

Cross-cultural Inspiration Detection and Analysis in Real and LLM-generated Social Media Data

Anonymous ACL submission

Abstract

Inspiration is linked to various positive outcomes, such as increased creativity, productivity, and happiness. Although inspiration has great potential, there has been limited effort toward identifying content that is *inspiring*, as opposed to just engaging or positive. Additionally, most research has concentrated on Western data, with little attention paid to other cultures. This work is the first to study cross-cultural inspiration through machine learning methods. We aim to identify and analyze real and AI-generated cross-cultural inspiring posts. To this end, we make publicly available the IN-SPAIRE dataset, which consists of 2,000 real inspiring posts, 2,000 real non-inspiring posts, and 2,000 generated inspiring posts evenly distributed across India and the UK. The real posts are sourced from Reddit, while the generated posts are created using the GPT-4 model. Using this dataset, we conduct extensive computational linguistic analyses to (1) compare inspiring content across cultures, (2) compare AI-generated inspiring posts to real inspiring posts, and (3) determine if detection models can accurately distinguish between inspiring content across cultures and data sources.

1 Introduction

Inspiration has been a part of our world for millennia, starting with ancient Greece, where Muses were responsible for delivering divine knowledge by whispering in a poet’s ear (Leavitt and Leavitt, 1997), all the way to today’s creativity domain, where it is still common for artists and scientists to attribute their best ideas to a higher power, independent of their own control.

According to Thrash and Elliot (2003, 2004), inspiration is a general construct that consists of three core characteristics: *evocation*, *transcendence*, and *approach motivation*. Evocation refers to the process of being triggered by a stimulus, either from within (such as a creative idea that comes from the

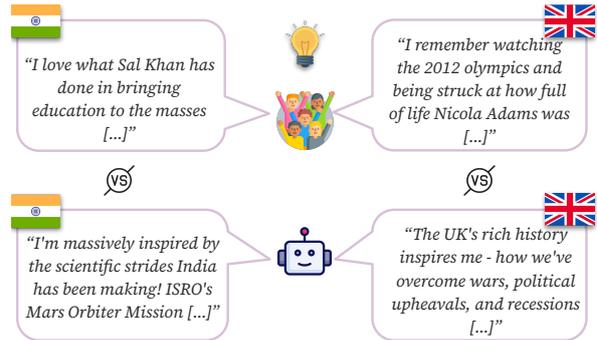


Figure 1: We compare AI-generated and human-written inspiring Reddit content across India and the UK. Although it is challenging for a person to distinguish between them, we find significant linguistic cross-cultural differences between generated and real inspiring posts.

subconscious) or from outside (such as a person, object, music, or nature). Transcendence allows one to perceive something beyond their usual concerns (Milyavskaya et al., 2012). Finally, an inspired person is motivated to express, transmit, or act on their inspiration (Elliot and Thrash, 2002).

Inspiration is an area of study with promising cross-disciplinary applications in creative fields (e.g., advertisement, storytelling), education, therapy, mentorship, coaching, or social media. For instance, social network recommendation systems can mitigate potential harms by showing more positive and inspiring content to users (Ignat et al., 2021). Access to inspiring content can have a positive impact on people’s lives by offering them a fresh perspective and motivating them to take action, particularly during periods of uncertainty and concern (Oleynick et al., 2014). Moreover, inspiration facilitates progress towards goals (Milyavskaya et al., 2012) and increases overall well-being (Thrash et al., 2010).

Despite the compelling motivation, little research has been done on the automatic identification of content that is *inspiring*, as opposed to simply positive (Ignat et al., 2021), and no studies have been conducted on how inspiration varies

068 across cultures, and whether it can be automatically
069 generated. At the same time, LLMs’ impressive
070 generative ability can create new opportunities for
071 automatically generating inspiring content.

072 In this work, we aim to address these research
073 gaps by focusing on three main research questions.
074 (RQ1:) How do inspiring posts compare across cul-
075 tures? (RQ2:) How do AI-generated inspiring posts
076 compare to real inspiring posts across cultures? and
077 (RQ3:) Can detection models differentiate inspir-
078 ing posts across cultures and data sources?

079 We summarize our contributions as follows.
080 First, we share a **novel dataset comprised of cross-**
081 **cultural real and generated inspiring and non-**
082 **inspiring posts**, for a total of 2,000 real inspiring
083 posts, 2,000 non-inspiring posts, and 2,000 LLM-
084 generated posts, balanced across India and the UK.
085 Second, we make use of the dataset to conduct
086 **computational linguistic analyses to compare in-**
087 **spiring content across cultures and data sources**,
088 i.e., India vs. the UK and AI-generated vs. human-
089 written. Finally, **we explore the effectiveness of**
090 **machine learning models for detecting inspira-**
091 **tion** across diverse data sources and cultures.

092 2 Related Work

093 **Automatic Inspiration Detection.** There have
094 been a limited number of research studies con-
095 ducted on inspiration, with the majority of them
096 being carried out by the psychology and sociol-
097 ogy communities. These studies, such as the ones
098 by Thrash and Elliot (2003, 2004) and Elliot and
099 Thrash (2002), have established the fundamental
100 characteristics of inspiration. Additionally, they
101 have developed a scale to measure the frequency
102 with which people experience inspiration in their
103 lives. These studies found that individuals who are
104 inspired tend to be more open to new experiences
105 and show greater absorption in their tasks. They
106 are also more intrinsically motivated, have a strong
107 drive to master their work, and are less competitive.

108 The work most similar to ours is from Ignat
109 et al. (2021), who are the first to study inspira-
110 tion through machine learning methods. To facili-
111 tate research in this domain, they release a weakly
112 labeled dataset consisting of inspiring and non-
113 inspiring posts collected from Reddit and anno-
114 tate these posts with their effect on the reader and
115 the emotions they transmit. They also provide a
116 RoBERTa (Liu et al., 2019) classifier fine-tuned on
117 human labels to provide a strong baseline for de-

termining whether a post is inspiring. Finally, they
perform data analyses to gain insights into which
topics inspire readers and how they influence them.
Our work builds on Ignat et al. (2021) by extend-
ing it to other cultures, India and the UK, and by
collecting AI-generated inspiring posts in order to
compare them to human-written posts.

Human vs. LLM-generated Cross-cultural Text.
With the rapid development of LLMs, these mod-
els demonstrate remarkable proficiency in generat-
ing human-like text across multiple languages and
styles (Wu et al., 2023; Tang et al., 2023).

In particular, LLMs excel at creative writing,
such as story generation (Yuan et al., 2022), ad-
vertising slogan creation (Murakami et al., 2023),
and news composition (Yanagi et al., 2020). Tools
like Yuan et al. (2022) can help users in their cre-
ative pursuits. By generating inspiring content, our
work is more indirect but still aims to help the user
express and act on their insights.

More similar to our work, LLMs have also
started to be applied to motivate people (Cox et al.,
2023). For example, Karinshak et al. (2023) used
GPT-3 to generate messages to persuade people
to receive the Covid-19 vaccine. To the best of
our knowledge, we are the first to generate inspir-
ing content using LLMs. Our work is also part of
the emerging work on modeling cultural factors in
LLMs (Huang and Yang, 2023; Fung et al., 2022;
Ramezani and Xu, 2023). Inspiration varies across
cultures. Therefore, we test the cultural knowledge
of LLMs about inspiration in India and the UK, and
compare it to inspiring Reddit posts from users in
these countries.

Computational Linguistics for Social Media
Analysis. The advent of computational linguis-
tics techniques has enabled researchers to analyze
vast amounts of social media data for various pur-
poses, including sentiment analysis, topic model-
ing, and linguistic variation across cultures (Pen-
nebaker et al., 2007; Pang et al., 2008; Imran et al.,
2020). These works have facilitated the extraction
of meaningful insights from diverse linguistic con-
texts, paving the way for studies on cross-cultural
communication in online environments.

