
Examining Race + Gender in Folktales and LLM-Generated Folk Illustration Descriptions

Rhea Kapur

Department of Computer Science, Linguistics
Stanford University
rheak@stanford.edu

1 Problem Statement

In this project, I am asking the following questions: how is race and gender expressed and discussed in Slavic and Mexican folktales? And, in terms of accompanying folktale illustrations in a story, how do GPT-4o-generated descriptions for these images discuss race, gender, and power? This comes from my broader goal to understand the role of race in traditional folklore (particularly when the stories are translated to the western lens) using the range and power of a computational approach. This project is more of a proof of concept that incorporating NLP techniques with traditional literary analysis of folktales will lend meaningful results.

2 Dataset

For the folktale texts, I constructed my own small pilot dataset with 8 tales: 4 from Mexico and 4 from Russia. Their titles are: Tlahuelpuchi, La Llorona, Charro Negro, La Lechuza, Vasilisa the Beautiful, Koschei the Deathless, Father Frost, and Tsar Saltan. I cut this dataset down significantly from what I had before (16 tales) for the podcast presentation for ease of analysis and so I could refer back to each tale. I sourced the texts from websites online that offered English translations (I held off on analyzing the texts in Russian and Spanish for lack of time and several issues I encountered, but I highly recommend doing this in the future to better represent for example word endings that might have different taggings for different genders, but that don't explicitly reference a man, woman, etc.).

For the folktale illustrations, I first downloaded 5 Russian folktale illustrations and 5 Mexican folktale illustrations (corresponding to the tales I selected above) from Google Images. I then programmatically prompted OpenAI's GPT-4o to generate image descriptions with the following instructions: "Describe who/what you see in this image and analyze race, gender and power dynamics at play with those present. Don't neglect any of these three areas. Keep it to one paragraph." I again kept this dataset small as my project is more of a proof-of-concept or exploration and I wanted to be able to examine generated descriptions myself as well.

3 Methods

For the textual analysis of the Mexican and Russian folktales, I first manually tagged each sentence in each folktale as mentioning men, women, both, or neither (nonbinary, nonhuman, etc.).

Next, I follow Bamman et al. [2014]: "Gender identity and lexical variation in social media," which was recommended to me by Myra. Specifically, I am analyzing emotion terms ("happy", "sad", "love", etc.) and kinship terms ("father", "mother", etc. as those authors do. They also look at pronouns, conjunctions, and other parts of speech of words in their data, but I am interested in how the distribution of nouns, adjectives, and verbs in folktale data, as from my cursory read I feel like the choices and amount of these parts of speech is telling in terms of the gender and racial dynamics in the

tales. First for Mexican stories and then for the Russian ones, I prompted GPT-4o to tag each sentence for any adjectives, nouns, verbs, emotion terms, and kinship terms (for the latter two, I recommended it use its own discretion as to what qualified after giving some preliminary recommendations). I then extracted counts of each of these word-level categories and the top 5 most frequent words in each (if they existed).

For my analysis of the LLM-generated folktale illustration descriptions, I simply repeated all that I did above but for each sentence in the 10 image descriptions.

Treating each sentence as a document and the gender tag as the collection type (male, female, both, neither), I also experimented with LDA Topic Modeling on the textual folktale data. I felt this didn't reveal any new results that weren't already visible from the analysis above, so I've simply included it in the appendix. I didn't repeat this for the folktale illustration descriptions.

I considered using CLIPScore (Hessel et al. [2021]) to evaluate how informative the GPT-4o-generated descriptions were of the folktale illustrations, but CLIPScore has a limit of 77 tokens and I prompted for a paragraph-length description. In the future, this could be an interesting avenue to explore (there are also variations that account for longer captions).

