

What Do LLMs Know About Human Emotions? The Russian Case Study



Olga Mitrofanova, Polina Iurevtseva (Saint Petersburg State University, St. Petersburg, Russia), Maxim Bakaev (Novosibirsk State Technical University, Novosibirsk, Russia)
o.mitrofanova@spbu.ru, st097486@student.spbu.ru, bakaev@corp.nstu.ru

The aim of our research: to test the hypothesis about the sensitivity of LLMs to linguistic markers of emotions in Russian contexts and the hypothesis on applicability of these markers in profiling Emotional AI users.

Our study includes:

- sets of experiments in order to verify psycholinguistic models of emotions (J. Russell’s circumplex model of affect, R. Plutchik’s wheel of emotions),
- development of synthetic personas differing in emotional states and socio-demographic features,
- experiments with LLM using personified emotion-aware prompts,
- evaluation of LLM assessments consistency,
- ANOVA procedures and verification of hypotheses on the differences in the reactions of various synthetic personas to the same emotional stimuli.

Experiment 1.

- Aim:** verification of the hypothesis that LLM and humans generally differ in recognition of emotional meanings of lexical items out of context. We compiled a dataset which included Russian nouns denoting emotions as target words, cf. Table 1. Lexical items selection was based on semantic tagging of contexts in the Russian National Corpus (RNC, <https://ruscorpora.ru>). The search query included *t:psych:emot* tag which provided context samples for 106 names of emotions with frequency over 1 ipm.
- We chose J. Russell’s two-dimensional circumplex model of emotion as it allows to oppose lexical meanings of emotional nouns using two distinct features, valence and arousal.
- We analysed the assessments of emotional meanings obtained from native speakers of Russian and LLM YandexGPT Pro 5 (<https://ya.ru/ai/gp>). Human assessors and LLM were asked to rate the meanings of each stimulus from the list of frequent names of emotions on two scales of valence and arousal in the integer interval from 1 to 10, cf. Fig 1 and Table 2. **The obtained results lead us to the conclusion that LLM and humans structure emotions differently as regards *valence* & *arousal* scale.**

Table 1. List of frequent names of emotions.

Russian	English Translation
антипатия, апатия, азарт, ажиотаж, бешенство, беспаяхитство, беспокойство, безнадежность, безысходность, благодарность, блаженство, влюбленность, волнение, восхищение, восторг, возбуждение, возмущение, знев, гордость, горечь, друсть, досада, доверие, жаслост, желание, экзальтация, забытие, задор, задухчивость, замешательство, зависть, злоба, злорадство, истисг, изумление, , конфуз, любовь, меланхолия, мучение, наорые, наслаждение, неооумение, недовольство, негодование, неустовство, неловкость, ненависть, неприятность, нерешительность, нетерпение, неовольство, неуверенность, нежелание, обида, облечение, осорочение, омерзение, опасение, оплчание, отпороть, операция, ожесточение, озлобление, паника, печаль, переживание, поияржение, презрение, прискорбие, признательность, радость, раскаяние, растерянность, разочаржение, раж, разочарование, ревность, счастье, стылтация, скорбь, скупка, смятение, смущение, сочувствие, сострадание, сожаление, спокойствие, спрадание, стпих, страсть, стыд, тоска, трепет, тревога, удивление, удовлетворение, удовольствие, умление, уныние, упоение, упущение, уважение, увлечение, ужас, экстаз, ярость	antipathy, apathy, excitement, agitation, frenzy, amnesia, anxiety, hopelessness, despair, gratitude, bliss, infatuation, anxiety, admiration, delight, arousal, outrage, anger, pride, bitterness, sadness, annoyance, trust, pity, desire, exaltation, oblivion, enthusiasm, contemplation, confusion, envy, malice, schadenfreude, fear, wonder, embarrassment, love, melancholy, torment, anguish, pleasure, perplexity, dissatisfaction, indignation, frenzy, awkwardness, hatred, trouble, indecision, impatience, discontent, insecurity, unwillingness, resentment, relief, disappointment, disgust, apprehension, despair, shock, aversion, bitterness, hostility, panic, sorrow, worry, trauma, contempt, grief, appreciation, joy, remorse, bewilderment, irritation, fervor, disappointment, jealousy, happiness, sympathy, mourning, boredom, turmoil, embarrassment, empathy, compassion, regret, tranquility, suffering, fear, passion, shame, longing, tremor, anxiety, astonishment, satisfaction, pleasure, tenderness, melancholy, ecstasy, consolation, respect, passion, horror, ecstasy, fury

Experiment 2.