Social media is a key source of inspiring content
for younger audiences (Raney et al., 2018). Fea-
tures such as hope and appreciation of beauty and
excellence trigger self-transcendent emotions in
videos tagged with “inspiration” on YouTube (Dale
et al., 2017), as well as in #inspiring and #mean-

ingful Tumblr memes and Facebook posts (Rieger and Klimmt, 2019; Dale et al., 2020). Similarly, we collect Reddit posts from subreddits related to inspiration across UK and Indian cultures and analyze them using computational linguistic tools such as Pennebaker et al. (2007).

3 The INSPAIRED Dataset

To answer our research questions, we compile a novel dataset, which we refer to as INSPAIRED - AI-generated Inspiring Reddit Content. Our dataset contains inspiring and non-inspiring posts from India and the UK, from two different sources: (1) crawled from Reddit and (2) generated by an LLM.

3.1 🌸 Real Inspiring Content

We collect 2,000 weakly labeled inspiring posts and 2,000 weakly labeled non-inspiring posts, balanced across India and the UK. We describe our data collection and annotation process below.

Data Collection. We scrape around 5,300 posts from Reddit, a popular online platform, specifically focusing on culturally inspiring content. Following the data collection process from Ignat et al. (2021), we conduct searches within culture-related or discussion-related flairs of the subreddits using keywords, such as “inspiration” and “motivation”, to identify the relevant data in the form of both posts and comments.

More specifically, for Indian data, we primarily target the regions of *Kerala*, *Karnataka*, *Maharashtra*, and *Tamil Nadu*. Besides the general “r/india”, we also explore subreddits at the *state* level, such as “r/Kerala”, “r/Karnataka”, “r/Maharashtra”, and “r/TamilNadu”, which serve as hubs for focused discussions on regional culture, traditions, and social issues. Finally, we also examine subreddits from specific *cities*, including capital cities like *Chennai* and *Bangalore*, to capture more local perspectives and experiences. For the UK data, we follow a similar strategy to collect Reddit posts, targeting both *state* level subreddits, such as “r/UnitedKingdom”, as well as *regional* subreddits representing areas and capital cities, such as “r/London”.

We group the collected posts into two main categories: inspiring posts from India and inspiring posts from the UK. Initially, we attempted to create a more fine-grained split at the region or city level, but we faced difficulties in finding annotators from those specific regions. However, we encourage future research to explore this direction and

Please read: [Inspiration Definition](#)

Something that fills people with an animating or exalting influence, or produces rousing feelings, novel thoughts, or a motivation to act.
 ! Note that we are not asking if the poster is inspired !

Example of inspiring post: “Never regret your decisions, every mistake makes u smarter and stronger”
Example of non-inspiring post: “Three months going on one year. He is such a light in the midst of all the chaos. We are so thankful for our little boo.”

Sir David Attenborough. He's brought the wonders of the natural world to millions. Inspiring people to pay more attention to life and it's beauty. The dude is in his 90s and still kicking it.

Is this text inspiring to you?

yes
 no

Move backward Move forward

Figure 2: Annotation guidelines for labeling inspiration. investigate how inspiration varies within a country or within a region, along with exploring other demographic information such as language, age, gender, or income.

Data Filtering. The collected posts are further filtered with an inspiration classifier model. We use the XLM-RoBERTa base (Conneau et al., 2019) from HuggingFace.¹ The model is a multilingual version of RoBERTa (Liu et al., 2019), and is pre-trained on 2.5TB of filtered Common-Crawl (Wenzek et al., 2020)². We fine-tune the model for five epochs using the dataset described below, with a learning rate of $2e-5$ and a batch size of 8. More implementation details can be found in the Appendix A.2.

For fine-tuning, we use the data from Ignat et al. (2021). The data contains around 12,000 annotated “general”, i.e., no target culture, balanced Reddit inspiring and non-inspiring posts. More details about the fine-tuning data can be found in Ignat et al. (2021). Next, we evaluate the quality of the fine-tuned model predictions by annotating a subset of the posts and comparing them with the model predictions.

Data Annotation. The posts are annotated by crowd-sourced workers from India and the UK. For each country, we select a sample of 200/ 1,000 posts equally sampled from inspiring and non-inspiring predicted posts, to be labeled by three annotators from that country.

We use Potato (Pei et al., 2022) to create the user interface and store the data. Following the guidelines from Ignat et al. (2021), the user interface

¹<https://huggingface.co/FacebookAI/xlm-roberta-base>

²We choose a multilingual model, as the Indian data contains code-mix data.

contains a definition of inspiration and examples of inspiring and non-inspiring posts, as seen in Figure 2. To find and hire UK annotators, we connect Potato to Prolific.³ We select three annotators who have the following qualifications: an approval rate >98%, living in the UK, and a right to vote (as a proxy for age). The annotators independently label 200 posts each, categorizing the posts based on their subjective judgment of whether they found the content inspiring or not. We conduct a similar annotation process with three reliable Indian annotators to label the data from India. To ensure consistency amongst the annotators, they are provided with the same interface as used for the UK annotators and shown in Figure 2.

Finally, we compute the agreement score between the annotators using the Fleiss Kappa measure (Fleiss and Cohen, 1973). We obtain a score of 0.24 for UK data and 0.29 for Indian data, indicating a fair agreement, aligning with Ignat et al. (2021) who obtained a 0.26 agreement score. Furthermore, by manually examining the low-agreement examples, we find that inspiration is highly subjective and *personal*, relying on one’s beliefs, interests, or upbringing (e.g., someone not interested in nature will probably not be inspired by David Attenborough). Finally, we make available all the annotations to encourage future work to research the general or personal nature of inspiration based on data and downstream applications.

Given that each post is annotated by three annotators, following Ignat et al. (2021), we mark a post as inspiring if at least one annotator labeled it as inspiring. We do not select the majority label to avoid excluding other annotators’ opinions. Disagreement among annotators in subjective tasks may reflect systematic differences in opinions across groups, not noise (Fleisig et al., 2023).

We find that the predictions made by the fine-tuned model are quite similar to human annotations. For Indian data, the accuracy rate is 72.9%, and the F1 score is 75.5%. Similarly, for UK data, the accuracy rate is 73.5%, and the F1 score is 80.1%. Therefore, we decide to not annotate more data and instead use the model predictions. Additionally, we further fine-tune the model on the annotated data and use it to collect more weakly labeled data. This approach aims to leverage the insights gained from the initial fine-tuning process and apply them to the user-annotated subset, thereby refining the model

³www.prolific.com

to better capture the nuances present in this specific dataset. We use the fine-tuned model predictions and the annotated subset as our final labeled data.

Quality Assurance. We remove posts that are classified as toxic or hate speech using two fine-tuned XLM-RoBERTa classifiers from HuggingFace.⁴⁵ We also remove profanity from posts using a profanity classifier.⁶ We acknowledge that automatic detection of profanity, toxicity, and hate speech are active research areas that have yet to be solved (Dale et al., 2021; Vidgen et al., 2021). Finally, for each country, we manually inspect 100 random posts to ensure syntactic and semantic correctness and find the remaining toxic or hate speech posts. We find that around 99% of the posts are of good quality.

Data Statistics. The final data statistics are shown in Table 1.

	# NON-INSPIRING ✗		# INSPIRING ♡	
	Annotated	Weakly-Labeled	Annotated	Weakly-Labeled
	100	900	100	900
	100	900	100	900

Table 1: Final number of inspiring (♡) and non-inspiring (✗) posts across India and the UK.

3.2 LLM-Generated Inspiring Content

We generate 2,000 inspiring posts with GPT-4,⁷ balanced across India and the UK. Our study can be conducted with any LLM. However, we chose GPT-4 because it is one of the largest LLMs available and has been shown to effectively emulate human texts. (Achiam et al., 2023)

3.2.1 Prompt Design and Robustness

GPT-4 takes as input a list of *message* objects, and returns an inspiring Reddit post. We use *messages*, which are more interactive and dynamic compared to the classical prompt style. Specifically, we use messages with three different roles: *system*, *user* or *assistant*.⁸

⁴<https://huggingface.co/textdetox/xlmr-large-toxicity-classifier>

⁵<https://huggingface.co/facebook/roberta-hate-speech-dynabench-r4-target>

⁶https://github.com/snguyenthanh/better_profanity

⁷<https://platform.openai.com/docs/guides/text-generation/chat-completions-api>

⁸<https://help.openai.com/en/articles/7042661-chatgpt-api-transition-guide>

The prompt is first formatted with a *system* role, which sets the behavior of the model. This is followed by a conversation between the *user* and *assistant*, in a **few-shot prompting** fashion. Prior work found that LLMs function better with few-shot prompts (i.e., instructions alongside example output) rather than using zero-shot prompts (with no examples). (Brown et al., 2020)

System Prompt. We find that we can obtain high-quality responses with additional context in our prompts. Therefore, we instruct the model to be a Reddit user from either the UK or India. To ensure that the generated data is diverse and reliable, we collect five versions of *system* prompts with different phrasing, with one example shown below.

