

Automated text analysis methods to identify the individual structure of motivation for sports and a healthy lifestyle

Vasily Kashkin¹, Diana J. Haladay

American University of Phnom Penh, Phnom Penh, Cambodia

Abstract In this study, we set two goals: develop and test a method for assessing a person's motivational profile, based on psychological and content analysis of the text and applicable to motivation for sports and a healthy lifestyle and test automated text analysis programs based on artificial intelligence and compare their performance with each other and with traditional expert content analysis. The selection of programs was made from 65 text analysis programs and 12 artificial intelligence-based chatbots. A total of 12 texts were used for the study, including three long interviews and speeches, and 9 short (2-6 sentences) posts on social media and online media. The evaluated texts contain 5,787 words, of which 91% are originally in Chinese and 9% in English. The content analysis methodology was substantiated in detail in our previous publications.

Keywords: personality type, automated text analysis, psychological text analysis, psychometrics, computational psychology, content analysis, psycholinguistics, psychology of sports and healthy lifestyle

1. Introduction

Many private and public organizations set themselves the task of involving people in sports and a healthy lifestyle. For this, it is important to understand the motivational structure of the individual in general and motivation in relation to sports and a healthy lifestyle in particular.

Traditionally, individual motivation is determined on the basis of scientifically validated psychometric tests (14-22). However, it is not always possible to obtain individual testing results. Especially when it comes to mass involvement of citizens in sports and a healthy lifestyle or planning marketing campaigns in this area.

In this case, it becomes especially important to be able to analyze and study the motivational structure of different people based on open data sources. Such open sources include individual author's text and other open information that people publish on social media, blogs and other Internet resources.

In this case, psychological content analysis of the text is used. As a research method, psychological content analysis of text has a long history and a number of scientifically based methods (7).

The novelty of the topic and research methods lies in the fact that psychological content analysis of text is now just beginning to be used in automated programs based on artificial

¹ Corresponding author: kashkin@kashkin.com.cn

intelligence. These are specialized text analysis programs (such as LIWC, Crystal, Symanto, etc.) and chatbots based on neural networks (ChatGPT, Gemini, GigaChat etc.).

In this study, we set the task of testing the applicability of the psychological content analysis method we developed for understanding the motivational structure for playing sports and a healthy lifestyle. And at the same time, we set the task of testing the capabilities of various programs and chatbots for automated text analysis and comparing their results with traditional expert (human) content analysis.

2. Materials and Methods

For preliminary selection, 65 text analysis programs and 12 chatbots based on artificial intelligence were used. Initially, we investigated how well these programs cope with psychological and content analysis of text. These programs have been extensively researched and tested in our previous studies (8,9,10). Therefore, here we will not provide a complete list of programs used for pre-selection. As a result of the selection process, we selected programs and chatbots that demonstrated the greatest capabilities for psychological content-analysis of text.

Only a few programs have options for psychological text analysis. The problems of psychological text analysis are partially solved by the **Crystal, Symanto AI, and Humantic AI** programs. However, we considered their functionality insufficiently complete for this study (we tested the functionality of psychological analysis of the text of these programs in other studies earlier: 8,9,10).

Of all the text analysis programs, we retained only **LIWC** for this study, since only this program specializes entirely in psychological text analysis and presents relevant results.

We also tested programs that extract key themes from text. The best results were shown by **Nocodfunctions, Luminoso Daylight, Intellexer** and **Netlytic**. But the very approach of identifying key topics turned out to be not very productive for the purposes of psychological analysis of the text.

Figure 1 Example of a keyword cloud for text



Source: prepared by the authors using the Netlytic program

Preliminary research on AI-powered chatbots has yielded much better results for content analysis and psychological text analysis. As a result, we chose **ChatGPT, Gemini (Bard)** and **GigaChat**.

The method of expert content analysis was developed by the authors of the article and substantiated in detail in our publication (6). It is based on academically recognized psychometric tests (14-22).