- Aim:** verification of the hypothesis that LLM is sensitive to instructions containing personified information on the emotional state and socio-demographic features of the speakers represented as synthetic personas.
- We restricted the dataset from Experiment 1 to 32 nouns included in R. Plutchik’s wheel of emotions.
- We expanded the dataset by adding collocations for names of emotions from RNC sketches (<https://ruscorpora.ru/page/tool-word>). Collocations provide relevant information on co-occurrence of lexical items in question. RNC sketches use morphosyntactic annotation in collocation analysis.
- For each noun denoting emotions we extracted top-30 collocates (10 adjectives, 10 verbs and 10 nouns) as minimal contexts transmitted to LLM within prompts, cf. example in Table 3.
- We reproduced interaction of humans differing in socio-demographic features and emotional states represented as synthetic personas. YandexGPT-5-Pro was chosen as the language model. Interaction with LLM was performed on behalf of four synthetic personas differing in age and gender (*child* – *adult*, *male* – *female*).

Experiment 2a. Zero-shot mode.

- In Experiment 2a for each persona we developed a series of role-based associative prompts corresponding to the target set of emotions: «Представь, что ты X. Приведи 10 прилагательных, ассоциирующихся со словом Y» / «Imagine you are X. Provide 10 adjectives associated with the word Y», where $X = \{\text{ребенок (child), взрослый (adult), мужчина (male), женщина (female)}\}$ and $Y = \{\text{names of emotions from R. Plutchik’s wheel}\}$. A total of 128 responses were obtained for the formed zero-shot queries containing a role projection of lexical associations with the names of emotions. Table 4 contains examples of LLM associations to the stimulus *восхищение (admiration)* for the personas *child*, *adult*, *male* and *female* respectively.

Table 3. Top-30 collocates (10 adjectives, 10 verbs and 10 nouns) for *восхищение (admiration)*.

Adjectival attributes	Verbs taking target noun as an indirect object	Coordinated nouns
1. неописанный 9,17 2. неописуемый 9,16 3. телячий 8,6 4. неистовый 8,37 5. неподдельный 8,32 6. неизъяснимый 7,55 7. бурный 7,54 8. немой 7,49 9. благоговейный 7,49 10. совершенный 7,2	1. сиять 8,4 2. наполнять 8,09 3. гореть 8,05 4. захлебываться 7,95 5. загораться 7,54 6. светиться 7,52 7. затрепетать 7,41 8. быть 7,28 9. наполняться 7,26 10. засиять 7,18	1. удивление 8,89 2. умиление 8,88 3. упоение 8,79 4. восторг 8,19 5. изумление 8,05 6. восхищение 7,94 7. ужас 7,91 8. благодарность 7,89 9. радость 7,75 10. благоговение 7,47

Table 4. The LLM associations to the stimulus *восхищение (admiration)* for 4 personas.

ребенок (child)	взрослый (adult)	мужчина (male)	женщина (female)
1. Восточное. 2. Волшебное. 3. Удивительное. 4. Яркое. 5. Сказочное. 6. Радостное. 7. Захватывающее. 8. Прекрасное. 9. Чудесное. 10. Незабываемое.	1. Искреннее. 2. Глубокое. 3. Безграничное. 4. Настоящее. 5. Восхищённое. 6. Пристрастное. 7. Грандиозное. 8. Восторженное. 9. Сильное. 10. Умилённое.	1. Искреннее. 2. Грандиозное. 3. Глубокое. 4. Сильное. 5. Восхищённое. 6. Настоящее. 7. Удивляющее. 8. Восторженное. 9. Сильное. 10. Подлинное.	1. Восторженное. 2. Искреннее. 3. Грандиозное. 4. Умилённое. 5. Восхищённое. 6. Неподдельное. 7. Глубокое. 8. Радостное. 9. Сильное. 10. Преклонение.

Experiment 2b. Few-shot mode.

- In experiment 2b for each persona we enriched the prompts, adding «Представь, что ты X и испытываешь эмоцию Y, которая характеризуется следующими контекстными маркерами Z. Назови эту эмоцию.» / «Imagine that you are X and you are experiencing emotion Y, which is characterized by the following contextual markers Z. Name this emotion.», where $X = \{\text{ребенок (child), взрослый (adult), мужчина (male), женщина (female)}\}$, $Y = \{\text{names of emotions from R. Plutchik’s wheel}\}$ and $Z = \{\text{top-30 context markers from RNC sketches}\}$. In response LLM generated 128 explanations of emotional scenarios. Table 5 contains examples of LLM-generated coherent texts corresponding to *восхищение (admiration)* for persona *ребенок (child)*, cf.. Each text generated by LLM contains persona description, physical state associated with the given emotion, factors causing the emotion, recommendations for experiencing emotions, etc.
- We assessed semantic similarity of LLM reactions to various prompts in Experiments 2a and 2b. Responses produced by LLM for each of the 4 personas and for each of the 32 emotion names were vectorized and represented as embeddings by means of Sentence-Transformers and a pre-trained multilingual transformer-based model ru-en-RoSBERTa. Cosine similarity was used as a similarity measure taking values in the range of [0, 1], provided that all vectors are normalized and have non-negative components.