4 Results for Analysis of Folktale Texts

4.1 From Mexico

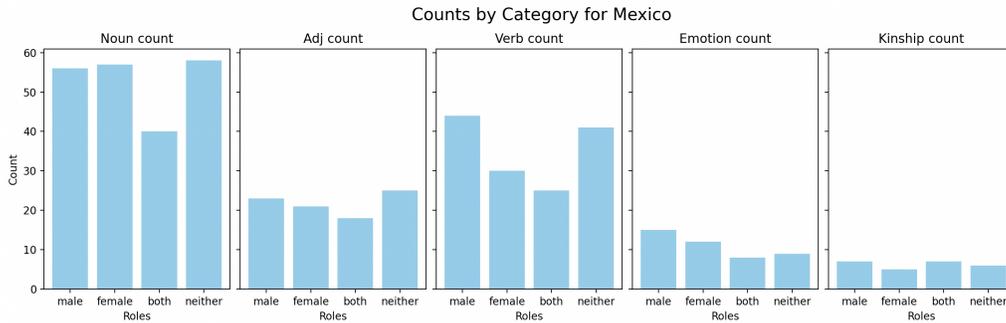


Figure 1: Counts by word-level category for Mexican folktale texts.

Table 1: Top Words By POS/Gender Segmentation in Mexican Tales

Category	Top 5 Words
Adjectives	
Male	Upper-class, good, few, unending, frightened
Female	Native, last, painful, deep, disappeared
Both	Native, Spanish, beautiful, constant, deep
Neither	Black, white, cursed, fatal, elegant
Nouns	
Male	Charro, devil, money, soul, parents
Female	Woman, children, love, life, legend
Both	Woman, children, relationship, love, Lechuza
Neither	Family, Tlahuelpuchi, blood, charro, death
Verbs	
Male	Loved, asked, said, took, went
Female	Found, gave, took, heard, believed
Both	Had, gave, cared, wanted, took
Neither	Said, believed, feed, appear, transform
Emotion Terms	
Male	Fear, loneliness, misery, love, greed
Female	Devastated, desperate, guilt, pain, regret

Category	Top 5 Words
Both	Love, sad, unconsciousness, death, fear
Neither	Fear, shame, loyal, brave, sad
Kinship Terms	
Male	Parents, man, master, brother, sister
Female	Children, mother, woman, families, community
Both	Mother, father, child, family, member
Neither	Family, member, infant, children, coward

From Figure 1, we see that in Mexican folktales, more verbs are used in sentences involving men than in sentences involving women. The number of adjectives used is near-equal. The number of nouns used also does not seem to vary greatly besides a relative drop in the "both" category. The kinship term counts seem relatively evenly distributed; there seems to be a slightly higher number of emotion terms used in contexts where men are described rather than women.

Table 1 offers a qualitative look at the data above; I will hold off on analyzing it until the discussion.

4.2 From Russia

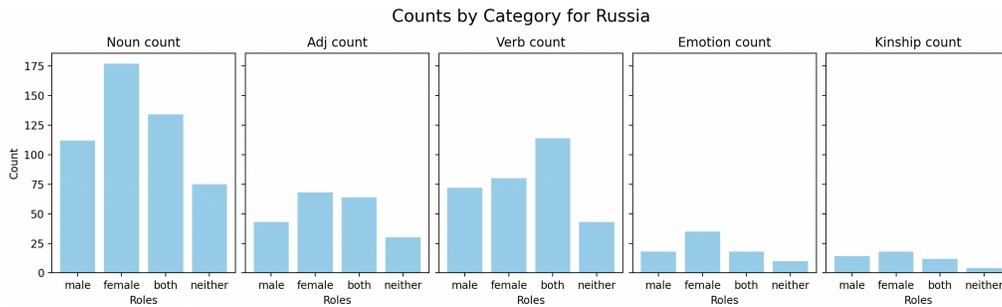


Figure 2: Counts by word-level category for Russian folktales.