The evaluation criteria were developed by the authors based on a detailed operationalization of the research topic. 6 main groups of assessed parameters were identified, then operationalization was also carried out for each group and the final assessment parameters were obtained. All evaluation parameters are presented below in the table with the research results.

In the text of this publication, we present complete tables for only two groups of parameters: 1) motivation structure 2) psychological characteristics of the individual. The full research material also included groups 3) cognitive styles and intelligence 4) individual goals and values 5) cultural level 6) type of leadership and professional orientation. All tables could not be included in this article due to their large volume, but the statistical data processing below uses materials from all six tables (a total of 215 assessment parameters and 5 methods of assessment, i.e. 1075 resulting data units).

The data for the study included texts of speeches at conferences, oral interviews, and posts on personal social media. All texts belong to one author - a top manager of one of the large Chinese IT companies.

The author's texts were used in both English and Chinese (since the author gave interviews/speaks in both languages).

A total of 12 texts were used for the study, including three long interviews and speeches, and 9 short (2-6 sentences) posts on social media and online media. The evaluated texts contain 5,787 words, of which 91% are originally in Chinese and 9% in English.

In some cases, chatbots and LIWC did a good job of understanding Chinese, in some cases it was necessary to first translate the text into English.

3. Research results

3.1. Structure of individual motivation

Table 1. Structure of individual motivation

Assessment parameter / question	Parameter categories / details	ChatGPT	Gemini (Bard)	Giga Chat	LIWC	Expert content-analysis
Scores on a scale from 0 to +3. 0 – indicator is not expressed; 1 – weakly expressed; 2 – moderately expressed; 3 – strongly expressed. X – indicator cannot be determined using text analysis						
3.1.1. Evaluate the motivation structure of the test author in accordance with the parameters of the Ritchie-Martin Motivational Profile Methodology (14)	Achievement:	3	2	2	X	3
	Recognition:	2	2	X	X	2
	Responsibility:	2	2	X	X	2
	Growth:	3	2	X	X	3
	Interest:	3	2	X	X	3
	Comfort:	0	1	X	X	0
3.1.2. Determine the severity of the value motivations of the author of the texts, using the Portrait Value Questionnaire scale, PVQ by S. Schwartz (15)	Self-Direction:	2	2	X	X	2
	Stimulation:	2	1	2	X	1
	Hedonism:	1	1	0	X	0
	Achievement:	3	2	2	2	2
	Power:	2	0	X	2	1
	Security:	1	1	1	X	1
	Conformity:	0	0	0	X	0
	Tradition:	0	1	0	X	0
3.1.3 Determine the severity of the value orientations of the assessed manager, in accordance with the classification of the Rokeach value orientation test (16)	<i>Terminal Values:</i>	1	1	0	X	0
	Equality:					
	Freedom:	2	1	X	X	1
	Social Recognition:	2	1	X	X	2
	Wisdom:	2	2	2	X	2

	<i>Instrumental Values:</i>	3	2	2	X	2
	<i>Ambition:</i>					
	<i>Responsibility:</i>	2	2	2	X	2
	<i>Competence:</i>	3	2	2	X	2
	<i>Independence:</i>	2	2	X	X	2

Source: developed by the authors

The motivation structure clearly expresses the desire for achievement, recognition, responsibility and personal growth. Interest in the work is also important. This is shown by all assessment tools used. The desire for comfort is rated fairly uniformly by all instruments, as zero or close to zero. In further tests, this assessment is further confirmed by low scores on the desire for hedonism.

This motivational structure is favorable for sports and a healthy lifestyle. The motives of achievement, recognition, and responsibility are expressed, but the motive of comfort and hedonism is minimized. For such an individual, traditional motivators used in sports will work well: orientation towards achieving goals, responsibility for one’s health, sensitivity to incentives and rewards.

Among the value motivations according to the Portrait Value Questionnaire scale, PVQ by S. Schwartz (15), self-orientation and achievements predominate. At the same time, such motivators as hedonism, conformism, security and traditionalism are not expressed at all (absent). This assessment once again confirms other assessments in our study, which characterize the author of the statements as an active person, committed to work, change and technical progress, but not to tradition, conformity, stability and security.