Imagine you're a person from {location} and use Reddit regularly.

User-Assistant Prompts. We design two rounds of conversations between a *user* and an *assistant*, where the *user* asks for a Reddit post or comment, and the *assistant* responds to the request. We use **few-shot prompting** by providing the *assistant* in the first round of conversation with an inspiring post, which is randomly extracted from the real posts annotated as inspiring by the majority of annotators. Finally, the answer to the second *user* is automatically generated by the *assistant* and used to collect the GPT-4 generated inspiring post. A *user* message is shown below.

Write a Reddit post or comment of maximum 100 tokens about what inspires you.

Quality Assurance. To ensure the quality of our generated data, we conduct sanity checks to review approximately 200 inspiring posts balanced across the UK and India. The posts are checked for cultural knowledge, factuality, semantic and syntax errors, and style. Based on the feedback, we find that the posts are semantically and syntactically accurate, possess cultural knowledge, and are often more complex than real posts. More information is provided in the Appendix A.1.

Cost. We generated 2,000 posts for two cultures, for a total cost of around 50\$. (0.03 per 1K input tokens and 0.06 per 1K output tokens).

4 Cross-Cultural Inspiration Analysis across Real and LLM-generated Posts

In line with previous work (Jakesch et al., 2023), we also find during manual inspection that it is chal-

lenging to differentiate between LLM-generated texts and those written by humans (see Table 2). Therefore, we perform computational linguistic analyses to compare real and LLM-generated text across cultures. This section addresses our first two research questions: *RQ1: How do inspiring posts compare across cultures?* and *RQ2: How do AI-generated inspiring posts compare to real inspiring posts across cultures?*.

4.1 Stylistic and Structural Features

We assess the linguistic style and structure of real and LLM-generated inspiring posts across India and the UK in terms of (1) analytic writing, (2) descriptiveness, and (3) readability.

Analytic Writing index measures the complexity and sophistication of the writing, which can be an indicator of advanced thinking. The formula for analytic writing is [*articles + prepositions – pronouns – auxiliary verbs – adverb – conjunctions – negations*] from LIWC scores (Jordan et al., 2019; Pennebaker et al., 2014) More information about LIWC can be found in Section 4.2. We display the scores in Table 3. The low complexity scores of Reddit posts are primarily negative due to their high usage of pronouns and lack of articles. We find that *LLM-generated inspiring posts from the UK and India are more complex than real posts*, which aligns with our initial observations from data quality checks. At the same time, there is *no significant difference in text complexity between real inspiring posts from India and those from the UK*.

Descriptiveness can be measured by the frequency of adjectives used in language patterns. Texts with high rates of adjectives tend to be more elaborate and narrative-like compared to texts with low rates of adjectives. (Chung and Pennebaker, 2008) In Table 3, we find that *LLM-generated inspiring posts from the UK and India are more descriptive than real posts*. Additionally, *real inspiring posts from India are more descriptive than those from the UK*.

Readability considers not only word count, but also word complexity. For instance, longer words are more complex than shorter ones. We use the Flesch Reading Ease metric (Flesch, 1948), which counts the number of words per sentence and syllables per word. In Table 3, we find that *LLM-generated inspiring posts are less readable than*

	Real Post 🗨️	GPT4-generated Post 🗨️
	<p>💡 <i>The youngest freedom fighter martyr of the country, the brave boy who refused to take the British cops across the river and was shot to death. His name was Baji Rout and he was 12 at the time of his death.</i></p> <p>❌ <i>What is a common meal there? I've always thought Indian palates to be quite spicy and rich.</i></p>	<p>💡 <i>Dr. APJ Abdul Kalam - His humble beginnings, insatiable thirst for knowledge, and absolute dedication to his country have been my greatest inspiration. It pushed me to work harder, dream big, and contribute to society.</i></p> <p>❌ N/A</p>
	<p>💡 <i>Dr. Helen Sharman. I'm very pleased to come from a country whose first astronaut isn't a man by default. Equality shouldn't be about women catching up, it should be about women being first 50% of the time.</i></p> <p>❌ <i>Mate all he does is play football and misses pens. Not that special</i></p>	<p>💡 <i>Absolutely love Sir David Attenborough's documentaries. His passion and commitment to preserving the environment is truly inspiring in these challenging times.</i></p> <p>❌ N/A</p>

Table 2: Random samples of inspiring (💡) and non-inspiring (❌) posts from India and UK.

		
	💡 🗨️	❌ 🗨️
Analytic	-16.7 ± 18.0	-6.9 ± 10.0
Descriptive	8.2 ± 7.8	8.9 ± 3.6
Readable	36.3 ± 63.1	12.1 ± 20.6
Word Count	61.1 ± 59.8	66.9 ± 20.3

Table 3: To what degree is the LLM-generated text (🗨️) stylistically and structurally different from the real text (🗨️)? We compute the mean and standard deviation for the inspiring posts, across cultures. The differences are statistically significant, based on the Student t-test (Student, 1908), p-value < 0.05.

real posts. Additionally, real inspiring posts from India are less readable than those from the UK. Furthermore, when measuring post length, LLM-generated UK posts are longer than real posts and UK posts are shorter than Indian posts.

4.2 Semantic and Psycholinguistic Features

We assess the semantic and psycholinguistic differences between LLM-generated and real inspiring posts across India and the UK, using topic modeling and LIWC psycholinguistic markers.

Data Pre-processing. We use `spacy` library⁹ to pre-process the data: tokenize each post, lowercase tokens, remove stop words, remove numbers, symbols, emojis, links, and lemmatize tokens.

Topic Modeling. We use Scattertext (Kessler, 2017) to create interactive visualizations of linguistic patterns. We use the n-gram representation of the text data to conduct topic extraction through sentence-level clustering. (Ramos et al., 2003)

Results. We analyze the n-gram and topic distributions by various dimensions: inspiring vs. non-inspiring, Indian vs. UK, and real vs. generated. We display the topic distribution across real and

generated UK inspiring posts in Figure 3. Further topic and n-gram distributions are in Appendix A.3.

💡 vs. ❌: Comparing inspiring to non-inspiring real posts, we find that, in the Indian data (Fig. 5, 7), the topic of *people* is frequently discussed in an inspiring context. Moreover, the most commonly occurring theme in non-inspiring posts is *bot*, encompassing content regarding Reddit rules.

In the UK data (Fig. 6, 8), the topic of *life* is amongst the most commonly occurring themes in inspiring posts, with discussions surrounding *career*, *luck*, and *pension*. On the other hand, the non-inspiring posts contain themes like *dark*, including conversations about *rain* and *winter*.

🗨️ vs. 🗨️: The LLM-generated Indian data (Fig. 10, 11) often places a significant emphasis on the topic of *inspiring*, within which common words include *dedication* and *motivational*. There is also a significant number of posts surrounding *isro* (*Indian Space Research Organisation*), featuring terms such as *space*, *mars*, and *mission*. In contrast, the real data from India contains mentions of *housing*, *summer*, and *living*, grouped under the category *live*. The *movie* topic is also popular, containing words such as *hype*, *stardom*, and *socialize*, hinting at the culture surrounding the film industry.

In the LLM-generated UK data (Fig. 3, top left), the topic *nhs* is emphasized, where common words include *pandemic*, *heroes*, and *staff*. Additionally, discussions often revolve around *adversity*, with mentions of *resilience*, *determination*, and *spirit*. In contrast, in the real data from the UK (Fig. 3, bottom right), discussions related to *job* dominate, with mentions of *salary*, *savings*, and *time*. Moreover, there is also a significant focus on *exercise*, featuring words like *discipline*, and *motivate*.