Table 5. LLM-generated texts corresponding to *восхищение (admiration)* for persona *ребенок (child)*.

<i>Представь, что ты ребенок и испытываешь эмоцию восхищение, которая характеризуется следующими маркерами (список). Назови эту эмоцию.</i>	<i>Imagine that you are a child and you are experiencing emotion admiration, which is characterized by the following contextual markers (list). Name this emotion.</i>
<i>Космос, ребенок испытывает восхищение, это очень важное чувство, которое помогает тебе расти и развиваться! Давай разберемся, что происходит, когда ты восхищаешься чем-то! 1. Что ты можешь почувствовать: твой взгляд широко открыт, дыхание становится глубоким, сердце бьется чаще, хочешь улыбаться или даже прыгать от радости, возникает желание поделиться своими чувствами с друзьями... 2. Чем может быть вызвано восхищение: красотой природы (редкие цветы, цветы), величием сказок и историй, удивительными вещами (искусства, картины), достижениями других детей и взрослых, своими собственными успехами...</i>	<i>When a child experiences admiration, it is a very important feeling that helps you grow and develop! Let's explore what happens when you are in admiration of something: 1. What you might feel: your eyes are wide open, your breathing becomes deep, your heart beats faster, you feel like smiling or even jumping for joy, you have a desire to share your feelings with others... 2. What can evoke admiration: the beauty of nature (rainbow, sunset, flowers), the magic of fairy tales and stories, amazing things (toys, pictures), the achievements of other children and adults, one's own successes...</i>

Pairwise comparison of the LLM responses provided 4032 cosine values in each of Experiments 2a and 2b. We calculated the mean values (*M*) and standard deviations (*SD*) of cosine similarity between the lists of associates displayed by the LLM and between coherent texts generated by the. The highest similarity value is observed within the group *ребенок (child)* ($M = 0.653$, $SD = 0.098$) and *мужчина (man)* ($M = 0.621$, $SD = 0.086$), however, for coherent texts generated by LLM results are slightly different: the highest similarity value is observed within the group *взрослый (adult)* ($M = 0.675$, $SD = 0.059$) and *женщина (woman)* ($M = 0.695$, $SD = 0.056$). High similarity values may indicate a more homogeneous and stable semantic structure of associative or textual LLM responses, that gives in perception of emotions.

ANOVA Procedure. In order to determine statistical significance of the differences in cosine similarity mean values in association groups, we put forward a set of hypotheses which were verified by means of ANOVA. In Experiment 2a we used association lists generated by LLM to nouns denoting emotions, in Experiment 2b we considered coherent texts generated by LLM as the *dependent variable*. The cosine similarity values based on embeddings obtained with ru-en-RoSBERTa model were treated as *derived dependent variables*.

The *independent variables* were represented as binary features:

- Is_diff_emotion* (0/1) – difference in the emotion value;
- Is_diff_gender* (0/1) – difference among personas by gender;
- Is_diff_age_group* (0/1) – difference among personas by age group.
- Three hypotheses were formulated and tested:**
- H1.** Association lists generated by LLM have statistically significant differences in the emotion value.
- H2.** Gender of personas influence cosine similarity values for associations.
- H3.** Age of personas influence cosine similarity values for associations.

Table 6. ANOVA results for the considered factors.

Experiment 2a (zero-shot mode)			Experiment 2b (few-shot mode)		
<i>Is_diff_emotion</i>	<i>Is_diff_gender</i>	<i>Is_diff_age_group</i>	<i>Is_diff_emotion</i>	<i>Is_diff_gender</i>	<i>Is_diff_age_group</i>
$F_{1,4030} = 503.23$	$F_{1,2014} = 0.96$	$F_{1,2014} = 6.44$	$F_{1,4030} = 948.29$	$F_{1,2014} = 165.59$	$F_{1,2014} = 101.16$
$p < 0.001$	$p = 0.328$	$p = 0.011$	$p < 0.001$	$p < 0.001$	$p < 0.001$

GitHub Repository:

https://github.com/polly-yu/llm_emotion

The study is performed with partial support of SPbSU research project 124032900006-1.

Conclusion: our findings demonstrate that semantic diversity of associations is determined primarily by differences in the emotional context and, to a lesser extent, by the age characteristics of personas, while gender does not have a significant effect.