In terms of sports and a healthy lifestyle, such a person may be inclined to romantic and adventure sports, such as outdoor hiking, sports tourism and orienteering, mountaineering, kayaking, and yachting. People with innovative and pioneering thinking tend to search for something new and feel like a pioneer. In addition, the author of the text clearly does not strive for comfort, safety, stability and predictability.

In accordance with the classification of the Rokeach test of value orientations (16), the most pronounced values are ambition, competence, wisdom, independence and responsibility. This once again presents us with the image of an independent, ambitious individual, focused on himself, his own strengths and competencies, and not afraid of responsibility. As in previous tests, the value of equality is absent or minimally expressed, which is quite natural in the value picture of an ambitious leader.

The absence of the value of equality in the individual motivational picture means that a person strives to be the first, the best, rather than inclined to team play. Such a person is more inclined to individual rather than team sports and is more focused on achieving higher results.

3.2. Psychological characteristics of personality

Table 2. Psychological characteristics of personality

Assessment parameter / question	Parameter categories / details	Chat GPT	Gemini (Bard)	Giga Chat	LIWC	Expert content-analysis
Scores on a scale from 0 to +3. 0 – indicator is not expressed; 1 – weakly expressed; 2 – moderately expressed; 3 – strongly expressed. X – indicator cannot be determined using text analysis						
3.2.1. To what extent can the author's style of speech	Straight	2	2	3	1 (20,81)	2
	Open	2	1	3	1 (20,81)	2
	Sincere	2	1	3	1 (20,81)	2

and writing be described as direct, open and sincere?						
3.2.2. To what extent can the author's style of speech and writing be described as manipulative, secretive and evasive? <i>(reverse test question to the previous one)</i>	Manipulative	0	0	X	0	
	Secretive	0	X	X	0	
	Evasive	0	X	X	0	
3.2.3. Determine the severity of the dominant types of behavioral reaction of the leader under study, according to the DISC (ODAT) classification (17)	Dominance:	1	1	2	2	1
	Impact:	2	2	2	2	2
	Consistency:	2	2	2	X	1
	Compliance:	1	1	2	X	1
3.2.4. Assess the severity of socio-psychological characteristics of the individual, in accordance with the parameters of the Sixteen Personality Factor Questionnaire by R. Cattell (18) (5 main groups of factors are presented)	Warmth:	2	2	X	2	2
	Reasoning:	3	2	2	3	3
	Emotional Stability:	2	1	X	X	2
	Dominance:	1	1	0	X	1
	Sensitivity:	1	1	0	X	1
3.2.5. Determine the pronounced accentuations of the personality of the author of the texts, in accordance with the Leonhard-Smishek test (19)	Character Accentuations: Demonstrative:	2	1	0	X	1,5
	Pedantic:	1	0	0	X	1
	Stuck:	0	0	X	X	0
	Excitable:	1	0	0	X	1
	Temperament Accentuations: Hyperthymic:	2	1	X	2	2
	Dythymic:	0	0	X	X	0
	Anxious-Fearful:	0	0	0	X	0
	Cyclothymic:	1	0	X	X	X
	Affective-Exalted:	2	0	0	X	0,5
	Emotive:	1	1	0	X	1
3.2.6. Rate the expression of the basic beliefs of the author of the texts, in accordance with the World Assumptions Scale, WAS (20)	World Benevolence:	2	1	X	2	2
	Justice:	2	2	X	X	1
	Self-Image:	2	2	X	X	2
	Luck:	1	1	X	X	1
	Control Beliefs:	2	1	X	X	2
3.2.7. Does the text rather allow one to define the author as a modest person or, on the contrary, a demonstrative one (prone to self-presentation)?	Demonstrative:	2	1	2	2** (41,46)	2
	Modest:	0	0	2	X	0

*LIWC uses its own indicator “Authenticity”, it is measured in % on a scale of 100

** Clout LIWC indicator, average level (41.46).

Source: developed by the authors

The level of openness, directness and sincerity of the text is assessed as quite high by all assessment tools used. The individual we studied does not shy away from questions and expresses his opinion unambiguously and clearly. The only factor of incomplete sincerity can be considered the propaganda of the services they provide, although even here there is perhaps excessive optimism, but there are no signs of deception and manipulation of facts.