LIWC Psycholinguistic Markers. We use Linguistic Inquiry and Word Count (LIWC), a popular

⁹<https://spacy.io/models>

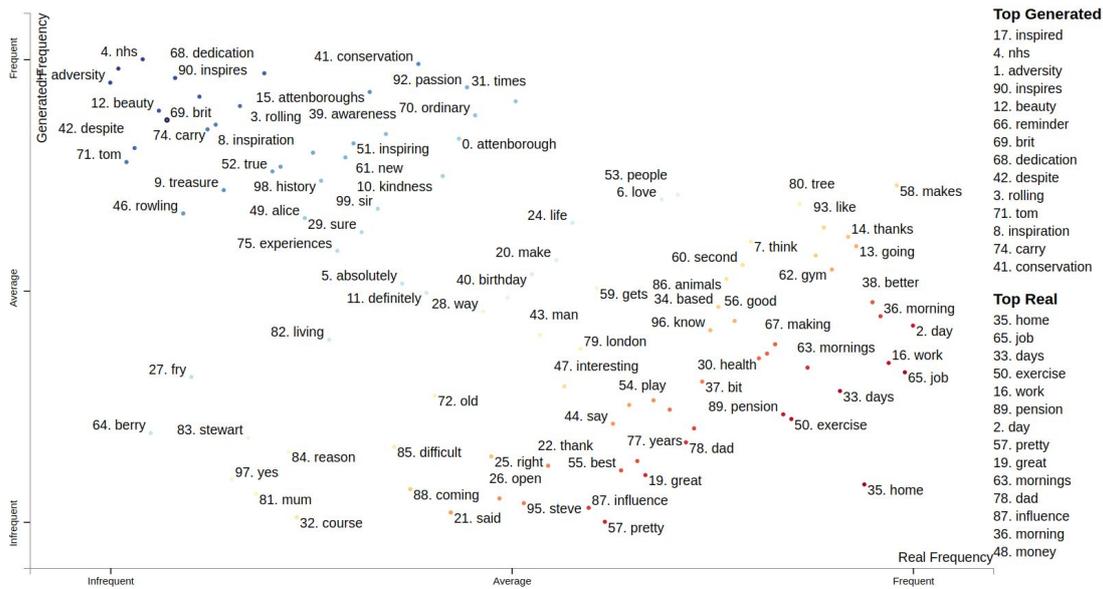


Figure 3: Visualization of **topics** used in the real and generated (🍷 vs. 🍷) inspiring posts from the UK. Points are colored red or blue based on the association of their corresponding terms with UK Real inspiring posts or UK LLM-Generated inspiring posts. The most associated topics are listed under **Top Generated** and **Top Real** headings. Interactive version: https://anonymous.4open.science/r/cross_inspiration.

text analysis tool (Pennebaker et al., 2007, 2015), to obtain the words related to human cognitive processes from each post. Specifically, we use the LIWC2015 dictionary, which contains 6,400 words and word stems, each related to a cognitive category. As an example, the word “mother” is assigned the following cognitive categories: *female*, *family*, *social*. Even though it is less biased than embedding-based models, LIWC has limitations which we discuss in the Limitation section.

Results. We display the top 24 categories and their LIWC scores, with the most significant differences across dimensions in Table 4. We find that *male* words are more frequent than *female*, the most common pronoun is *I*, the most frequent tense is *present*, and the most common emotion is *positive*.

💡 vs. ❌: Comparing inspiring to non-inspiring real posts, we find that the Indian inspiring posts are more related to *socializing*, *insight*, *feelings*, *perception*, *affection*, and contain *more positive emotions*. Additionally, Indian inspiring posts tend to have *more comparisons* and use more words related to *achievement*, *health*, *reward*, and *work*.

UK real inspiring posts contain more words related to *affection*, *comparisons*, *feelings*, *achievement*, *health*, *home*, *leisure*, *money*, *reward*, and *work* than non-inspiring posts. Furthermore, compared to Indian inspiring posts, UK inspiring posts have fewer words related to *family*, *socializing*, *affection*, *perception*, *religion*, *less positive emotions* and more words related to *achievement*, *health*,

home, *leisure*, *money*, *rewards*, and *work*.

🍷 vs. 🍷: Comparing real to LLM-generated inspiring posts, we find that real Indian posts tend to include more words related to *family*, *social interactions*, *comparisons*, *feelings*, *perceptions* as well as *home*, *leisure* and *rewards*. Conversely, real Indian posts contain fewer words related to *affection*, *insight*, *achievement*, *health*, and *religion*.

UK real inspiring posts contain more words related to *comparisons*, *feelings*, *health*, *home*, *leisure*, *money*, *rewards* and *work* than LLM-generated UK posts. Conversely, real UK posts contain fewer words related to *socializing*, *affection*, *insight*, *perception*, *achievement* and *religion* than LLM-generated UK posts. Furthermore, compared to LLM-generated Indian posts, LLM-generated UK posts have fewer words related to *socializing*, *achievement*, *leisure*, *money*, *religion*, *work* and more words related to *affection*, *perception* and *more positive emotions*.

5 Cross-Cultural Inspiration Detection across Real and LLM-generated posts

To answer our last research question – *RQ3: Can detection models effectively differentiate inspiring posts across diverse cultures and data sources?* – we fine-tune several classification models to identify if a post is inspiring, represents India or the UK, and whether it is real or LLM-generated.

Implementation Details. As a baseline, we use a Random Forest classifier (Ho, 1995) with the

LIWC	🇮🇳		🇬🇧			
	👉	👎	👉	👎	👉	👎
Social Processes						
FAMILY	0.4	0.4	0.2	0.2	0.1	0.2
FRIEND	0.3	0.4	0.2	0.2	0.4	0.2
FEMALE	0.4	0.3	0.2	0.4	0.4	0.2
MALE	1.1	1.0	0.9	0.9	1.3	1.7
SOCIAL	9.3	7.7	8.6	7.4	8.1	8.1
Affective Processes						
AFFECT	6.1	4.3	8.4	5.4	4.8	9.4
NEGEMO	1.4	1.3	1.0	1.3	1.6	0.9
POSEMO	4.5	3.0	7.3	4.1	3.1	8.3
Cognitive Processes						
COMPARE	3.0	2.4	1.8	3.0	2.7	1.8
INSIGHT	2.3	2.0	4.8	2.2	2.1	4.9
Perceptual Processes						
FEEL	0.6	0.4	0.4	0.7	0.3	0.4
PERCEPT	2.4	2.0	2.1	2.1	2.1	2.4
Personal Concerns						
ACHIEV	1.9	1.0	3.5	2.4	1.6	2.6
HEALTH	0.7	0.4	0.9	1.0	0.8	0.8
HOME	0.3	0.2	0.1	0.7	0.4	0.2
LEISURE	1.5	1.4	1.2	1.7	1.4	0.9
MONEY	0.7	0.8	0.6	1.2	0.8	0.4
RELIG	0.4	0.5	1.1	0.1	0.1	0.6
REWARD	2.0	1.1	1.0	2.2	1.6	1.1
WORK	3.0	1.9	3.0	3.9	2.7	1.9

Table 4: Comparing LIWC scores across culture (India and the UK) in inspiring vs. non-inspiring (👉 vs. 👎) and real vs. LLM-generated posts (👉 vs. 👎).

default settings from `sklearn` (Pedregosa et al., 2011) and TF-IDF (Ramos et al., 2003) as features.

As a main model, we use the XLM-RoBERTa base model (Conneau et al., 2019), pre-trained on 2.5TB of filtered CommonCrawl (Wenzek et al., 2020). We fine-tune the model on our dataset for five epochs with a learning rate of $2e-5$ and a batch size of 8. Finally, we also use the pre-trained Llama 2.7b model, from HuggingFace.¹⁰ We fine-tune the model on our dataset using LoRA (Hu et al., 2022) for five epochs with a rank of 4, learning rate of $2e-4$ and a batch size of 2.¹¹

Model Training Setup. We experiment with two setups to split the data into training, validation, and test sets: a *default* train-val-test split of 64-16-20% and a *few-shot* split of 8-2-90%.