Table 2: Top Words By POS/Gender Segmentation in Russian Tales

Category	Top 5 Words
Adjectives	
Male	White, black, red, lonely, happy
Female	Beautiful, ugly, little, heavy, terrible
Both	Old, good, beautiful, happy, kind
Neither	Terrifying, human, sharp, black, golden
Nouns	
Male	Prince, Ivan, father, horse, brothers
Female	Vasilisa, mother, doll, stepmother, Baba-Yaga
Both	Mother, Prince, Ivan, Koshchei, daughter
Neither	Forest, witch, skulls, light, tasks
Verbs	
Male	Said, went, came, saw, let
Female	Said, asked, could, gave, was
Both	Said, went, asked, gave, came
Neither	Show, see, give, eat, carry
Emotions	
Male	Happy, sad, surprised, frightened, unwell
Female	Fear, love, hate, joy, sadness
Both	Love, sad, happy, jealous, grieved
Neither	Gloomy, tears, happiness, love, fear
Kinships	

Category	Top 5 Words
Male	Father, sons, brothers, king, mother
Female	Mother, daughter, stepmother, child, sister
Both	Mother, father, daughter, son, sister
Neither	Hands, youth, sailors, knights

From Figure 2, we see that in Russian folktales, in contrast to the above, slightly more verbs are used in sentences involving women than in sentences involving men. This pattern also applies to adjectives and nouns quite strongly, as well as for emotion terms and kinship terms (less strongly). Again, Table 2 offers a qualitative look at the data above; I will hold off on analyzing it until the discussion.

I will note that the Russian tales tended to be much longer than the Mexican ones, which explains why the highest counts are nearly double that of the graph in Section 4.1. This is an important difference/limitation to keep in mind when interpreting results in the discussion.

5 Results for Analysis of LLM-Generated Folk Illustration Descriptions

5.1 From Mexico

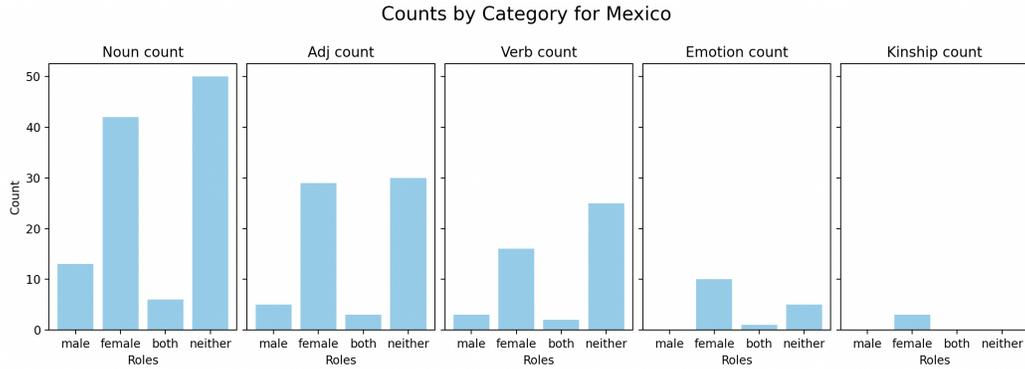


Figure 3: Counts by word-level category for LLM-generated Mexican folktale illustration descriptions.

Table 3: Mexico

Category	Top 5 Words
Adjectives	
Male	Mexican, spectral, aggressive, exaggerated, dominant
Female	Female-presenting, ghostly, dark, emotional, tragic
Both	Crying, generational, marginalized
Neither	Ambiguous, powerful, systemic, traditional, black
Nouns	
Male	Image, figure, legend, horseman, deals
Female	Figure, women, power, sorrow, moon
Both	Boy, woman, hand, element, trauma
Neither	Figure, power, race, identity, culture
Verbs	
Male	Depicts, associated, reflects
Female	Suggest, reflect, clutches, underscore, evoking
Both	Holding, adds
Neither	Depicts, reflects, evoke, are, symbolize
Emotions	
Male	None
Female	Sorrow, grief, desperation, violence, displacement

Category	Top 5 Words
Both	Trauma
Neither	Grief, anguish, menace, haunting, vulnerability
Kinships	
Male	None
Female	Mother, baby, widows
Both	None
Neither	None

From Figure 3, we see that in LLM-generated descriptions for Mexican folktale illustrations, far more of each part of speech or category analyzed (noun, adjective, verb, emotion term, and kinship term count) occur in sentences describing women than in sentences describing men. Outside of the gender binary, the exception to this is the "neither" category: more occurrences are here for the noun, adjective, and verb categories. Again, Table 3 offers a qualitative look at the data above; I will hold off on analyzing it until the discussion.

