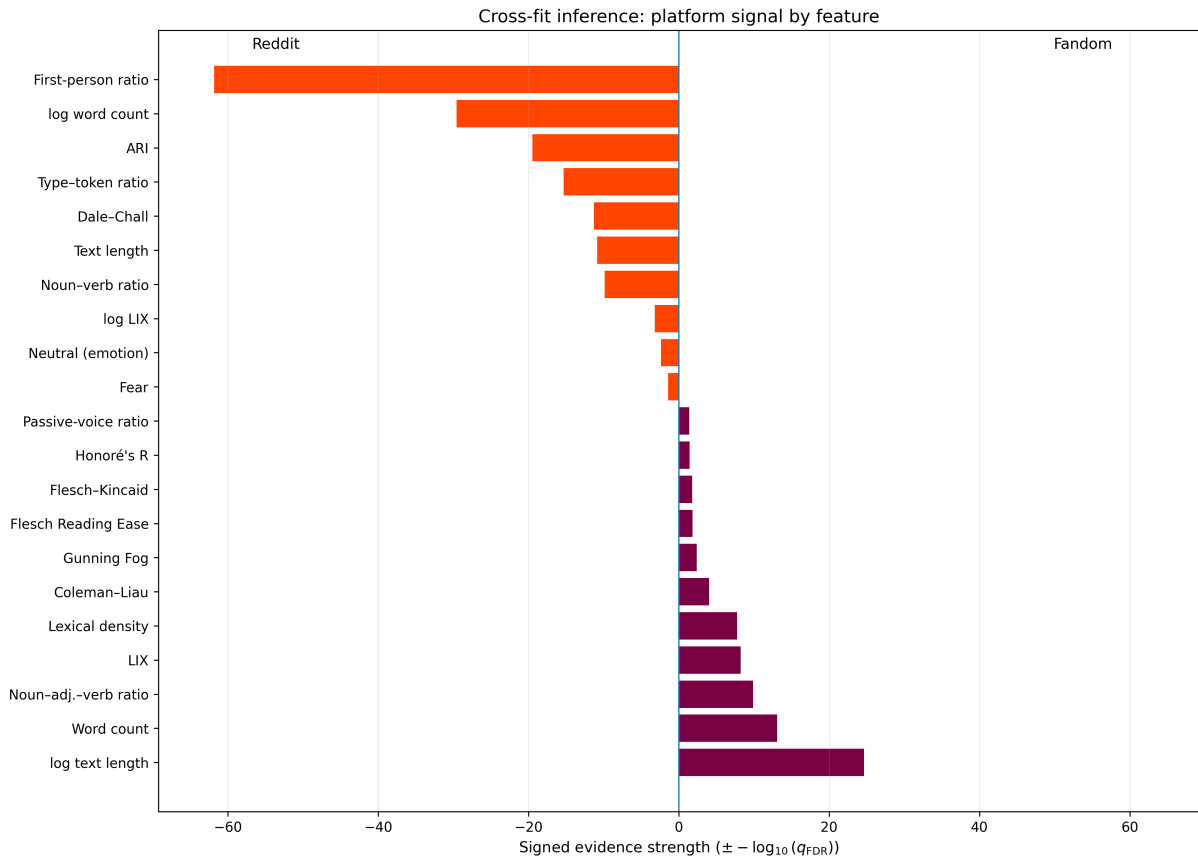


Figure 2: Cross-fit inference by feature. Bar length represents signed $-\log_{10}(q_{FDR})$: negative values indicate features associated with higher odds of Reddit, positive values indicate features associated with higher odds of Fandom. Greater distance from zero indicates stronger evidence after FDR correction. Exact values for odds ratios, aggregated p -values, and corrected q -values are reported in Table 3.

3.2.2 H2: Readability standardisation on r/NoSleep

H2 is only weakly supported. A few readability indices point to slightly greater standardisation on r/NoSleep (Coleman-Liau: $d = 0.374$, Dale-Chall: $d = 0.206$), but the pattern is not consistent across the full set of measures. Most indices show negligible differences. Overall, the two platforms remain close on readability, and no strong platform-specific profile emerges from these indices alone.

3.2.3 H3: Lower lexical density on r/NoSleep

H3 is clearly supported. r/NoSleep texts show substantially lower lexical density ($M = 0.46$ vs. $M = 0.48$, $d = 0.398$) and significantly higher first-person pronoun ratio ($M = 0.549$ vs. $M = 0.398$, $d = 0.746$) than Fandom texts. This pattern is fully consistent with the platform's mandatory first-person immersive regime. Critically, the large effect on first-person ratio ($d = 0.746$, classified as "moderate") suggests that this single narratorial constraint drives much of the observable platform

signal.