Results. Since the dataset is evenly distributed across classes, a random baseline results in a 50% accuracy score. Figure 4 displays the disaggregated results for each label. The Random Forest baseline with TF-IDF scores, trained in a *default* setup achieves high accuracy scores (88 – 99), suggesting that important words and topics contribute significantly to inspiration detection across culture and source. This finding is also supported by the topic modeling results in Section 4.2, which shows

¹⁰<https://huggingface.co/meta-llama/Llama-2-7b>

¹¹All the model parameters are shared in the repository.

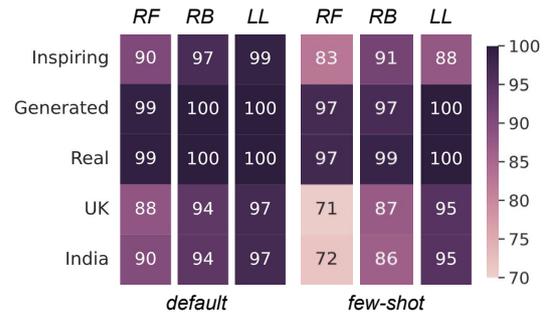


Figure 4: Classification test accuracy with the *few-shot* and *default* setups with the Random Forest TF-IDF (RF), XLM-RoBERTa base (RB), and Llama 2.7b (LL) models.

different topic distributions for each class. Given that Random Forest is an interpretable model, we encourage future work to further analyse the importance of different features for inspiration detection. Finally, we find that even with very few training data (600 posts), in the *few-shot* setup, the main models, Llama 2.7b and XLM-RoBERTa, learn to accurately distinguish inspiring content across cultures (India, UK) and data sources (real and generated), without significant differences in performance across cultures and data sources.

6 Conclusion

In this paper, we introduced the task of cross-cultural inspiration detection and generation in social media data. To facilitate research in this domain, we released INSPAIRED, a dataset of 2,000 real inspiring posts, 2,000 real non-inspiring posts, and 2,000 LLM-generated inspiring posts, evenly distributed across India and the UK. We performed extensive linguistic data analyses to gain insight into what topics inspire each culture and compare AI-generated inspiring posts to real inspiring posts across various linguistic dimensions. Despite the difficulty humans have in distinguishing between real posts and those generated by LLMs, we found that these posts have noticeable differences in style, structure, and semantics and that, even with little data, fine-tuned models accurately distinguish inspiring content across cultures and data sources.

We hope our work will enable the exploration of various applications to improve creativity and motivation, including storytelling, advertising, and social media, as well as therapy and coaching. Our dataset is available to test, fine-tune, and analyze other models, and it can be accessed alongside our classification models at https://anonymous.4open.science/r/cross_inspiration.

615 **Limitations**

616 **A more fine-grained data split.**

617 Our posts are divided into two main categories -
618 inspiring posts from India and inspiring posts from
619 the UK. Initially, we tried to create a more detailed
620 classification at the region or city level, but we
621 faced difficulties in finding annotators from those
622 specific regions. However, we encourage future
623 research to explore this direction and investigate
624 how inspiration varies within a country or region,
625 as well as explore other demographic information,
626 such as language, age, gender, or income.

627 **Study case for India.** The extensive discussions
628 on Indian subreddits, like r/AskIndia and r/India,
629 offer straightforward access to text-based motiva-
630 tional narratives. Additionally, region-based sub-
631 Reddits, like r/Chennai and r/Mumbai, provide di-
632 verse insights into localized experiences and dis-
633 cussions. The nature of posts on this topic is mostly
634 anecdotal. Reddit users contribute by sharing per-
635 sonal experiences that have inspired them.

636 **A limited number of cross-cultural inspiring** 637 **posts.**

638 It is crucial to recognize that inspiration is not al-
639 ways explicitly articulated through words like “in-
640 spiring” or “motivation”, especially within cultures
641 that aren’t as open about talking about such topics.
642 In many instances, it manifests more subtly, em-
643 bedded within narratives, imagery, or cultural ex-
644 pressions. This implicit nature of inspiration adds
645 another layer of complexity to the data collection
646 process.

647 The language barrier also introduces an addi-
648 tional barrier, as it requires linguistic fluency and
649 cultural understanding to interpret and analyze in-
650 spiring content effectively. Posts in languages other
651 than English may contain nuances and cultural re-
652 ferences that are not easily translatable. Moreover,
653 it becomes more challenging to collect inspiring
654 content when data collection is restricted to certain
655 countries. That is why we collected fewer posts
656 compared to Ignat et al.’s “general” inspiring posts.

657 **Data generated with a closed-source model.**

658 We use GPT-4 to generate the inspiring posts,
659 which is not an open-source LLM. We believe it
660 would be valuable for future work to consider gen-
661 erating more data using open-source models like
662 LLaMA or BLOOM. At the same time, we pub-
663 licly released the data generated with GPT-4 so

664 that others can build on this dataset, and chose this
665 model due to its SOTA performance, worldwide
666 accessibility, and popularity.

667 **Relevance of LLM-based data to current times.**

668 In our reliance on LLM-based data, it’s imperative
669 to recognize the temporal limitations inherent in its
670 training corpus. While these models offer remark-
671 able capabilities in understanding and generating
672 text, they might not fully capture the current cul-
673 tural and societal context in their outputs. This
674 limitation can be primarily attributed to the fact
675 that these models are trained on data only up to a
676 specific cutoff date, i.e., GPT-4 only learned from
677 data dated up to January 2022.

678 **Limitations of the LIWC-based data analysis.**

679 LIWC is a dictionary where each word is asso-
680 ciated with a psycholinguistics category. There-
681 fore, it is less biased than embedding-based models
682 that often contain biased word associations (e.g.,
683 man-programmer, woman-homemaker). Based on
684 a simple and scalable word count approach, LIWC
685 is one of the most commonly used data analysis
686 tools. However, it has several limitations. First,
687 it does not consider word sense disambiguation
688 based on context, which can lead to inaccurate re-
689 sults (Schwartz et al., 2013). Second, the model
690 was developed and validated for the analysis of
691 long-form writing, which may impact its valid-
692 ity on short social media posts (Panger, 2016).
693 Third, ongoing research is being conducted to
694 measure the extent to which LIWC demonstrates
695 cross-linguistic and cross-cultural generalizability,
696 given that most studies have been conducted in
697 English (Dudău and Sava, 2021).

698 **References**

- 699 Josh Achiam, Steven Adler, Sandhini Agarwal, Lama
700 Ahmad, Ilge Akkaya, Florencia Leoni Aleman,
701 Diogo Almeida, Janko Altschmidt, Sam Altman,
702 Shyamal Anadkat, et al. 2023. Gpt-4 technical report.
703 *arXiv preprint arXiv:2303.08774*.
- 704 Tom Brown, Benjamin Mann, Nick Ryder, Melanie
705 Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind
706 Neelakantan, Pranav Shyam, Girish Sastry, Amanda
707 Askell, et al. 2020. Language models are few-shot
708 learners. *Advances in neural information processing*
709 *systems*, 33:1877–1901.
- 710 Cindy K Chung and James W Pennebaker. 2008. Re-
711 vealing dimensions of thinking in open-ended self-
712 descriptions: An automated meaning extraction