5.2 From Russia

Table 4: Russia

Category	Top 5 Words
Adjectives	
Male	Older, male, bright, richly, elaborate
Female	Young, ornate, delicate, passive, reliant
Both	Older, young, traditional, protective, dominant
Neither	Cultural, ethnic, homogenous, racial, traditional
Nouns	
Male	Man, power, authority, figure, attire
Female	Woman, garments, safety, predicament, gender
Both	Man, woman, scene, figure, gender
Neither	Representation, setting, narrative, traditions, figures
Verbs	
Male	Signifies, holding, appears, reinforces, convey
Female	Is, stands, wrapped, evoking, represent
Both	Depicts, appears, offer, embody, reinforce
Neither	Is, tied, depicts, engaged, underscores
Emotions	
Male	Power, control, authority, strength, benevolence
Female	Admiration, curiosity, wariness, safety, deference
Both	Dominance, benevolence, subservience, assistance, power
Neither	None
Kinships	
Male	King, noble, man
Female	Woman, man
Both	Man, woman
Neither	None

From Figure 4, we see that in LLM-generated descriptions for Russian folktale illustrations, the counts of nouns and adjectives are highest for sentences solely including men. The verb, emotion terms, and kinship terms counts are fairly evenly distributed across the board of available data. Again, Table 4 offers a qualitative look at the data above; I will hold off on analyzing it until the discussion.

6 Discussion

These results become interesting when I examine the counts and the top 5 word extractions together. Beginning with the textual analysis, we see that sentences referencing men in Mexican folktales

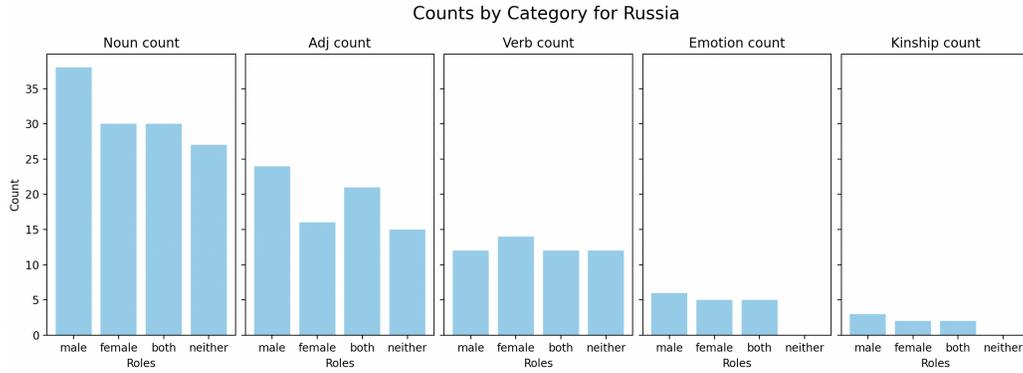


Figure 4: Counts by word-level category for LLM-generated Russian folktale illustration descriptions.

have more verbs than sentences referencing women, both men and women, or neither. Looking at the most frequent verbs in these categories themselves, we see that they are more action-oriented for the male-referring sentences ("went", "asked") than for the female-referring ones ("heard", "gave", "believed"). This seems to me like evidence of how these folktales in the language that they use reinforce gendered stereotypes where men are the "doers" and women are more "emotional". Speaking of emotions, it is often the case that both emotion and adjective counts are higher for women-focused sentences than for other types. This occurs with the generated descriptions for the Mexican stories as well as in the Russian stories themselves. The adjectives chosen for women are very appearance focused (looking at Table 2, there's "beautiful", "ugly" for the female-related sentences and more focus on feelings for the male ones like "lonely", "happy"). This also plays into gendered stereotypes of women being emotional. Additionally, "white" is a common adjective descriptor in the Russian tales; this reflects how a lot of the characters in these stories (if not all) are white.