For comparison, we previously conducted a similar study of the speech of a European politician. This speech was characterized by pronounced evasiveness and insincerity: the politician regularly avoided the question asked; did not answer what was asked; reformulated the interviewer's question in his own words, changing the content of the question. In the politician's answers, there were obvious logical inconsistencies and the selection from the general context of only topics that were beneficial to him. In this case, chatbots, expert content analysis and the LIWC system also showed very similar assessment results, rating the politician's speech as insincere, secretive and manipulative. This suggests that the sincerity and openness dimension lends itself well to assessment based on automated text analysis.

For the purposes of sports and a healthy lifestyle, directness, openness and sincerity are positive personality characteristics. Such a person is most capable of achieving his goals, not deceiving himself, and taking responsibility for his own and his family members health.

According to the DISC method (17), the Impact and Consistency parameters were expressed above average.

The Cattell test (18) shows above average levels of the Warmth indicator - this indicator characterizes openness, friendliness, and sociability. The Reasoning parameter is rated at a high level. This parameter means a developed ability for logical and rational thinking. Emotional stability is rated above average in this test.

Analysis of Leonhard-Smishek's temperamental accentuations (19) shows a picture of a mentally stable personality. This is a good indicator for sports, including difficult ones that require perseverance and overcoming oneself.

Of all the accentuations, demonstrativeness is expressed above average. This means that when involving such a person in sports, it is important to show him the visual side of sports achievements - photographs of awards, videos and photos demonstrating sports achievements (awarding athletes, photos of climbing mountains, on a yacht, etc.).

Hyperthymia is expressed above average - and this is also a very favorable indicator for playing sports and leading an active lifestyle.

Basic beliefs are expressed quite clearly in accordance with the World Assumptions Scale, WAS (20). In particular, the World Benevolence, Justice, Self-Image and Control Beliefs indicators are above average. World Benevolence means openness to the world, faith in the goodness of the world. Although the LIWC indicator Emotional Tone does not literally mean the same thing, but rather measures the overall optimism of speech, we believe that in this case they are very similar indicators. The positive Self-image also corresponds to the author's beliefs about his own importance and superiority demonstrated in other tests.

3.3. Indicators according to LIWC classification

Separately, it is necessary to consider the classification of personality characteristics LIWC. LIWC has its own classifier, which does not always coincide with the indicators of other psychometric tests we used in this study.

Table 3. Text evaluation according to LIWC parameters

WC	3387	Word count	WC	Total word count
Analytic	90.88	Analytical thinking	Analytic	Metric of logical, formal thinking
Clout	41.46	Clout	Clout	Language of leadership, status
Authentic	20.81	Authentic	Authentic	Perceived honesty, genuineness
Tone	80.25	Emotional tone	Tone	Degree of positive (negative) tone
WPS	22.14	Words per sentence	WPS	Average words per sentence
BigWords	35.58	Big words	BigWords	Percent words 7 letters or longer
Dic	79.01	Dictionary words	Dic	Percent words captured by LIWC

Source: developed by the authors in the LIWC program based on research data

Most of the results that LIWC provides (in full versions of the tables, brief excerpts are presented here) are percentages of the total number of words in the text. In Table 3 above, several metrics are calculated in a different way: WC word count, WPS average word count, and overall metrics (Analytical Thinking, Clout, Authenticity, and Emotional Tone). Overall scores are based on the LIWC methodology, based on previously studied empirical data, presented in percentiles, ranging from 1 to 99.

Analytical thinking reflects the extent to which people use words that involve formal, logical, and hierarchical patterns of thinking. In our study, the indicator of analytical thinking is high – 90.88.