713	method for natural language. <i>Journal of research in personality</i> , 42(1):96–132.	767
714		768
715	Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale . <i>CoRR</i> , abs/1911.02116.	769
716		770
717		771
718		772
719		773
720		774
721	Samuel Rhys Cox, Ashraf Abdul, and Wei Tsang Ooi. 2023. Prompting a large language model to generate diverse motivational messages: A comparison with human-written messages. In <i>Proceedings of the 11th International Conference on Human-Agent Interaction</i> , pages 378–380.	775
722		776
723		777
724		778
725		779
726		780
727	David Dale, Anton Voronov, Daryna Dementieva, Varvara Logacheva, Olga Kozlova, Nikita Semenov, and Alexander Panchenko. 2021. Text detoxification using large pre-trained neural models . In <i>Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing</i> , pages 7979–7996, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.	781
728		782
729		783
730		784
731		785
732		786
733		787
734		788
735	Katherine R Dale, Arthur A Raney, Sophie H Janicke, Meghan S Sanders, and Mary Beth Oliver. 2017. Youtube for good: A content analysis and examination of elicitors of self-transcendent media. <i>Journal of Communication</i> , 67(6):897–919.	789
736		790
737		791
738		792
739		793
740	Katherine R Dale, Arthur A Raney, Qihao Ji, Sophie H Janicke-Bowles, Joshua Baldwin, Jerrica T Rowlett, Cen Wang, and Mary Beth Oliver. 2020. Self-transcendent emotions and social media: Exploring the content and consumers of inspirational facebook posts. <i>New Media & Society</i> , 22(3):507–527.	794
741		795
742		796
743		797
744		798
745		799
746	Diana Paula Dudău and Florin Alin Sava. 2021. Performing multilingual analysis with linguistic inquiry and word count 2015 (liwc2015). an equivalence study of four languages . <i>Frontiers in Psychology</i> , 12.	800
747		801
748		802
749		803
750		804
751	Andrew J Elliot and Todd M Thrash. 2002. Approach-avoidance motivation in personality: approach and avoidance temperaments and goals. <i>Journal of personality and social psychology</i> , 82(5):804.	805
752		806
753		807
754		808
755	Eve Fleisig, Rediet Abebe, and Dan Klein. 2023. When the majority is wrong: Modeling annotator disagreement for subjective tasks . In <i>Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing</i> , pages 6715–6726, Singapore. Association for Computational Linguistics.	809
756		810
757		811
758		812
759		813
760		814
761	Joseph L Fleiss and Jacob Cohen. 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. <i>Educational and psychological measurement</i> , 33(3):613–619.	815
762		816
763		817
764		818
765	Rudolph Flesch. 1948. A new readability yardstick. <i>Journal of applied psychology</i> , 32(3):221.	819
766		820
		821
		822
	Yi R Fung, Tuhin Chakraborty, Hao Guo, Owen Rambow, Smaranda Muresan, and Heng Ji. 2022. Normsage: Multi-lingual multi-cultural norm discovery from conversations on-the-fly. <i>arXiv preprint arXiv:2210.08604</i> .	
	Tin Kam Ho. 1995. Random decision forests. In <i>Proceedings of 3rd international conference on document analysis and recognition</i> , volume 1, pages 278–282. IEEE.	
	Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-rank adaptation of large language models . In <i>International Conference on Learning Representations</i> .	
	Jing Huang and Diyi Yang. 2023. Culturally aware natural language inference. In <i>Findings of the Association for Computational Linguistics: EMNLP 2023</i> , pages 7591–7609.	
	Oana Ignat, Y-Lan Boureau, Jane A. Yu, and Alon Y. Halevy. 2021. Detecting inspiring content on social media . <i>2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)</i> , pages 1–8.	
	Ali Shariq Imran, Sher Muhammad Daudpota, Zenun Kastrati, and Rakhi Batra. 2020. Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on covid-19 related tweets. <i>Ieee Access</i> , 8:181074–181090.	
	Maurice Jakesch, Jeffrey T Hancock, and Mor Naaman. 2023. Human heuristics for ai-generated language are flawed. <i>Proceedings of the National Academy of Sciences</i> , 120(11):e2208839120.	
	Kayla N Jordan, Joanna Sterling, James W Pennebaker, and Ryan L Boyd. 2019. Examining long-term trends in politics and culture through language of political leaders and cultural institutions. <i>Proceedings of the National Academy of Sciences</i> , 116(9):3476–3481.	
	Elise Karinshak, Sunny Xun Liu, Joon Sung Park, and Jeffrey T Hancock. 2023. Working with ai to persuade: Examining a large language model’s ability to generate pro-vaccination messages. <i>Proceedings of the ACM on Human-Computer Interaction</i> , 7(CSCW1):1–29.	
	Jason Kessler. 2017. Scattertext: a browser-based tool for visualizing how corpora differ . In <i>Proceedings of ACL 2017, System Demonstrations</i> , pages 85–90, Vancouver, Canada. Association for Computational Linguistics.	
	J.H. Leavitt and P.J. Leavitt. 1997. <i>Poetry and Prophecy: The Anthropology of Inspiration</i> . Studies in Literature and Science. University of Michigan Press.	
	Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. <i>ArXiv</i> , abs/1907.11692.	

823	Marina Milyavskaya, Iana Ianakieva, Emily Foxen-Craft, Agnes Colantuoni, and Richard Koestner. 2012. Inspired to get there: The effects of trait and goal inspiration on goal progress.	877
824		878
825		879
826		880
827	Soichiro Murakami, Sho Hoshino, and Peinan Zhang. 2023. Natural language generation for advertising: A survey . <i>Preprint</i> , arXiv:2306.12719.	881
828		882
829		883
830	Victoria C. Oleynick, Todd M. Thrash, Michael C. LeFevre, Emil G. Moldovan, and Paul D. Kieffaber. 2014. The scientific study of inspiration in the creative process: challenges and opportunities. <i>Frontiers in Human Neuroscience</i> , 8.	884
831		885
832		886
833		887
834		888
835	Bo Pang, Lillian Lee, et al. 2008. Opinion mining and sentiment analysis. <i>Foundations and Trends® in information retrieval</i> , 2(1–2):1–135.	890
836		891
837		
838	Galen Panger. 2016. Reassessing the facebook experiment: critical thinking about the validity of big data research . <i>Information, Communication & Society</i> , 19:1108 – 1126.	892
839		893
840		894
841		
842	F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. <i>Journal of Machine Learning Research</i> , 12:2825–2830.	895
843		896
844		897
845		
846		
847		
848		
849	Jiaxin Pei, Aparna Ananthasubramaniam, Xingyao Wang, Naitian Zhou, Apostolos Dedeloudis, Jackson Sargent, and David Jurgens. 2022. Potato: The portable text annotation tool. In <i>Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations</i> .	898
850		899
851		900
852		901
853		
854		
855	James W. Pennebaker, Roger John Booth, and Martha E. Francis. 2007. Linguistic inquiry and word count (LIWC2007) .	902
856		903
857		904
858	James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, and Kate G. Blackburn. 2015. The development and psychometric properties of LIWC2015 .	905
859		906
860		
861	James W Pennebaker, Cindy K Chung, Joey Frazee, Gary M Lavergne, and David I Beaver. 2014. When small words foretell academic success: The case of college admissions essays. <i>PloS one</i> , 9(12):e115844.	907
862		908
863		909
864		910
865	Aida Ramezani and Yang Xu. 2023. Knowledge of cultural moral norms in large language models. <i>arXiv preprint arXiv:2306.01857</i> .	911
866		912
867		913
868	Juan Ramos et al. 2003. Using tf-idf to determine word relevance in document queries. In <i>Proceedings of the first instructional conference on machine learning</i> , volume 242, pages 29–48. Citeseer.	914
869		915
870		916
871		917
872	Arthur A Raney, Sophie H Janicke, Mary Beth Oliver, Katherine R Dale, Robert P Jones, and Daniel Cox. 2018. Profiling the audience for self-transcendent media: A national survey. <i>Mass Communication and Society</i> , 21(3):296–319.	918
873		919
874		920
875		921
876		922
	Diana Rieger and Christoph Klimmt. 2019. The daily dose of digital inspiration: A multi-method exploration of meaningful communication in social media. <i>New Media & Society</i> , 21(1):97–118.	923
		924
		925
		926
		927
		928
	H Andrew Schwartz, Johannes Eichstaedt, Eduardo Blanco, Lukasz Dziurzynski, Margaret Kern, Stephanie Ramones, Martin Seligman, and Lyle Ungar. 2013. Choosing the right words: Characterizing and reducing error of the word count approach. In <i>Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 1: Proceedings of the Main Conference and the Shared Task: Semantic Textual Similarity</i> , pages 296–305.	929
		930
		931
		932
	Student. 1908. The probable error of a mean. <i>Biometrika</i> , pages 1–25.	
	Ruixiang Tang, Yu-Neng Chuang, and Xia Hu. 2023. The science of detecting llm-generated texts. <i>arXiv preprint arXiv:2303.07205</i> .	
	Todd M. Thrash and Andrew J Elliot. 2003. Inspiration as a psychological construct. <i>Journal of personality and social psychology</i> , 84 4:871–89.	
	Todd M. Thrash and Andrew J Elliot. 2004. Inspiration: core characteristics, component processes, antecedents, and function. <i>Journal of personality and social psychology</i> , 87 6:957–73.	
	Todd M. Thrash, Laura A. Maruskin, Scott E. Cassidy, James W. Fryer, and Richard M. Ryan. 2010. Meditating between the muse and the masses: inspiration and the actualization of creative ideas. <i>Journal of personality and social psychology</i> , 98 3:469–87.	
	Bertie Vidgen, Tristan Thrush, Zeerak Waseem, and Douwe Kiela. 2021. Learning from the worst: Dynamically generated datasets to improve online hate detection. In <i>ACL</i> .	
	Guillaume Wenzek, Marie-Anne Lachaux, Alexis Conneau, Vishrav Chaudhary, Francisco Guzmán, Armand Joulin, and Edouard Grave. 2020. CCNet: Extracting high quality monolingual datasets from web crawl data . In <i>Proceedings of the Twelfth Language Resources and Evaluation Conference</i> , pages 4003–4012, Marseille, France. European Language Resources Association.	
	Junchao Wu, Shu Yang, Runzhe Zhan, Yulin Yuan, Derek F. Wong, and Lidia S. Chao. 2023. A survey on llm-generated text detection: Necessity, methods, and future directions . <i>ArXiv</i> , abs/2310.14724.	
	Yuta Yanagi, Ryohei Orihara, Yuichi Sei, Yasuyuki Tahara, and Akihiko Ohsuga. 2020. Fake news detection with generated comments for news articles. In <i>2020 IEEE 24th International Conference on Intelligent Engineering Systems (INES)</i> , pages 85–90. IEEE.	
	Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: story writing with large language models. In <i>27th International Conference on Intelligent User Interfaces</i> , pages 841–852.	