Figure 5: Russian Image #2. The LLM-generated caption was: "This image, likely from Russian folklore or history, depicts a group of figures engaged in what seems to be an official or ceremonial interaction. On the left side, older male figures, adorned in richly patterned robes and elaborate hats, convey power and authority through their grand attire and commanding postures. On the right, younger men in bright red uniforms, holding axes, appear to be soldiers or guards, signifying their subordinate role in the hierarchy. The central figure's interaction—likely giving orders or pronouncing judgment—reinforces traditional patriarchal power structures, where older, wealthier men hold authority over younger subordinates. The scene lacks female representation, emphasizing a gendered distribution of roles within this narrative. The cultural and racial uniformity in the figures, rooted in Slavic or Russian traditions, underscores a lack of diversity, portraying a homogenous group aligned with the historical and nationalistic aesthetic of the time. The visual cues and poses create a clear stratification, where wealth, age, and role determine one's position in the power dynamic."

The top 5 most frequent words for the LLM-generated folk illustration descriptions are very telling of the LLM's own biases as well as inherent bias in the folktale image descriptions. The adjectives associated with the female-referring sentences are very focused on drastic emotions (see "emotional"

and "tragic" in Table 3) while they reinforce macho-ism for the male-referring sentences ("dominant", "aggressive"). These are very dangerous stereotypes that come to light in the generated image descriptions. We see a similar pattern with the verbs used as with the textual analysis; looking at Table 4, verbs like "evoking" and "is" are used for the female-referring sentences while ones like "holding", "reinforces", and "signifies" are used for the male-referring sentences. I feel as though the latter verbs carry a much stronger tone and connotation with them; they convey a sense of action (again reinforcing the sexist difference I discussed above). The emotions follow the same pattern as the adjectives: in Table 4, terms like "safety" and "deference" are used to describe female characters in illustrations while ones like "power", "control", and "authority" are used to describe male ones. In my eye, I do think that there is a way in which the image descriptions themselves have the differently gendered characters oriented in a way that reinforces these same stereotypes, but the LLM-generated descriptions definitely reinforce them to a degree that is heavily exaggerated.

With regard to race, I would like to point out that the LLM-generated descriptions are not at all good at analyzing nonwhite characters when other characters in the illustrations appear to be white. GPT-4o also makes assumptions around the cultural origin of images (and gets it wrong / assumes Eurocentricity) in the caption. For example, consider Figure 5. Though there is someone who appears to be a person of color (with the black hair), GPT-4o states that everybody in the image is racially uniform or white.



Figure 6: Mexican Image #5. The LLM-generated caption was: "This illustration depicts a medieval-style court scene, where a king or ruler, seated on a throne and adorned with luxurious fur and a crown, interacts with a male figure in armor and a woman in a red dress. The king exudes power and authority, underscored by his position at the center and higher elevation, as well as his commanding posture. The male figure, possibly a soldier or noble, stands respectfully before the king, denoting his subordinate but still active role in the power structure. His weaponry and stance convey readiness and strength, reinforcing a traditional masculine archetype. The woman, with her ornate dress and crown, stands slightly aside, representing beauty and possibly a ceremonial or secondary role, reflecting gender norms that often limit women to symbolic or passive positions in such contexts. The race of all figures aligns with a European aesthetic typical of Russian or Slavic folklore, and there is little to no racial diversity. The surrounding courtiers, mostly male and similarly dressed, reinforce a patriarchal and hierarchical power dynamic. The imagery emphasizes the traditional gendered roles of men as protectors or rulers and women as objects of display or alliances, within a racially homogenous elite."

In the case of Figure 6, the LLM actually assumed that the image was from a Slavic or Russian tale rather than another and perceives the character in the image to be white (when to my eye they clearly appear to be someone of color). I'll note that I had the memory setting turned off when I was prompting (so it was not influenced by my previous prompts for the Russian images — but even if it was!). This is highly concerning and dangerous, and it can reinforce the power paradigm of whiteness when used for downstream image description tasks.