The influence metric, **Clout**, refers to the relative social status, confidence, or leadership that people demonstrate in their writing or conversations. The author of speech and text in our study has an average score of 41.46. This means that the individual is sufficiently aware of his status and has self-confidence.

The measure of authenticity, **Authenticity**, has less to do with “deception” in the traditional sense, but rather reflects the degree to which a person exercises self-control. Examples of texts that score low on the Authenticity scale include prepared texts (i.e., speeches written in advance) and texts in which a person is socially cautious. In our study, this indicator is 20.81, that is, the author of the speech controls what he is talking about.

Emotional Tone - Although the LIWC-22 includes measures of both positive and negative tone, the Tone variable combines these two measures into one summary variable. The algorithm is designed in such a way that the higher the number, the more positive the tone. Numbers below 50 suggest a more negative emotional tone. In our study, the tone is 80.25, so the overall tone is optimistic and positive. In his speech, the author does not use words associated with conflicts and enmity, as well as curse words. He doesn't show negative emotions.

The **Big Words** indicator (in our case is 35.58) – words consisting of more than 6 letters. People who frequently use long words tend to be less emotional and often psychologically detached.

In the texts used in our study, a large content of **functional words** can be noted. It can be noted that the author often speaks about “us” rather than “himself.” **Social indicator** = 6.02. Social words are words that relate to other people (eg, they, she, we, talk, friends). Generally, people who use a lot of social words are more sociable and more socially connected with others. In our study, this indicator is above average, as is the **Socrefs** indicator, which measures attitudes towards others.

LIWC revealed the active use of words denoting drivers (motivation). The author of the speech often talks about what needs to be achieved, or what the company has already achieved - work, better, best, working, etc.

We can compare our scores to averages across different text categories. LIWC authors publish averages for hundreds of thousands of texts from various sources: Twitter, Conversation, Blogs, New York Times (13, liwc.app/static/documents/). Among the presented categories, the texts of our research can be largely attributed to Conversations and Blogs.

Table 4. Average indicators of LIWC texts in groups Twitter, Conversation, Blogs, New York Times

Category	Twitter		Conv		Blogs		NYT		Others	
Texts (N)	1,000		1,000		1,000		1,000		11,000	
Summary Variables	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
Word count	4442.1	2858.2	586.39	510.42	2144.4	1919.6	744.1	493.6	2101.1	2511.7
Analytic	42.86	27.48	9.46	8.21	38.70	23.20	87.62	12.41	51.27	28.08
Clout	49.10	28.36	38.41	23.27	29.99	23.45	53.90	17.59	49.99	29.27
Authentic	52.33	25.58	69.56	19.97	68.08	23.50	28.90	19.61	49.52	28.39
Tone	68.00	26.36	58.63	27.54	44.99	20.48	37.08	22.30	46.24	25.95
Words/sentence	30.79	100.62	21.51	99.50	14.27	6.03	19.88	6.21	17.01	26.22
Big words	15.98	5.63	9.76	2.46	14.19	4.21	24.77	4.54	17.48	6.70
Dictionary words	83.53	7.54	93.52	2.50	88.30	6.36	80.34	5.61	88.50	4.98

Source: LIWC, liwc.app/static/documents/

If we make a comparison only by general categories, we can note that the author’s speech reflects his analytical thinking to a greater extent, more than the average for each of the presented categories. In addition, more than the average for Conversations and Blogs, the author’s speech reveals his status and self-confidence. On the other hand, the author has a lot of control over his speech; the Authentic score is lower than the average for any of the presented categories. The author also speaks in a more positive tone than the average for all categories presented.

4. Discussion

Evaluations in our study were given on a scale from 0 to 3, where 0 means the indicator is not expressed; 1 – weakly expressed; 2 – moderately expressed; 3 – strongly expressed. X – indicator cannot be determined using text analysis. There are also intermediate ratings of the type 2.5, since in some cases chatbots gave ratings of this nature.

The result is a table with estimates of 215 rows by 6 columns, of which 5 are variables and 1 is formally a row identifier. Data type – nominative/ordinal data. A total of 1075 resulting assessments were analyzed.