A Appendix

A.1 Data Quality Assurance

We mention in the paper that the posts are checked for cultural knowledge, factuality, semantic and syntax errors, and style. Specifically, the authors each check 200 real and generated posts from each culture. One author is from India and is, therefore, able to authenticate the cultural and factual aspects of the sampled posts (e.g., Sal Khan and his education contributions, the ISRO, and its scientific progress in India) and check the correctness of the few (<10% of the data) code switch posts between English and Tamil. All the authors possess cultural knowledge about the UK, to recognize the cultural and factual elements of the posts. When unsure, they manually searched Wikipedia for the factuality of the posts (e.g., Sir David Attenborough’s contributions to nature, the British Nicola Adams, the first-ever female boxer to retain an Olympic title).

Finally, all the authors possess proficient English skills to check for semantic and syntax errors in the sampled posts. We find that apart from a few typos (only in real posts), very common in social media posts, the vast majority of the posts (98%) are correct. In terms of style, we observe that LLM-generated sampled posts are more formal and complex than real posts (e.g., there are no typos, and there are more articles and prepositions). Our analysis also confirmed this by automatically measuring the analytic index.

A.2 Fine-tuning Process

XLm-RoBERTa Model. Implementation Details. We experiment with three different fine-tuned classification models for the purpose of weakly labeling the presence of inspiration in the dataset. The first attempt involved fine-tuning the model on the general inspiration dataset introduced in (Ignat et al., 2021). In each of the experiments, the data is split into three subsets: training, validation, and test sets, using an 80:10:10 ratio. The fine-tuning process involved training the model for 5 epochs with a learning rate of $2e - 5$ and a batch size of 8. We monitored the model performance on a separate validation set and selected the best model checkpoint based on accuracy metric.

In the second experiment, we focused on a subset of the dataset used in annotation, consisting of 200 posts. These posts were labeled as inspiring if at least one user considered them to be so. We

fine-tuned the base XLm-RoBERTa model using this subset, maintaining the same training configurations as in the initial experiment. In the last experiment, we leveraged the user-annotated posts to further refine the model trained on the general inspiration dataset from the first experiment. This approach aimed to leverage the insights gained from the initial fine-tuning process and apply them to the user-annotated subset, thereby refining the model to better capture the nuances present in this specific dataset.

A.3 Topic Analysis

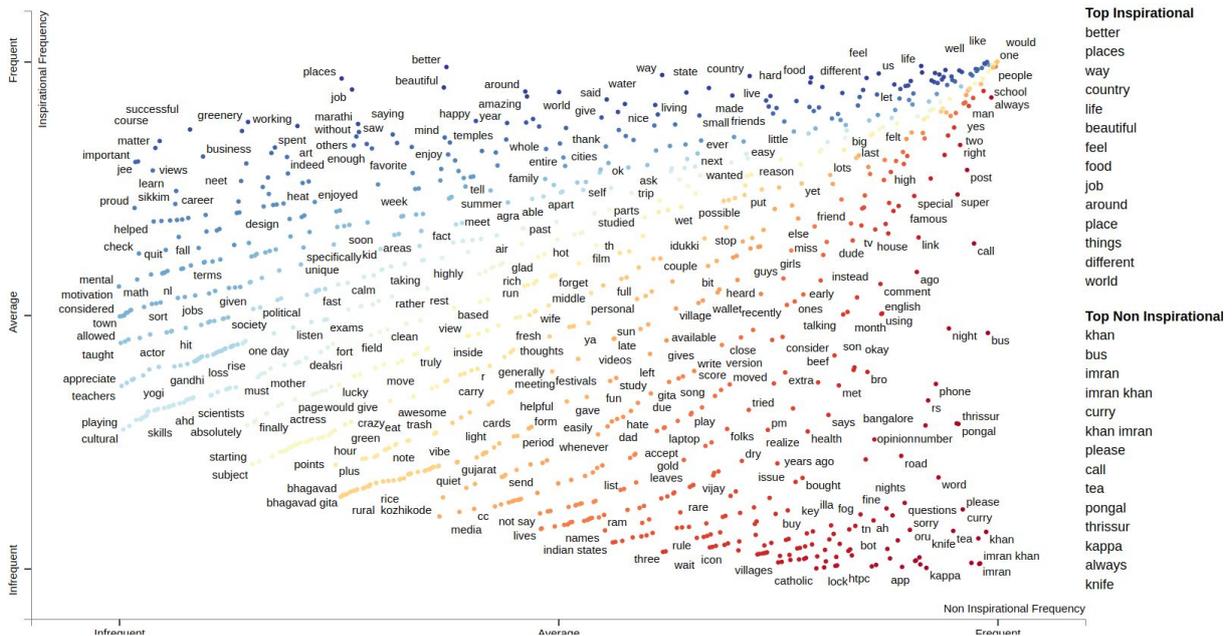


Figure 5: Scattertext visualization of **unigrams** used in the real inspiring and non-inspiring (💡 vs. ❌) Reddit posts from **India**. Points are colored in red or blue based on the association of their corresponding terms with Indian Non-inspiring posts or Indian inspiring posts. The most associated terms are listed under “Top inspiring” and “Top Non-inspiring” headings.