3.2.4 H4: Lower nominal/passive style on r/NoSleep

H4 is supported in direction, but the magnitude of the effect remains modest. Relative to Fandom, r/NoSleep shows lower noun-adjective-verb ratios ($d = 0.321$) and lower passive-voice ratios ($d = 0.277$), which fits its immersive first-person storytelling norms. Still, these differences are small-to-medium and do not amount to a sharply distinct platform style.

3.2.5 H5: Greater register heterogeneity on Fandom

H5 is only partially supported. Fandom is clearly more heterogeneous in document-level properties, especially text length (variance $\approx 5.2\times$ higher), but this greater heterogeneity is much less visible in sentence-level measures (sentence length: $d = 0.001$). The result therefore points to broader formal looseness on Fandom rather than to a strongly differentiated internal stylistic landscape.

Table 3: Cross-fit inference results for the 21 features that remain significant after FDR correction ($q < 0.05$). All reported features were selected in all five folds. $OR > 1$ indicates higher odds of Fandom; $OR < 1$ indicates higher odds of Reddit. The first-person ratio dominates the signal, confirming the primacy of narrator stance as the main platform discriminator.

Feature	OR	P_{agg}	q_{FDR}
First-person ratio	0.519	4.6×10^{-64}	1.5×10^{-62}
log word count	2.7×10^{-7}	1.7×10^{-31}	2.6×10^{-30}
log text length	1.1×10^6	2.2×10^{-26}	2.3×10^{-25}
ARI	0.071	4.3×10^{-21}	3.4×10^{-20}
Type-token ratio	0.668	7.3×10^{-17}	4.7×10^{-16}
Word count	126.5	1.5×10^{-14}	8.0×10^{-14}
Dale-Chall	0.454	1.1×10^{-12}	4.9×10^{-12}
Text length	0.015	3.2×10^{-12}	1.3×10^{-11}
Noun-adj.-verb ratio	1.990	3.9×10^{-11}	1.3×10^{-10}
Noun-verb ratio	0.469	4.0×10^{-11}	1.3×10^{-10}
LIX	8.675	2.1×10^{-9}	6.1×10^{-9}
Lexical density	1.342	6.8×10^{-9}	1.8×10^{-8}
Coleman-Liau	1.779	4.0×10^{-5}	9.8×10^{-5}
log LIX	0.330	2.7×10^{-4}	6.3×10^{-4}
Neutral (emotion)	0.643	2.0×10^{-3}	4.2×10^{-3}
Gunning Fog	3.152	2.2×10^{-3}	4.3×10^{-3}
Flesch Reading Ease	1.820	7.8×10^{-3}	1.5×10^{-2}
Flesch-Kincaid	3.784	9.8×10^{-3}	1.7×10^{-2}
Honoré’s R	1.818	2.3×10^{-2}	3.8×10^{-2}
Fear	0.527	2.4×10^{-2}	3.8×10^{-2}
Passive-voice ratio	1.100	2.9×10^{-2}	4.3×10^{-2}

Summary of univariate results Globally, most features show either negligible effects or no meaningful divergence, and only the first-person pronoun ratio stands out as a strong and unambiguous discriminator. The clearest platform effect is therefore not a broad stylistic signature, but a narrow architectural constraint: *r/NoSleep*’s mandatory first-person narration. Beyond that rule, the two corpora remain surprisingly similar.

3.3 Logistic regression with cross-fit inference

To obtain more defensible inferential statistics, we implement a cross-fit procedure: an L1-regularised logistic regression selects features on each of five training folds, and held-out GLM inference is then conducted on the corresponding test fold alone. P -values are aggregated conservatively (Bonferroni across folds) and then corrected for multiple testing via the Benjamini-Hochberg procedure.

The out-of-fold AUC of the full logistic model is 0.795, indicating moderate but not overwhelming discriminability from the engineered feature set. This is consistent with the broader pattern observed throughout the paper: platform differences are real, but they are concentrated in a limited number of dimensions rather than distributed across the full stylistic feature space.

Figure 2 and Table 3 show that the cross-fit sig-

nal is dominated by a small number of features. The strongest and most stable predictor is the first-person ratio, which is selected in every fold and aligns directly with *r/NoSleep*’s narratorial rule. Length-related variables also contribute meaningfully, reflecting the broader dispersion of Fandom texts. Other readability, lexical, and emotion features remain significant after correction, but their contribution is weaker and more diffuse. The main lesson is therefore not that many independent stylistic dimensions sharply separate the two platforms, but that a narrow cluster of cues—above all narrator stance—accounts for most of the learnable difference.

Several features that appear significant in the univariate comparisons do not survive the cross-fit procedure. This suggests that part of their apparent signal is either redundant with stronger predictors (like first-person ratio) or too weak to remain robust once features are considered jointly. In that respect, the cross-fit analysis reinforces the paper’s main conclusion: platform differences exist, but they are concentrated in a limited number of dimensions.