For statistical analysis, we set the task to study and compare how consistent the assessments given by chatbots (ChatGPT, Gemini (Bard), GigaChat), LIWC software and human authors of the article are (expContAn - expert content analysis). Variables were analyzed in pairs.

The R language and vcd library were used for data analysis.

Due to the fact that the variables are qualitative, the most popular way to analyze them is to analyze contingency tables. Based on contingency tables, various statistics and measures of connection were calculated:

1) Pearson chi-square. This is a nonparametric method that allows you to assess the significance of differences between the actual (revealed by the study) number of qualitative characteristics of the sample that fall into each category and the theoretical number that can be expected in the studied groups if the null hypothesis is true. The method allows you to assess the statistical significance of differences between two or more relative indicators (frequencies, proportions).

$$\chi_n^2 = n \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

where $O_i = n_i/n$ – the observed frequency of observations n falling into each of the intervals,

$E_i = P_i(\theta)$ – the expected frequencies of observations falling into each interval (theoretical

probability), which correspond to the chi-square distribution. If the difference between the expected and observed frequencies is small (the chi-square has not reached its critical value), we accept the null hypothesis of no differences. If the differences turn out to be significant (the critical chi-square value is reached for a given number of degrees of freedom), we reject the null hypothesis and say that there are statistically significant differences.

2) Chi-square of likelihood ratios. This criterion is not based on the observed-expected frequency difference, but on their observed/expected ratio. Also called G-test. G tests are likelihood ratio or maximum likelihood tests of statistical significance that are increasingly used in situations where chi-square tests were previously recommended.

The general formula for the G test is as follows:

$$G = 2 \sum_{i=1}^n O_i \cdot \ln\left(\frac{O_i}{E_i}\right)$$

A likelihood ratio (LR) test is a statistical test used to test limitations on the parameters of statistical models estimated from sample data. It is one of the three basic tests for checking restrictions, along with the Lagrange multiplier test and the Wald test. The idea of the test is based on comparing the likelihood functions for a long model (without restrictions) and a short model (with restrictions). The likelihood ratio statistic is calculated as follows:

$$G = 2 \sum_{i=1}^n n_i \cdot \ln\left(\frac{n_i}{E_i}\right)$$

3) Phi coefficient (only for 2x2 contingency tables). Obtained by transforming Cramer's V coefficient (if the number of categories in the contingency table is 2, that is, the table is 2x2).

$$\Phi = \sqrt{\frac{\chi^2}{n}}$$

4) Pearson contingency coefficient (contingency coefficient). The disadvantage of the contingency coefficient is that its maximum value depends on the size of the table. This coefficient can reach a value of 1 only if the number of categories is not limited.

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

5) Cramer's contingency coefficient V is more common. Unlike Pearson's contingency coefficient, which is always less than 1, Cramer's coefficient is equal to 1 in a situation of strictly determined relationship between variables.

$$C = \sqrt{\frac{\chi^2}{n(k-1)}}$$

where $k = \min(q, m)$, where in turn q and m represent the number of categories contained in the contingency table of two characteristics (usually X and Y).

To make an unambiguous judgment about the presence of a relationship between variables, Cramer's V coefficient and the p-value of the significance level were used.

Cramer's V ranges from 0 (no relationship between variables) to 1 (complete relationship) and can only reach 1 when each variable is completely determined by the other.

Cramer's V coefficient values were interpreted according to the recommendations of Ray and Parker.

During the analysis, the following values of Cramer's V coefficients and p-level of significance were obtained.