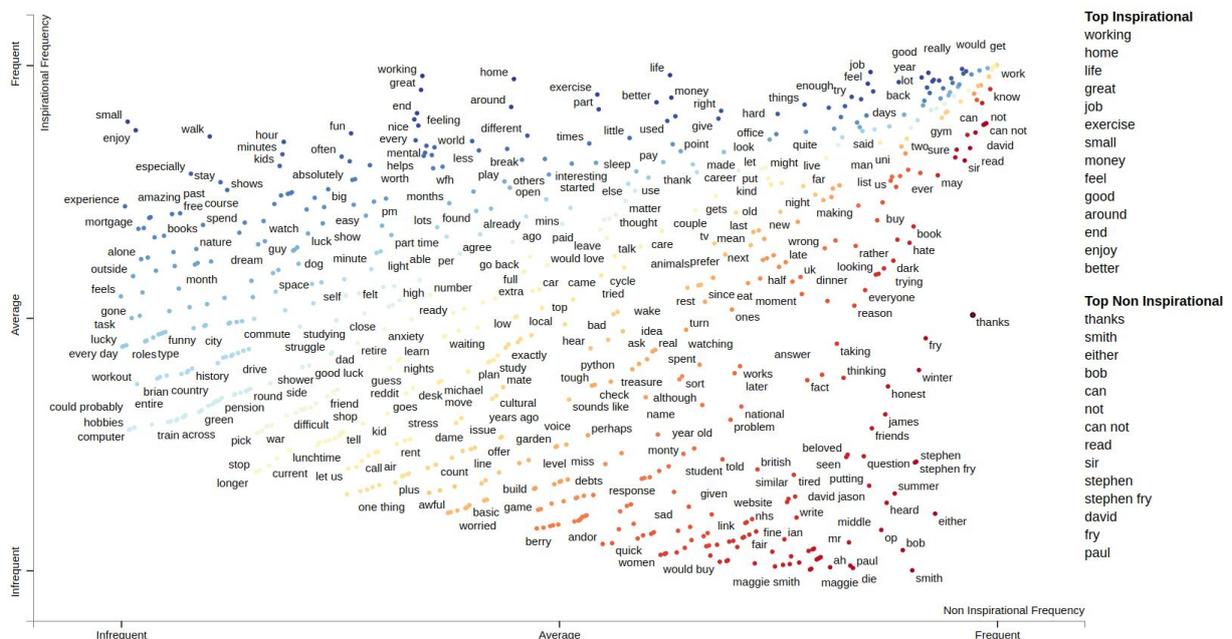


Figure 6: Scattertext visualization of **unigrams** used in the real inspiring and non-inspiring (💡 vs. ❌) Reddit posts from the **UK**. Points are colored in red or blue based on the association of their corresponding terms with UK Non-inspiring posts or UK inspiring posts. The most associated terms are listed under “Top inspiring” and “Top Non-inspiring” headings.

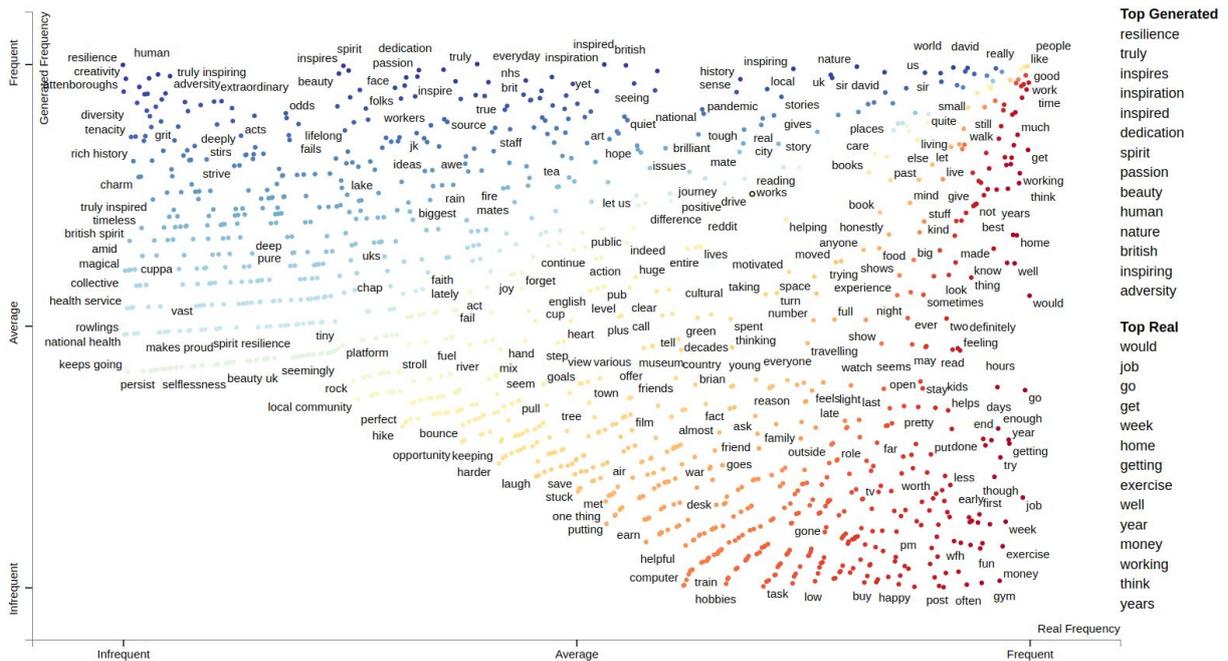


Figure 9: Scattertext visualization of **unigrams** used in the real and generated (🍄 vs. 📄) inspiring posts from the **UK**. Points are colored **red** or **blue** based on the association of their corresponding terms with the UK Real inspiring posts or the UK LLM-Generated inspiring posts. The most associated topics are listed under **Top Generated** and **Top Real** headings.

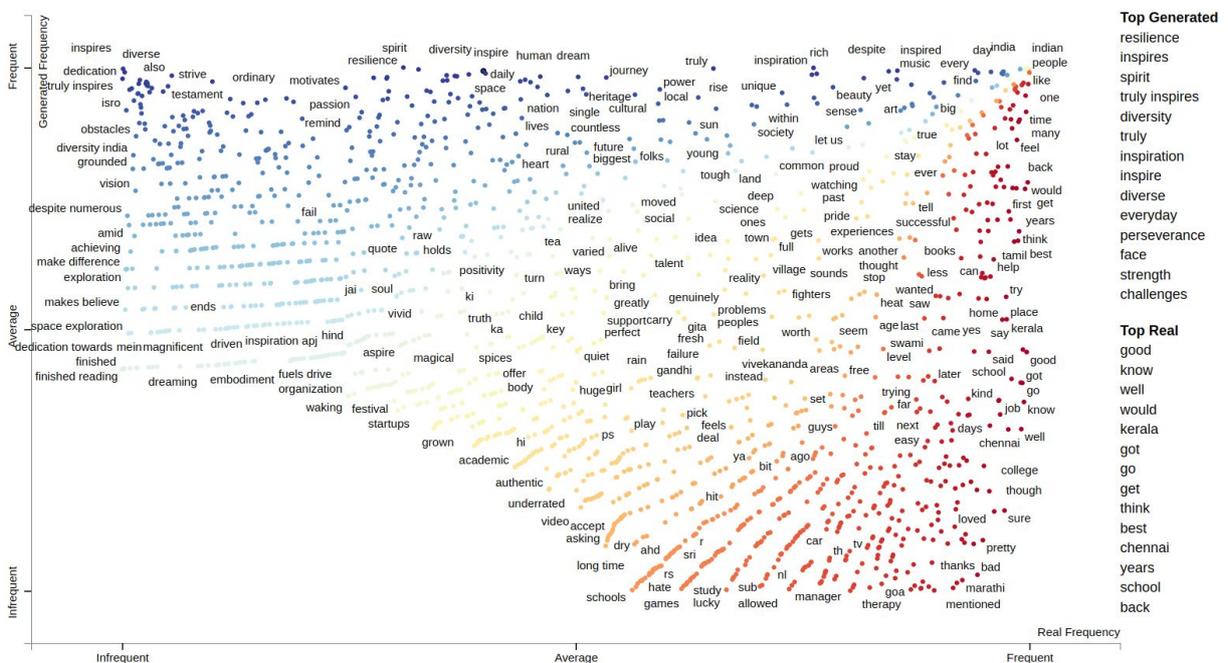


Figure 10: Scattertext visualization of **unigrams** used in the real and generated (🍄 vs. 📄) inspiring posts from **India**. Points are colored **red** or **blue** based on the association of their corresponding terms with India Real inspiring posts or India LLM-Generated inspiring posts. The most associated topics are listed under **Top Generated** and **Top Real** headings.