3.4 Ablation study: disentangling first-person narration

To directly test whether the first-person pronoun ratio drives most of the platform signal, we run a series of ablation experiments on the logistic classifier. Each experiment removes or isolates a block of features and evaluates out-of-fold AUC. A key additional experiment *residualises* all remaining features on the first-person ratio within each training fold before classification on the held-out fold, thus removing the variance of every feature that is linearly attributable to narrator stance.

Results are summarised in Figure 3.

The ablation results clarify the hierarchy of signals. Removing the first-person ratio alone reduces performance only modestly ($AUC \approx 0.78$), because part of its effect is absorbed by correlated features such as lexical density and nominal style. The more revealing test is residualisation: once the variance attributable to first-person narration is removed from the remaining features, classification performance drops much more sharply ($AUC \approx 0.70$). This shows that narrator stance is not just one predictor among others, but the main organising signal in the model. Other feature blocks still contribute, especially readability and length, but their role is clearly secondary.

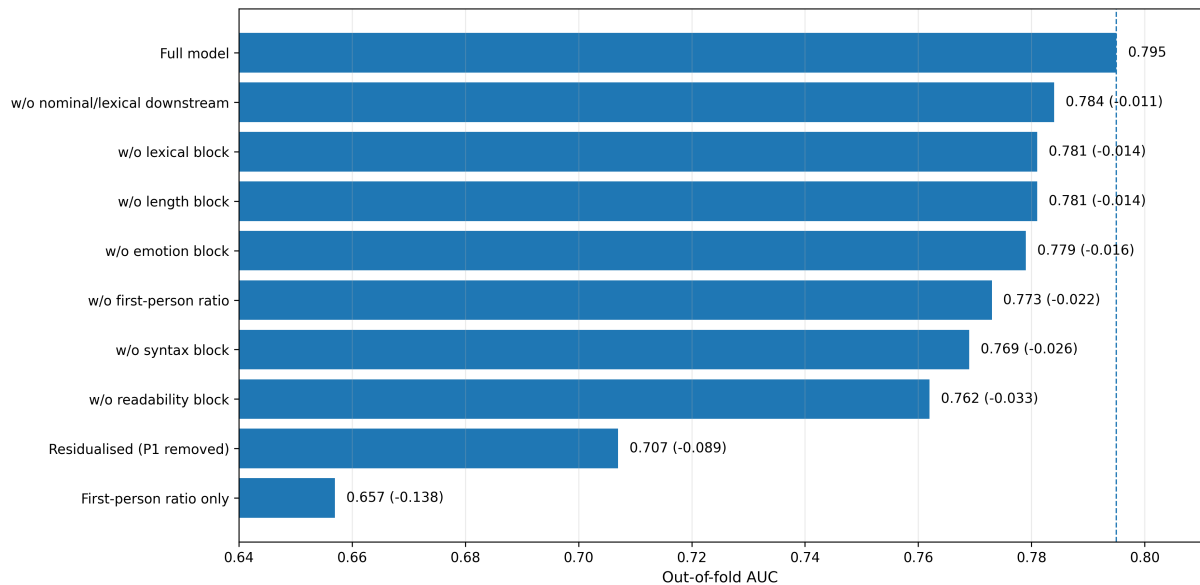


Figure 3: Ablation study across all experimental conditions. Bars report out-of-fold AUC, and labels indicate the change relative to the full model. The figure separates first-person-related conditions from broader feature-block ablations. The largest performance drop is observed when all remaining features are residualised on the first-person ratio, confirming that narrator stance accounts for a substantial share of the platform signal.

4 Discussion

The point of this paper was not simply to measure platform effects, but to ask how far a genre can resist them. In the case of creepypasta, the answer seems mixed. Some rules do leave a strong mark: the clearest example is *r/NoSleep*’s first-person requirement, which produces the most visible and robust difference in the corpus. But beyond such explicit and easily enforceable constraints, platform rules seem to shape style much less strongly. Most readability, lexical, syntactic, and emotional measures remain relatively close across the two corpora, and they do not add up to two clearly distinct prose styles.

This suggests that genre can prevail against platform rules, though not completely. The platforms do not produce identical texts, but neither do they fully reorganise the genre at the level of prose. A plausible interpretation is that many writers are responding not only to platform rules, but also to shared genre expectations: they read across sites, imitate successful stories, and carry writing habits from one environment to another.