Table 5. Cramer's V coefficients and p-levels of significance

```
[1] "ChatGPT_GeminiBard_V Cramer's = 0.6355 , p-value = 0"  
[1] "ChatGPT_GigaChat_V Cramer's = 0.2723 , p-value = 0.0206629042307633"  
[1] "ChatGPT_LIWC_V Cramer's = 0.6113 , p-value = 0.00115170533428588"  
[1] "ChatGPT_ExpContAn_V Cramer's = 0.6856 , p-value = 0"  
[1] "GeminiBard_GigaChat_V Cramer's = 0.3294 , p-value = 2.15073297548107e-05"  
[1] "GeminiBard_LIWC_V Cramer's = 0.5574 , p-value = 0.00138034794760122"  
[1] "GeminiBard_ExpContAn_V Cramer's = 0.7432 , p-value = 0"  
[1] "GigaChat_LIWC_V Cramer's = 0.446 , p-value = 0.0534792664053121"  
[1] "GigaChat_ExpContAn_V Cramer's = 0.3785 , p-value = 1.76587588551769e-07"  
[1] "LIWC_ExpContAn_V Cramer's = 0.7051 , p-value = 1.13977893561046e-05"
```

Source: developed by the authors

Statistically significant relationships were identified for couples:

- 1 **ChatGPT_GeminiBard** (on the Ray-Parker scale from 0.6 to 0.8 - **strong**);
 - 2 **ChatGPT_GigaChat** (on the Ray-Parker scale from 0.2 to 0.4 - **average**);
 - 3 **ChatGPT_ExpContAn** (on the Ray-Parker scale from 0.6 to 0.8 - **strong**);
 - 4 **GeminiBard_GigaChat** (on the Ray-Parker scale from 0.2 to 0.4 - **average**);
 - 5 **GeminiBard_ExpContAn** (on the Ray-Parker scale from 0.4 to 0.6 - **relatively strong**);
 - 6 **GigaChat_ExpContAn** (on the Ray-Parker scale from 0.2 to 0.4 - **average**);
- Also, statistically significant relationships were identified for 3 pairs with the LIWC variable, however, these pairs had 28-30 observations.
- 7 **ChatGPT_LIWC** – connection strength on the Ray-Parker scale from 0.6 to 0.8 – **strong**;
 - 8 **GeminiBard_LIWC** – on the Ray-Parker scale from 0.2 to 0.4 – **average**;
 - 9 **LIWC_ExpContAn** – on the Ray-Parker scale from 0.6 to 0.8 – **strong**.

To more accurately assess the relationship in small samples, the Chi-square test and p-level of significance were simulated using the **Monte Carlo method** with 100,000 repetitions. As a result, all relationships between variables and LIWC were statistically significant.

Based on the analysis, we can conclude that ChatGPT has a strong correlation with chatbots such as Gemini (Bard), specialized software for psychological text analysis LIWC and expert judgments.

5. Conclusion

The conducted research, in our opinion, confirmed the basic hypothesis that the analysis of texts from open sources allows us to form an idea of the important features of an individual psychological profile and the structure of motivation.

This assessment cannot be complete since we are limited to open source texts only. The themes of the texts do not affect all aspects of the author's life. However, the study showed that at a minimum we can identify the most important features of the motivational structure and psychological profile.

This allows us to assume with a high degree of probability some psychological and motivational patterns of the individual in relation to sports and a healthy lifestyle. The individual studied has a pronounced tendency towards hyperthymia, optimism, and activity.

Motivating factors are achievement, demonstrativeness and responsibility. Hedonism, comfort, stability and security do not play a role in the picture of motivation.

Based on this picture of motivation, it is possible to create individualized proposals (for example, marketing ones) for involvement in sports. In this particular case, it should be a sentence that emphasizes the connection of sport with achievements, superiority over others, search and pioneer tasks. At the visual level, these can be pictures of mountaineering, yachting, or awarding champions.

Such a proposal, drawn up taking into account the motivational structure of the individual, will have a much higher chance of attracting the interest of a particular person and involving him in playing sports and leading a healthy lifestyle.

The hypothesis about the applicability of artificial intelligence tools for psychological text analysis was also confirmed in the study. The results of ChatGPT, Gemini and LIWC for the same parameters showed a significant correlation with each other, as well as with expert content analysis. GigaChat showed a lower level of correlation, indicating that the neural network itself appears to be at an earlier stage of development than the others used in the study.