The GPT-4o-classification of the kinship terms seems dubious at best (notice how in Table 4, "man" and "woman" is included). I won't comment specifically on these as a result, and I'll instead recommend that manual classification or better prompting/methods are needed.

7 Conclusion and Future Work

In conclusion, my textual analysis of Mexican and Russian folktales reveals that the tales themselves do reinforce gender stereotypes in their choices of nouns, adjectives, verbs, emotion terms, and kinship terms (for different contexts, each of these apply to varying degrees). My analysis of LLM-generated descriptions of accompanying folktale images show that not only do the images themselves seem to exhibit some degree of positional gender/racial biasing (at least to my eye), the generated descriptions multiply this increasingly and to an alarming degree with how they reinforce similar gender stereotypes as the text and then also make mistakes that favor whiteness in their racial descriptions.

In the future, I would want to expand this work to a much larger subset of folktales (with similar lengths) and images. For the image descriptions, I would want to crowdsource descriptions as well to compare with the LLM-generated ones. I want to use CLIPScore or similar metrics to analyze caption-image informativity. Additionally, I would want to do more complex textual analysis around sentiment, framing, and more in the different gender segmentations.

References

- David Bamman, Jacob Eisenstein, and Tyler Schnoebelen. Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, 18(2):135–160, 2014. doi: <https://doi.org/10.1111/josl.12080>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/josl.12080>.
- Jack Hessel, Ari Holtzman, Maxwell Forbes, Ronan Le Bras, and Yejin Choi. CLIPScore: a reference-free evaluation metric for image captioning. In *EMNLP*, 2021.

A Appendix

A.1 LDA Topic Modeling Results for Folktale Text Analysis

Table 5: Topics and Their Keywords for Mexican Folktales

Male	
Topic 1	possibly, coins, charro, man, horse, fear, gold, bag, well, left
Topic 2	money, never, parents, loved, days, horse, could, time, go, enough
Topic 3	would, devil, place, take, charro, appeared, time, one, however, night
Female	
Topic 1	woman, one, legend, owl, guilt, lechuza, bruja, ritual, also, versions
Topic 2	woman, people, children, said, heard, owl, desperately, cry, deep, wandering
Topic 3	lechuza, owl, believed, women, bruja, witch, legend, families, revenge, next
Both	
Topic 1	lechuza, woman, locals, reference, spanish, take, kill, plan, nuevo, following
Topic 2	child, cry, killed, sometimes, family, lechuza, take, creature, legend, drunkards
Topic 3	child, see, sleep, woman, bright, family, glow, lured, using, trees
Neither	
Topic 1	charro, black, lechuza, said, salt, appears, travel, home, devil, owl
Topic 2	lechuza, also, death, say, charro, due, specific, nature, nahua, still
Topic 3	like, family, tlahuelpuchi, time, however, blood, fire, also, must, belief

Table 6: Topics and Their Keywords for Russian Folktales

Male	
Topic 1	ivan, koshchei, prince, swan, came, deathless, tsar, see, black, hare
Topic 2	prince, ivan, went, time, brothers, sea, came, swan, old, good
Topic 3	prince, tsar, ivan, gvidon, sailors, father, island, saw, came, told
Female	
Topic 1	vasilisa, said, doll, woman, old, one, dog, girl, wide, almost
Topic 2	old, woman, vasilisa, said, doll, go, ask, princess, daughter, beautiful
Topic 3	said, one, doll, vasilisa, stepdaughter, little, last, always, girl, child
Both	
Topic 1	mother, old, ivan, one, koshchei, prince, son, woman, way, time
Topic 2	tsar, old, prince, mother, daughter, see, frost, said, woman, made
Topic 3	ivan, old, prince, mother, said, daughter, man, father, two, wife
Neither	
Topic 1	made, ship, island, corn, casket, one, duck, oak, egg, hare
Topic 2	city, would, forest, come, asked, ring, time, sea, little, give
Topic 3	said, swan, go, golden, city, need, skulls, gate, bless, shall