Looking at what remains similar across platforms is therefore just as important as identifying what differs. Empirically, the two corpora share a common textual profile: they occupy broadly similar readability ranges (both around grade 5–8), show near-identical average sentence length (14.41

words), remain fairly close on most lexical and syntactic measures, and draw on a comparable affective repertoire dominated by fear and disgust rather than sharply divergent emotional modes. Taken together, these regularities suggest that creepypasta is held together by more than topic alone. Across platforms, the genre appears to preserve a recognisable stylistic core: accessible narrative prose, strong experiential framing, and a relatively stable horror texture.

We are not claiming that platform effects are weak in general. Rather, we believe that more systematically examining texts from the same genre across platforms would be beneficial for computational literary studies, to detect commonalities that help clarify what gives a genre its continuity and coherence.

5 Limitations

This interpretation remains provisional. We compare only two platforms, and both belong to the same broad online horror ecology. Future work could extend this design to other genres and platforms to test whether the “narrow platform signal” we observe here generalises more widely. Our feature set also remains mostly surface-level; stronger contrasts might emerge from measures tied more closely to narrative structure, discourse organisation, or interaction with readers. Finally, we anal-

use published texts rather than rejected or removed ones, so some platform effects may be more visible in moderation outcomes than in the surviving corpus itself.

6 Data and code availability

Code is available at this repository: https://anonymous.4open.science/r/acl_nlp4DH_creepy-B3CF/. Data are available upon reasonable request but cannot be reshared for legal reasons.

7 Acknowledgements

This work received support from the CultureLab: Computational Science of Culture Grand Research Programme of Université PSL, funded under the Investissements d’Avenir programme launched by the French Government and implemented by ANR, under reference ANR-10-IDEX-0001-02 PSL. Errors remain our own.

References

- Peter Boot. 2023. ‘A pretty sublime mix of WTF and OMG’. four explorations into the practice of evaluation on online book reviewing platforms. *Journal of Cultural Analytics*, 7(2).
- Jacob Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2 edition. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. *GoEmotions: A dataset of fine-grained emotions*. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4040–4054, Online. Association for Computational Linguistics.
- John Guillory. 1993. *Cultural Capital: The Problem of Literary Canon Formation*. University of Chicago Press, Chicago.
- Maria Kraxenberger and Gerhard Lauer. 2022. *Wreading on online literature platforms*. *Written Communication*, 39(3):462–496.
- Alexandre Lionnet-Rollin and Florian Cafiero. 2025. *What makes a creepypasta stick? on the recipe for canonicity of online literary productions*. HAL preprint. HAL Id: hal-05406444, version 1.
- Tess McNulty. 2025. The some and the many: A distant reading of twitter’s most popular #poems. In *Expressive Networks: Poetry and Platform Cultures*, pages 43–62. Amherst College Press.
- Franco Moretti. 2000. *The slaughterhouse of literature*. *Modern Language Quarterly*, 61(1):207–227.
- David B. Nieborg, Thomas Poell, and José van Dijck. 2022. *Platforms and platformization*. In Terry Flew, Jennifer Holt, and Julian Thomas, editors, *The SAGE Handbook of the Digital Media Economy*, pages 29–49. SAGE, London.
- Federico Pianzola, Simone Rebor, and Gerhard Lauer. 2020. *Wattpad as a resource for literary studies: Quantitative and qualitative examples of the importance of digital social reading and readers’ comments in the margins*. *PLOS ONE*, 15(1):e0226708.
- Andrew Piper. 2018. *Enumerations: Data and Literary Study*. University of Chicago Press, Chicago.
- Thomas Poell, David B. Nieborg, and Brooke Erin Duffy. 2021. *Platforms and Cultural Production*. Polity, Cambridge, UK.
- Nicholas Proferes, Naiyan Jones, Sarah Gilbert, Casey Fiesler, and Michael Zimmer. 2021. *Studying reddit: A systematic overview of disciplines, approaches, methods, and ethics*. *Social Media + Society*, 7(2):1–16.
- Simone Rebor, Peter Boot, Federico Pianzola, Brigitte Gasser, J. Berenike Herrmann, Maria Kraxenberger, Moniek M. Kuijpers, Gerhard Lauer, Piroška Lendvai, Thomas C. Messerli, and Paul Sorrentino. 2021. *Digital humanities and digital social reading*. *Digital Scholarship in the Humanities*, 36(Supplement_2):ii230–ii250.
- Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. *Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter*. *ArXiv*, abs/1910.01108.
- Fabienne Silberstein-Bamford. 2024. *“thank god for tags”: Fanfiction as a reading paradigm*. *New Review of Hypermedia and Multimedia*, 30(1–2):129–147.
- José van Dijck, Thomas Poell, and Martijn de Waal. 2018. *The Platform Society: Public Values in a Connective World*. Oxford University Press, New York.
- Yohanna Joseph Waliya and Mufutau Adebawale Tijani. 2024. *Technodiscursive analysis of twitterbot poetry*. *Digital Studies in Language and Literature*, 2(1):1–24.