References

1. Kennedy, B., Ashokkumar, A., Boyd, R. L., & Dehghani, M. (2022). Text analysis for psychology: Methods, principles, and practices. In M. Dehghani & R. L. Boyd (Eds.), *Handbook of language analysis in psychology* (pp. 3–62). The Guilford Press
2. Sumit Pandey, Srishti Sharma, A comparative study of retrieval-based and generative-based chatbots using Deep Learning and Machine Learning, *Healthcare Analytics, Volume 3*, 2023
3. Van Lissa C. J. Developmental data science: How machine learning can advance theory formation in developmental psychology // *Infant and Child Development*. – 2023. – T. 32. – №. 6. – C. e2370
4. Demszky D. et al. Using large language models in psychology // *Nature Reviews Psychology*. – 2023. – T. 2. – №. 11. – C. 688-701
5. Dhingra S. et al. Mind meets machine: Unravelling gpt-4's cognitive psychology // *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*. – 2023. – T. 3. – №. 3. – C. 100139
6. Vasily Kashkin, Yury Androsik. Methods of comprehensive marketing and behavioral economic research based on automated text analytics tools. // *International Business Management*, vol. 14, issue 12, 2020
7. Vasily Kashkin, Yuri Androsik. Review of the development of text analysis methods: from manual processing in psycholinguistics to modern automated programs in marketing. *Digital Economy* vol. 5 (26), 2023.
8. Valeria Zamoshnikova, Vasily Kashkin. Automated analysis of psychometric characteristics of Chinese customers. *Digital Economy*, issue 9, 2023
9. Vasily Kashkin, Valeriya Paliy. Automated Determination of a Behavioral Personality Type Using the Disc Method: Comparative Research of Programs and Chatbots Based on Artificial Intelligence. *Asian Social Science* (Vol.20, No.1 (2024)).
10. Vasily Kashkin, Valeriya Paliy. Automated LinkedIn Analysis to Determine Psychometric Characteristics of a Client. *Asian Social Science* (Vol.20, No.2 (2024)).
11. Boyd, R., Ashokkumar, A., Seraj, S. & Pennebaker, J. The Development and Psychometric Properties of LIWC-22 (Univ. Texas at Austin, 2022).
12. Ashokkumar, A. & Pennebaker, J. W. Social media conversations reveal large psychological shifts caused by COVID-19's onset across US cities. *Sci. Adv.* 7, eabg7843 (2021).

13. Seraj, S., Blackburn, K. G. & Pennebaker, J. W. Language left behind on social media exposes the emotional and cognitive costs of a romantic breakup. *Proc. Natl Acad. Sci.* 118, e2017154118 (2021).
14. S. Ritchie, P. Martin, *Motivation Management* (Gower Pub Co, London, 1999)
15. Schwartz, S.H. (1992), "Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries", In *The Psychology and Values: the Ontario Symposium* Zanna MP (ed.) Erlbaum: Hillsdale, NJ; Vol.8, 1-24.
16. Rokeach, M. *The Nature of Human Values*, Free Press, 1973, 438 p.
17. Marston, W. "The Emotions of Normal People", 1928, 622 p.
18. Cattell, H. E. P., & Mead, A. D. (2008). The Sixteen Personality Factor Questionnaire (16PF). In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment*, Vol. 2. Personality measurement and testing (pp. 135–159). Sage Publications
19. Schmieschek H. Fragebogen zur Ermittlung akzentuierter Persönlichkeiten *Psychiatrie, Neurologie und medizinische Psychologie* Vol. 22, No. 10 (Oktober 1970), pp. 378-381
20. Janoff-Bulman, R. (1989). Assumptive worlds and the stress of traumatic events: Applications of the schema construct. *Social Cognition*, 7, 13–136. <https://doi.org/10.1521/soco.1989.7.2.113>
21. Harrison, A. F., Bramson R. M. *The Art of Thinking: Strategies for Asking Questions, Making Decisions, and Solving Problems*, Berkley Books, 1984, 208 p.
22. Kirton, M. (1976) Adaptors and innovators. A description and measure *Journal of Applied Psychology*, 61, 622-629